mediaLAWS

Rivista di diritto dei media

Numero speciale
I-2024

L'impatto di Internet e dell'intelligenza artificiale sul diritto all'informazione e alla comunicazione

SEDE
CONTACTS
Studio legale Melzi d'Eril Vigevani
Via San Barnaba 32 - 20122 Milano

Università Bocconi - Dipartimento di Studi Giuridici
Via Roentgen 1 - 20136 Milano
e-mail: redazione@rivistadidirittodeimedia.it

ANNO 2024 - CODICE ISSN 2532-9146
Impaginazione a cura di Chiara Pavesi

MediaLaws - Rivista di diritto dei media è una rivista quadrimestrale telematica, ad accesso libero, che si propone di pubblicare saggi, note e commenti attinenti al diritto dell'informazione italiano, comparato ed europeo.

La rivista nasce per iniziativa di Oreste Pollicino, Giulio Enea Vigevani, Carlo Melzi d'Eril e Marco Bassini e raccoglie le riflessioni di studiosi, italiani e stranieri, di diritto dei media.

La rivista fa proprio il Code of Conduct and Best Practice Guidelines for Journal Editors elaborato dal COPE (Comittee on Publication Ethics).

La qualità e il rigore scientifici dei saggi della Rivista sono garantiti da una procedura di *double-blind peer review* affidata a un comitato di esperti per la valutazione individuato secondo criteri di competenza e rotazione e aggiornato ogni anno.

MediaLaws - Rivista di diritto dei media
Regolamento per la pubblicazione dei contributi

1. "MediaLaws – Rivista di diritto dei media" è una rivista telematica e ad accesso aperto che pubblica con cadenza quadrimestrale contributi attinenti al diritto dell'informazione.

2. Gli organi della rivista sono il Comitato di direzione, il Comitato scientifico e il Comitato degli esperti per la valutazione. L'elenco dei componenti del Comitato di direzione e del Comitato scientifico della rivista è pubblicato sul sito della stessa (rivista.medialaws.eu). Il Comitato degli esperti per la valutazione è sottoposto ad aggiornamento una volta l'anno.

3. La rivista si compone delle seguenti sezioni: "Saggi", "Note a sentenza" (suddivisa in "Sezione Europa", "Sezione Italia" e "Sezione comparata"), "Cronache e commenti" e "Recensioni e riletture". I singoli numeri potranno altresì ospitare, in via d'eccezione, contributi afferenti a sezioni diverse.

4. La sezione "Saggi" ospita contributi che trattano in maniera estesa e approfondita un tema di ricerca, con taglio critico e supporto bibliografico.

5. La sezione "Note a sentenza" ospita commenti alle novità giurisprudenziali provenienti dalle corti italiane, europee e straniere.

6. La sezione "Cronache e commenti" ospita commenti a questioni e novità giuridiche di attualità nella dimensione nazionale, europea e comparata.

7. La sezione "Recensioni e riletture" ospita commenti di opere rispettivamente di recente o più risalente pubblicazione.

8. La richiesta di pubblicazione di un contributo è inviata all'indirizzo di posta elettronica submissions@medialaws.eu, corredata dei dati, della qualifica e dei recapiti dell'autore, nonché della dichiarazione che il contributo sia esclusiva opera dell'autore e, nel caso in cui lo scritto sia già destinato a pubblicazione, l'indicazione della sede editoriale.

9. La direzione effettua un esame preliminare del contributo, verificando l'attinenza con i temi trattati dalla rivista e il rispetto dei requisiti minimi della pubblicazione.

10. In caso di esito positivo, la direzione procede ad assegnare il contributo alla sezione opportuna.

11. I saggi sono inviati alla valutazione, secondo il metodo del doppio cieco, di revisori scelti dall'elenco degli esperti per la valutazione della rivista secondo il criterio della competenza, della conoscenza linguistica e della rotazione. I revisori ricevono una scheda di valutazione, da consegnare compilata alla direzione entro il termine da essa indicato. Nel caso di tardiva o mancata consegna della scheda, la direzione si riserva la facoltà di scegliere un nuovo revisore. La direzione garantisce l'anonimato della valutazione.

12. La direzione comunica all'autore l'esito della valutazione.
Se entrambe sono positive, il contributo è pubblicato.
Se sono positive ma suggeriscono modifiche, il contributo è pubblicato previa revisione dell'autore, in base ai commenti ricevuti, e verifica del loro accoglimento da parte della direzione. La direzione si riserva la facoltà di sottoporre il contributo così come modificato a nuova valutazione, anche interna agli organi della rivista. Se solo una valutazione è positiva, con o senza modifiche, la direzione si riserva la facoltà di trasmettere il contributo a un terzo valutatore. Se entrambe le valutazioni sono negative, il contributo non viene pubblicato.

13. Per pubblicare il contributo, l'Autore deve inviare una versione definitiva corretta secondo le regole editoriali della rivista pubblicate sul sito della stessa, un abstract in lingua italiana e inglese e un elenco di cinque parole chiave. Il mancato rispetto dei criteri editoriali costituisce motivo di rigetto della proposta.

14. Le valutazioni vengono archiviate dalla direzione della rivista per almeno tre anni.

15. A discrezione della direzione, i saggi di autori di particolare autorevolezza o richiesti dalla direzione possono essere pubblicati senza essere sottoposti alla procedura di referaggio a doppio cieco ovvero essere sottoposti a mero referaggio anonimo, previa segnalazione in nota.

Indice
Index

Introduzione

Saggi

Note a sentenza

Cronache

Indice
Index

Introduction

Essays

Case notes

Comments

Co-funded by the
Erasmus+ Programme
of the European Union

Finanziato dall'Unione europea. Le opinioni espresse appartengono, tuttavia, al solo o ai soli autori e non riflettono necessariamente le opinioni dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono esserne ritenute responsabili

Sono stati sottoposti a referaggio a doppio cieco i saggi di Matilde Bellingeri-Federica Delaini, Luca Catanzano, Francesco Cirillo, Daniel Foà, Martina Palazzo, Matteo Paolanti, Nicoletta Pica e Urbano Reviglio-Sofia Verza-Konrad Bleyer-Simon, Lorenzo Ricci. Sono stati sottoposti a referaggio anonimo i contributi di Manuela Luciana Borgese, Martina Iemma, Alberto Orlando, Bruno Pitingolo, Giuseppe Proietti, Andrea Ruffo e Dora Trombella.

Introduzione

L'impatto di Internet e dell'intelligenza artificiale sul diritto all'informazione e alla comunicazione

Palmina Tanzarella

Non appena insediatosi alla Casa Bianca, il neo rieletto Presidente degli Stati Uniti Donald Trump ha annunciato l'investimento di almeno 500 miliardi di dollari per lo sviluppo delle infrastrutture dell'intelligenza artificiale. Contestualmente, al Forum economico mondiale di Davos il Segretario generale dell'Onu Antonio Guterres ha avvertito che "ci troviamo di fronte a due nuove e profonde minacce" che rischiano "di sconvolgere la vita come la conosciamo: la crisi climatica e l'espansione scarsamente governata dell'intelligenza artificiale". Inoltre, il Primo ministro spagnolo Pedro Sánchez ha sottolineato come il dominio della tecnologia porti con sé il rischio di erodere la democrazia, denunciando come i social media – nati quali strumenti di unità e giustizia sociale - siano ora usati per manipolare e dividere la società.

L'allarme non è stato lanciato soltanto dal mondo politico. Già da tempo, i giuristi hanno avvertito dei seri rischi che il mondo digitale porta con sé. È indubbio che l'avvento di Internet ha profondamente trasformato il modo di comunicare privatamente e in pubblico, con inevitabili ripercussioni sulla sfera dei diritti fondamentali costituzionalmente protetti, nonché sulla stessa definizione di democrazia.

Tanti sono i diritti che subiscono radicali cambiamenti. Ad esempio, per quanto riguarda l'esercizio della manifestazione del pensiero si possono tracciare differenti nodi problematici: da una prima fase, permeata dall'illusione che finalmente la diffusione delle idee da parte di chiunque non avrebbe incontrato più ostacoli dovuti alla scarsezza delle infrastrutture, è seguita una fase in cui si è presa coscienza di come l'illimitatezza di tale diffusione costituisca un pericolo sia per la compressione di altri diritti sia per la stessa democrazia. Si assiste, infatti, alla proliferazione di fenomeni come la diffamazione online, l'*hate speech*, le *fake news*. Non affatto trascurabile è l'impatto sempre maggiore che avrà l'utilizzo dell'intelligenza artificiale. I sistemi di *machine learning* comportano cambiamenti di enorme portata nell'ambito della produzione, circolazione e ricezione dell'informazione. La selezione degli algoritmi, la loro attività di rimozione e la loro capacità di generare contenuti generano nuovi problemi, quali la difficoltà di riconoscere una protezione del diritto d'autore o la problematica individuazione del regime di responsabilità per la creazione e la diffusione di contenuti che si rivelino dannosi.

Si è posta dunque la necessità di una regolazione della rete e degli strumenti da essa utilizzati che, tuttavia, fatica a trovare risposte adeguate visti i repentini cambiamenti tecnologici. Si assiste così a una costante attività di tipo giuridico, in cui, oltre ai sin-

goli ordinamenti nazionali, l'Unione Europea gioca un ruolo cruciale sia sul versante legislativo che su quello giurisdizionale. Con l'approvazione dell'AI Act, l'Europa si è spinta più avanti di tutti gli altri sulla scena internazionale per assicurare un sistema digitale il più rispettoso possibile di diritti fondamentali; e forse proprio per questo motivo la vigente normativa europea sconta il rischio di rivelarsi, in alcune scelte, precocemente obsoleta.

È tornato in auge il problema di delineare il "giusto" confine tra la diffusione di idee lecite e le idee illecite. Si affacciano nel panorama giuridico nuove norme penali per arginare i discorsi d'odio o falsi, norme che impongono una responsabilità diretta in capo agli *Internet Service Providers*, norme di co-regolamentazione tra autorità pubbliche e gestori della rete. Si tratta di un passaggio davvero significativo, perché i rischi posti dall'evoluzione tecnologica degli ultimi decenni sta spingendo anche i regimi democratici, in rottura con le loro radicate tradizioni costituzionali, a intraprendere forme di controllo e incanalamento dei flussi informativi, specie attraverso forme di collaborazione con le grandi piattaforme digitale.

Come e in che termini questi rimedi possono rivelarsi adeguati? Più in generale, come può rispondere il diritto a queste innumerevoli sfide?

Il presente numero speciale di MediaLaws ha l'ambizione di dare una qualche risposta concreta a tali domande, col pregio di mettere in luce più nel concreto le diverse sfaccettature di una realtà sempre più complessa.

Rispondendo alla *call for papers* – pubblicata dal Dipartimento di Giurisprudenza dell'Università degli Studi di Milano Bicocca nell'ambito del progetto di ricerca europeo *Jean Monnet module* "Diritto a Internet e Diritti fondamentali in Internet nell'era dell'algoritmo" (IntHuR) e dalla sottoscritta coordinato – giovani studiosi si sono cimentati nell'elaborazione di contributi che, sia sotto forma di saggi sia sotta forma di cronaca o nota a sentenza, tentano di mettere ordine alla confuse congerie di tesi, regolamentazioni e giurisprudenza che pervadono il mondo dell'online, senza nondimeno trascurare la visione prospettica rispetto al futuro che ci attende.

Tanti gli interrogativi che permangono. Tuttavia, l'auspicio è quello di alimentare ulteriormente il dibattito grazie a questa *special issue*, la quale non avrebbe potuto vedere la nascita senza il prezioso contributo del Dott. Nicola Canzian, del Dott. Marco Bassini, nonché della redazione, in particolare delle Dott.sse Marina Cazzaniga e Giulia Napoli. A loro tutti porgo i miei più sinceri ringraziamenti.

Milano, 23 gennaio 2025

Saggi

La tutela della libertà di informazione nel *Digital Services Act* tra contrasto alle "manipolazioni algoritmiche" e limiti alla *content moderation*[*]

Nicoletta Pica

Abstract

Il saggio si propone di esaminare le modalità di tutela della libertà di informazione delineate dal Regolamento UE 2022/2065 (*Digital Services Act*).
Dopo aver esaminato il quadro dei rischi cui la libertà di informazione è esposta nel contesto dell'economia digitale, anche per effetto dell'utilizzo sempre più massiccio di algoritmi e modelli di IA, lo scritto si sofferma sulle caratteristiche dell'impianto regolatorio definito dal DSA, individuando due principali direttrici nella regolazione a tutela della libertà di informazione: il contrasto alle "manipolazioni algoritmiche" e l'introduzione di limiti alla *content moderation*.

The paper analyzes the methods of protection of freedom of information outlined by EU Regulation 2022/2065 (Digital Services Act).
After examining the framework of risks to which freedom of information is exposed in the context of the digital economy, also due to the increasingly massive use of algorithms and AI models, the paper focuses on the characteristics of the regulatory system defined by the DSA, identifying two main guidelines in the regulation to protect freedom of information: the fight against "algorithmic manipulations" and the introduction of limits to content moderation activity.

Sommario

[*] L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

Keywords

libertà di informazione – *Digital Services Act* – coregolazione – intelligenza artificiale – *content moderation*

1. Introduzione

Sulla base dell'interpretazione ormai consolidata nella giurisprudenza costituzionale[1], la libertà di informazione trae fondamento dalla libertà di manifestazione del pensiero[2] (art. 21 Cost.)[3], e si declina – nella sua accezione attiva – come diritto di diffondere e comunicare informazioni (diritto di informare), nonché – nella sua accezione passiva – come diritto di ricevere ed essere destinatari di informazioni (diritto di essere

[1] Corte cost., 26 marzo 1993, n. 112.

[2] *Ex multis*, v. C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958; P. Costanzo, *Informazione nel diritto costituzionale*, in *Digesto delle discipline pubblicistiche*, VIII, Torino, 1993, 319 ss.; V. Crisafulli, *Problematica della "libertà d'informazione"*, in *Il Politico*, 1964, 297 ss.; P. Barile, voce *Libertà di manifestazione del pensiero*, in *Enciclopedia del diritto*, XXIV, Milano, 1974, 424 ss.; Id., *Libertà di manifestazione del pensiero*, Milano, 1975; V. Italia, *Considerazioni su propaganda e libertà di manifestazione del pensiero*, in *Scritti in onore di Vezio Crisafulli*, vol. II, Padova, 1985; A. Pace-M. Manetti, *Art. 21. La libertà di manifestazione del proprio pensiero*, in G. Branca-A. Pizzorusso (a cura di), *Commentario della Costituzione*, Bologna-Roma, 2006.

[3] A livello sovranazionale, vengono in rilievo: l'art. 10 CEDU, ai sensi del quale «1. Ogni persona ha diritto alla libertà d'espressione. Tale diritto include la libertà d'opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera. Il presente articolo non impedisce agli Stati di sottoporre a un regime di autorizzazione le imprese di radiodiffusione, cinematografiche o televisive. 2. L'esercizio di queste libertà, poiché comporta doveri e responsabilità, può essere sottoposto alle formalità, condizioni, restrizioni o sanzioni che sono previste dalla legge e che costituiscono misure necessarie, in una società democratica, alla sicurezza nazionale, all'integrità territoriale o alla pubblica sicurezza, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, alla protezione della reputazione o dei diritti altrui, per impedire la divulgazione di informazioni riservate o per garantire l'autorità e l'imparzialità del potere giudiziario»; l'art. 11 della Carta dei diritti fondamenti dell'Unione Europea, in forza del quale: «1. Ogni individuo ha diritto alla libertà di espressione. Tale diritto include la libertà di opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera. 2. La libertà dei media e il loro pluralismo sono rispettati». Per l'esame della giurisprudenza della Corte di Strasburgo e della Corte di Giustizia, v. M. Bassini, *Internet e libertà di espressione. Prospettive costituzionali e sovranazionali*, Roma, 2019, 191 ss., che, analizzando le suddette disposizioni anche in chiave comparata, evidenzia il diverso approccio dell'ordinamento statunitense: «Mentre il Primo Emendamento (non a caso, per l'appunto, "primo" nell'elencazione contenuta nel *Bill of Rights*) riverbera l'estremo afflato libertario proprio di quella cultura giuridica, premurandosi di vietare al potere pubblico ogni interferenza con il diritto di parola, l'art. 10 CEDU, in linea del resto con le costituzioni nazionali, sceglie una via diversa, frutto probabilmente anche del diverso periodo storico in cui tali documenti sono stati formati. […] si assiste a una relativizzazione per definizione, vale a dire "in partenza", dell'ambito di tutela di questa libertà, costretta entro limiti che sono normalmente tipizzati espressamente dai parametri costituzionali o convenzionali (si pensi all'art. 10, par. 2, CEDU) di riferimento, oltre che ricavabili implicitamente dall'ordinamento giuridico. Emerge così una diversa connotazione rispetto alla natura "assolutizzante" del Primo Emendamento. Pur conoscendo anche negli Stati Uniti possibili restrizioni, la libertà di espressione nasce in Europa "limitata". È paradigmatico quanto prevede il già richiamato par. 2 dell'art. 10 della CEDU, che indica le tre condizioni alle quali le restrizioni possono considerarsi legittime».

informati)[4].

Invero, nonostante la disposizione costituzionale si riferisca espressamente al solo versante "attivo", si è notato come la libertà di manifestare il proprio pensiero non possa non comprendere anche la «libertà di prendere conoscenza del pensiero manifestato, così come nella libertà di informare non può non essere compresa la libertà di accedere alle informazioni»[5].

Quest'ultima, peraltro, espressamente tutelata dall'art. 10 CEDU[6], trova un ulteriore aggancio costituzionale nel principio democratico, che concorre ad attuare nella misura in cui consente la formazione di un'opinione pubblica libera e consapevole.

È significativo, a tal proposito, che la prevalente impostazione dottrinale desuma il fondamento costituzionale del diritto all'informazione non tanto dall'art. 21 Cost.[7], che ------ si è detto - garantirebbe espressamente il solo diritto di informare quale corollario della libertà di espressione, quanto dall'interpretazione sistematica del quadro costituzionale[8], configurandolo come principio costituzionale, oltre che come situazione giuridica soggettiva[9].

Segnatamente, secondo la dottrina in parola, essendo il diritto all'informazione strumento di «formazione di un'opinione pubblica documentata»[10] e, perciò, imprescindibile presupposto del principio democratico[11], a venire in rilievo sarebbero tutte le disposizioni rivolte al pieno sviluppo della persona (art. 2, 3, II comma), l'eguaglianza (art. 3), la sovranità popolare (art. 1) nonché la partecipazione all'organizzazione del

---

[4] G. Gardini, *Le regole dell'informazione*, Torino, 2021, 43, che richiama anche una terza accezione della libertà di informazione: «una libertà in senso riflessivo, per cui il titolare è legittimato a informarsi, andare alla ricerca (cd. *inspectio*) di notizie concernenti fatti, stati e situazioni rispetto alle quali intende esercitare il proprio diritto all'informazione e alla conoscenza».

[5] M. Cuniberti, *Costituzione e mezzi di comunicazione*, in G. E. Vigevani-O. Pollicino-C. Melzi d'Eril-M. Cuniberti-M. Bassini, *Diritto dell'informazione e dei media*, Torino, 2019, 242.

[6] M. Bassini, *Internet e libertà di espressione* cit., 194: «Non è priva di pregio e di conseguenze pratiche la scelta di abbracciare nel perimetro di tutela sia il c.d. profilo attivo, relativo alla divulgazione di informazioni e opinioni, sia il c.d. profilo passivo, relativo alla loro ricerca e recezione. Per esempio, in tempi recenti, nell'ambito del dibattito inerente al contrasto alla disinformazione in rete, si è evocato il diritto a ricevere informazioni (su un presupposto carattere "qualificato" delle stesse) come un possibile fondamento per l'adozione di misure volte a contrastare la circolazione di notizie false o tendenziose».

[7] Come, invece, sostenuto da: V. Cuffaro, *Profili civilistici del diritto all'informazione*, Napoli, 1986, 33 ss.; N. Lipari, *Libertà di informare o diritto ad essere informati?*, in *Diritto radiodiff. e telecom.*, 1978, 4 ss.; A. Pace, *Libertà di informare e diritto ad essere informati: due prospettive a confronto nell'interpretazione dell'art. 7, primo comma, del T.U. della radiotelevisione*, in *Dir. pubbl.*, 2007, 312 ss.

[8] A. Loiodice, *Il diritto all'informazione: segni ed evoluzione*, in M. Ainis (a cura di), *Informazione Potere Libertà*, Torino, 2005, 38; id, *Contributo allo studio sulla libertà di informazione*, Napoli, 1969, 64 ss.; id, (voce) *Informazione* (*diritto alla*), in Enc. dir., vol. XXI, Milano, 1971, 478 ss.

[9] A. Loiodice, *Il diritto all'informazione* cit., 36.

[10] Ivi, 25. In merito al rapporto tra informazione e opinione pubblica, v., A. Papa, *Democrazia della comunicazione e formazione dell'opinione pubblica*, in *Federalismi.it*, 2017, 1.

[11] La libertà di informazione è stata definita dalla Corte «pietra angolare dell'ordinamento democratico» (Corte Cost., 17 aprile 1969 n. 84, in *Consulta OnLine*); essa «nei suoi risvolti attivi e passivi (libertà di informare e diritto ad essere informati), esprime [...] una condizione preliminare (o, se vogliamo, un presupposto insopprimibile) per l'attuazione ad ogni livello, centrale o locale, della forma propria dello Stato democratico» (in termini, Corte Cost. 20 luglio 1990 n. 348; 12 febbraio 1996 n. 29; 15 ottobre n. 312, in *Consulta OnLine*).

Paese (art. 3, II comma)[12].

Ed infatti, come più volte ribadito dalla Corte Costituzionale[13], il diritto in parola, che non può essere concepito come mera libertà negativa, ovvero come diritto di accedere alle informazioni senza ingiustificate restrizioni da parte dell'autorità pubblica, deve essere «determinato e qualificato in riferimento ai principi fondanti della forma di Stato delineata dalla Costituzione, i quali esigono che la nostra democrazia sia basata su una libera opinione pubblica e sia in grado di svilupparsi attraverso la pari concorrenza di tutti alla formazione della volontà generale».

Da tale premessa scaturiscono conseguenze sul piano qualitativo, atteso che – lo ha chiarito ulteriormente la Corte - la funzione assolta dal diritto all'informazione può pienamente realizzarsi solo se quest'ultimo risulti caratterizzato dal pluralismo delle fonti informative «in modo tale che il cittadino possa essere messo in condizione di compiere le sue valutazioni avendo presenti punti di vista differenti e orientamenti culturali contrastanti».

In altri termini, se, come osservato dalla dottrina, pare difficile configurare il diritto a un'informazione di qualità o veritiera[14], «l'ideale punto di equilibrio del sistema» può essere colto «nella garanzia del pluralismo dei media, condizione in grado di soddisfare l'aspettativa dei cittadini a che il sistema dell'informazione corrisponda massimamente ai criteri di imparzialità, obiettività e completezza»[15].

In altri termini, il pluralismo informativo, consentendo che le molteplici posizioni culturali e politiche presenti nella società possano aver voce (pluralismo interno) e che le posizioni dominanti nel settore dell'informazione non rendano marginali quelle minoritarie (pluralismo esterno)[16], permette di realizzare l'intima connessione del diritto all'informazione con il concetto di democrazia fondata sull'opinione pubblica[17].

Si tratta di un principio che tanto la Corte costituzionale, quanto la giurisprudenza sovranazionale, sovente hanno avuto modo di affermare, evidenziando il «valore centrale del pluralismo in un ordinamento democratico»[18] e la conseguente necessità che

---

[12] A. Loiodice, *Il diritto all'informazione*, cit., 36. In tal senso anche P. Barile, *Diritti dell'uomo e libertà fondamentali*, Bologna, 1984, 232; R. Zaccaria, *Diritto dell'informazione e della telecomunicazione*, Padova, 2007, 18 ss.; P. Barile-E. Cheli-S. Grassi, *Istituzioni di diritto pubblic*o, Padova, 2007, 416; P. Caretti, *Diritto dell'informazione e della comunicazione*, Bologna, 2004, 23 ss. Recentemente, v. O. Pollicino, *Tutela del pluralismo nell'era digitale: ruolo e responsabilità degli Internet service provider*, in *Consulta OnLine*, 6 osserva: «Nell'ordinamento italiano, l'art. 21 della Costituzione non offre, sotto un profilo testuale, una chiara base normativa al diritto di informazione. Piuttosto, è soprattutto alla luce dell'art. 10 della CEDU che diviene possibile individuare, nell'ambito di tutela offerto alla libertà di manifestazione del pensiero, anche la libertà di comunicare e ricevere informazioni».

[13] Corte cost. 26 marzo 1993, n. 112.

[14] M. Bassini, *Internet e libertà di espressione*, cit., 297.

[15] M. Bassini, *Internet e libertà di espressione*, cit., 298, che, sul punto, richiama M. Cuniberti, *Costituzione e mezzi di comunicazione*, cit., 244.

[16] V. Pampanin, *Tutela del pluralismo informativo e regolazione economica nel mercato convergente della comunicazione*, in G. Avanzini-G. Matucci (a cura di), *L'informazione e le sue regole*, cit., 170 evidenzia come «questi due modi di intendere e attuare il pluralismo informativo si rivelano peraltro chiaramente collegati tra loro, nella misura in cui è intuitivo che l'esigenza di più voci e di più operatori rappresenta una precondizione per avere nel panorama dell'informazione anche una diversità di contenuti».

[17] A. Loiodice, *Il diritto all'informazione*, cit., 25.

[18] *Ex multis*, Corte cost., 14 luglio 1988 n. 826, in *Consulta OnLine*.

lo Stato[19] approntì «un quadro legislativo e amministrativo appropriato per garantire un pluralismo effettivo nei media»[20].

Se nel settore radiotelevisivo, storicamente investito dal problema, il panorama normativo si è progressivamente allineato ai moniti delle Corti[21], la Rete apre scenari inediti in ordine all'attuazione di un principio ormai consolidato[22].

Invero, non v'è dubbio che l'accesso a Internet abbia rafforzato la libertà di manifestazione del pensiero, se non altro per le peculiari modalità con cui viene esercitata in Rete; quest'ultima, infatti, non solo consente di divulgare le proprie opinioni, ma permette anche di condividere quelle altrui e discuterle con una platea potenzialmente illimitata di interlocutori[23], confermando – oggi ancor più che in passato – la necessità di ricomprendere nell'ambito oggettivo dell'art. 21 Cost., non solo le proprie idee e le notizie frutto della propria conoscenza (ovvero il "proprio pensiero" cui espressamente si riferisce l'art. 21 Cost.), ma anche la citazione o diffusione di pensieri altrui[24].

Al contempo, con l'avvento del web 2.0 anche la libertà di informazione ha assunto connotati del tutto nuovi, sia sotto il profilo "passivo" (diritto di essere informati), in considerazione dell'enorme mole di informazioni, dati e notizie fruibili in rete[25],

---

[19] In dottrina si esclude che tale diritto possa essere fatto valere nei confronti degli operatori privati dell'informazione, pena il rischio di «trasformare la posizione "attiva" garantita primariamente dall'art. 21 (cioè il diritto di informare) da "libertà", come la qualifica la norma costituzionale, in "funzione" o "dovere"» (in termini, M. Cuniberti, *Costituzione e mezzi di comunicazione* cit., 243).

[20] CEDU, Grande Camera*, Centro Europa 7 e Di Stefano c. Italia*, ric. 38433/09 (2012), in *giustizia.it*.

[21] Per un'accurata ricostruzione dell'evoluzione normativa in materia v. V. Pampanin, *Tutela del pluralismo informativo e regolazione economica* cit., 163 ss. Sull'argomento, v. G.E. Vigevani, *I media di servizio pubblico nell'età della rete. Verso un nuovo fondamento costituzionale, tra autonomia e pluralismo*, Torino, 2018; M. Manetti, *Pluralismo dell'informazione e libertà di scelta,* in *Rivista AIC*, 2012, I; E. Apa, *Il nodo di Gordio: informazione televisiva, pluralismo e Costituzione,* in *Quaderni costituzionali,* 2, 2004, 335 ss. In merito alla giurisprudenza costituzionale, v. B. Tonoletti, *Principi costituzionali dell'attività radiotelevisiva*, Torino, 2003, 215 ss.

[22] V. O. Pollicino, *Tutela del pluralismo nell'era digitale* cit..; F. Donati, *Democrazia, pluralismo delle fonti di informazione e rivoluzione digitale*, in *Federalismi.it*, 20 novembre 2013; O. Grandinetti, *La par condicio al tempo dei social, tra problemi "vecchi" e "nuovi" ma, per ora, ancora tutti attuali*, in questa *Rivista*, 3, 2019, 126 ss.; M. Monti, *Le internet platforms, il discorso pubblico e la democrazia*, in *Quaderni costituzionali*, 4, 2019, 822 ss.

[23] C. Capolupo, *Informazione e partecipazione democratica nell'era dei social media*, in M. Villone-A. Ciancio-G. De Minico-G. Demuro-F. Donati (a cura di), *Nuovi mezzi di comunicazione e identità: omologazione o diversità?*, Roma, 2012, 604.

[24] Sul punto, v. M. Orofino, *Art. 21 Cost.: le ragioni per un intervento di manutenzione ordinaria*, in questa *Rivista*, 2, 2019, 85: «La specifica qualificazione del pensiero espressa dal pronome possessivo ha fatto sorgere il dubbio che l'ambito oggettivo della libertà comprendesse solo espressioni, idee e notizie frutto del proprio personale pensiero o della propria diretta cognizione e non coprisse, dunque, la mera ripetizione, citazione o diffusione di pensieri altrui. La Corte costituzionale, chiamata ad esprimersi sul punto, non ha assecondato una lettura così restrittiva della libertà in questione, argomentando, al contrario, che nulla vieta che un pensiero o un'informazione altrui sia fatta lecitamente propria e diffusa. Per cui anche le notizie, i fatti di attualità, le conoscenze e, più in generale, le informazioni acquisite da altri rientrano nell'ambito oggettivo tutelato dalla norma anche quando esse sono diffuse da altre persone. Così facendo ha offerto una lettura interpretativa della norma secondo la quale "il possessivo proprio, riferito al pensiero, non intende esprimere un'appartenenza» ma «sottolineare il valore dell'autonomia del singolo, l'indipendenza di giudizio". È intuitivamente evidente come tale precisazione sia di grande rilevanza oggi. […] L'interpretazione dell'art. 21 Cost. offerta dalla Corte, volta a non conferire particolare rilevanza all'articolo possessivo, ha il pregio di consentire oggi di ricondurre all'ambito oggettivo della libertà in questione tanto i *retweet* e gli *sharing* che caratterizzano i *social network*, quanto la pubblicazione di contenuti altrui sui siti di condivisione»

[25] Secondo A. Ciancio, *Il pluralismo alla prova dei nuovi mezzi di comunicazione*, in A. Ciancio (a cura

sia sotto il profilo "attivo" (diritto di informare), atteso che gli utenti, da meri destinatari dei contenuti informativi trasmessi dai media tradizionali, divengono fautori dell'informazione veicolata nella Rete, che, ormai trasformatasi nel principale canale informativo[26], consente loro di diffondere notizie autonomamente[27], senza costi[28] e – apparentemente - senza intermediazione[29].

Ciononostante, è sempre più evidente come, nel contesto dell'economia digitale, la libertà di informazione risulti esposta a svariati rischi.

Pur nella consapevolezza dei molteplici profili di interesse sotto cui la questione può essere indagata, l'analisi si svilupperà esaminando, nella prima parte, i più rilevanti fenomeni da cui detti rischi traggono origine, per poi verificare, nella seconda parte dello scritto, come la recente regolazione europea abbia reagito alle minacce cui la libertà di informazione è esposta nei mercati digitali, anche per effetto dell'utilizzo sempre più massiccio di algoritmi e modelli di IA.

2. Il quadro dei rischi per la libertà di informazione

Il primo fenomeno è strettamente connesso all'affermazione di un *business model* incentrato sullo sfruttamento dei dati mediante l'utilizzo di algoritmi e tecniche di intelligenza artificiale (*machine learning* e *deep learning*) da parte delle piattaforme operanti nel web[30].

Invero, è ormai noto come la profilazione[31], solitamente preordinata a finalità com-

di), *Il pluralismo alla prova dei nuovi mezzi di comunicazione*, Torino, 2012, 31, l'aspetto più significativo dell'esercizio della libertà di informazione risiederebbe proprio sotto il profilo "passivo", nell'esercizio della libertà di informarsi, «poiché il web rende disponibili e attingibili una quantità oggi incalcolabile di informazioni, di dati, di notizie, sui più svariati oggetti e argomenti, fruibili per effetto di un semplice atto di volontà del "cybernavigatore", praticamente senza limiti geografici, di tempo o di materia».

[26] V., sul punto, i dati riportati da G. Pitruzzella, *La libertà di informazione nell'era di Internet*, in G. Pitruzzella-O. Pollicino-S. Quintarelli, *Parole e potere. Libertà d'espressione, hate speech e fake news*, Milano, 2017, 61.

[27] Come rilevato da M. Orofino, *Art. 21 Cost.: le ragioni per un intervento di manutenzione ordinaria*, cit., 84, quest'aspetto conferma l'impossibilità di distinguere concettualmente l'attività di manifestazione del proprio pensiero e l'attività di informazione, atteso che nel web 2.0 «ogni espressione pubblica tende a divenire informazione alla luce dell'enorme platea di destinatari potenzialmente raggiungibili». Invero, «se si concorda sul fatto che ogni espressione del pensiero umano contenga in sé elementi informativi per i destinatari della comunicazione, non può che concludersi che la distinzione tra manifestazione e informazione sia artificiale ed essenzialmente legata all'idea che l'informazione sia solo quella veicolata attraverso i tradizionali mezzi di comunicazione (stampa, radio, televisione). Come si è già detto, nel web 2.0 da un lato si assiste all'emersione di una pluralità di nuovi servizi che consentono una diffusione anche maggiore rispetto ai mezzi tradizionali e dall'altro viene meno quel confine, già di per sé labile, tra la volontà di effettuare una semplice manifestazione del pensiero e la diffusione a scopo informativo».

[28] F. Donati, *Il principio del pluralismo delle fonti informative al tempo di Internet*, in *Diritto e Società*, 4, 2013, 663.

[29] Di «*informazione disintermediata*» parla C. Capolupo, *Informazione e partecipazione democratica* cit., 622.

[30] Per una approfondita ricostruzione, v. F. Chirico-A. Manganelli, *Mercati e servizi digitali*, in C. Cambini-A. Manganelli-G. Napolitano-A. Nicita (a cura di), *Economia e diritto della regolazione*, Bologna, 2024, 367 ss.

[31] Pertanto può ritenersi condivisibile l'affermazione secondo cui «l'attività di profilazione merita un'osservazione diversa e specifica da parte del diritto, in funzione della caratteristica, propria di tale tecnica, di coinvolgere situazioni giuridiche che vanno oltre la compromissione della privacy − pur

merciali, sia utilizzata anche per la personalizzazione dei contenuti informativi[32], che, selezionati dall'algoritmo[33] alla stregua di criteri costantemente aggiornati, come le precedenti ricerche dell'utente o il tempo di visualizzazione di un singolo contenuto[34], vengono fatti convergere nelle cosiddette "bolle di filtraggio" (*filter bubble*[35]).

Come posto in luce dalla dottrina, i meccanismi preordinati alla personalizzazione dei contenuti informativi sono costruiti sul paradigma della «sovranità del consumatore»[36], il quale può effettivamente trarre benefici da un'offerta commerciale ritagliata sul suo profilo[37]; si tratta, però, di un modello suscettibile di esiti preoccupanti se trasposto al settore dell'informazione, atteso che ciascun utente, confinato nella propria bolla informativa, sarà indotto alla *confirmation bias*, tenderà cioè ad acquisire solo notizie, pareri e informazioni coerenti con le proprie opinioni[38], e lo farà all'interno di gruppi attestati su posizioni univoche, che, incontaminati da visioni alternative, tendono a trasformarsi in casse di risonanza (*eco chambers*)[39] in cui le opinioni si amplificano e radicalizzano, polarizzandosi.

Certo, il fenomeno della polarizzazione non è nuovo[40], ma risulta indubbiamente acu-

di per sé grave − per investire altri diritti legati alle libertà fondamentali» (in terminis, R. De Meo, *Autodeterminazione e consenso nella profilazione dei dati personali*, in *Dir. inf.*, 3, 2013, 587 ss.).

32 Facebook, ad esempio, propone le informazioni sulla base delle affinità dichiarate (ad. es. le liste di amici) o desunte attraverso i *like* (M. Gambaro, *Concorrenza e pluralismo nel mercato di internet*, in T. E. Frosini-O. Pollicino-E. Apa-M. Bassini (a cura di), *Diritti e libertà in Internet*, Milano, 2017, 279).

33 Il funzionamento degli algoritmi di ricerca e *story selection*, utilizzati ad esempio da Google e Facebook, sono dettagliatamente spiegati da P. Costa, *Motori di ricerca e social media* cit., 259 ss.

34 G. De Gregorio, *The market place of ideas nell'era della post-verità: quali responsabilità per gli attori pubblici e privati online?*, in questa *Rivista*, 1, 217, 94.

35 La definizione risale allo scritto di E. Parisier, *The Filter bubble: what the Internet is hiding from you*, New York, 2011.

36 G. Pitruzzella, *La libertà di informazione nell'era di Internet*, cit., 68.

37 Tuttavia, profili di criticità potrebbero emergere anche sotto tale profilo. Invero, la piattaforma «colloca i consumatori in un mercato secondario dell'informazione (*information aftermarket*), ovvero un mercato secondario in cui i consumatori effettuano le loro scelte dopo aver acquistato un prodotto o un servizio primario, con una libertà di scelta *ex post* che è vincolata dalle scelte precedenti nel mercato primario. In un certo senso, quando i consumatori scelgono una piattaforma digitale, scelgono anche un *gatekeeper* che li indirizza vero il mercato (secondario). Lì, i consumatori sono indotti a esercitare la loro libertà di scelta entro i limiti decisi dal *gatekeeper*, che preseleziona offerte su misura basate sulle informazioni digitali che gli algoritmi delle piattaforme possono estrapolare dai dati dei consumatori stessi» (F. Chirico-A. Managanelli, *Mercati e servizi digitali*, cit., 374).

38 A. Peruzzi-F. Zollo-A. L. Schmidt-W. Quattrociocchi, *From Confirmation Bias to Echo-Chamber*, in *Sociologia e Politiche Sociali*, 3, 2018, 54. Tra i numerosi contributi v. anche M. Mezzanotte, *Fake news nelle campagne elettorali digitali. Vecchi rimedi o nuove regole?*, in *Federalismi.it*, 19 dicembre 2018, 3, che richiama le osservazioni di W. Quattrocchi, *La babele dell'Internet*, su *Le Scienze*, aprile 2018, 39.

39 P. Costa, *Motori di ricerca e social media* cit., 254 osserva come «anziché connettere individui con punti di vista e ideologie differenti, i *social media* tendono a rafforzare i pregiudizi, a causa dell'effetto di riverbero ("*echo chamber effect*"), ossia la tendenza dell'informazione a rimbalzare all'interno di sistemi chiusi».

40 A. Ciancio, *Il pluralismo alla prova dei nuovi mezzi di comunicazione*, cit. 16-17 osserva che già la pay-tv ha determinato la frammentazione del pubblico, consentendo all'utenza di scegliere i programmi e le informazioni cui attingere e per cui pagare, facendo sorgere «il rischio di autoisolamento degli individui e, più a monte, dei gruppi di cui i primi fanno parte, i quali, attingendo al campo delle notizie che prediligono e ritengono interessanti, si precludono la possibilità di aprirsi al confronto con opinioni ed idee differenti, con intuibili conseguenze circa l'effettiva affermazione di una circolazione plurale delle

ito dalla fruizione dei *social network*, al cui interno gruppi affini alle proprie idee non solo possono essere più agevolmente rintracciati, ma vengono altresì suggeriti dall'algoritmo della piattaforma, che ha interesse ad inserire l'utente in gruppi omogenei e radicalizzati[41] «perché questo importa maggiori interazioni, spinge gli utenti ad un interesse frenetico e sempre nuovo per i temi più disparati, che diventano più attrattivi se semplici, condensati, divisivi, settari»[42].

I rischi per la libertà di informazione sono evidenti se solo si considera che, specie quando l'informazione è di tipo politico[43], solo il pluralismo e la libertà di autodeterminazione in ordine alle informazioni fruibili giovano alla «sovranità del cittadino»[44], caposaldo dell'ordinamento democratico.

Non a caso, parte della dottrina si appella alla valorizzazione della sovranità popolare di cui all'art. 1 della Carta costituzionale quale strumento di adeguamento dell'intelligenza artificiale al quadro costituzionale, potendo «il primo articolo della Costituzione italiana […] costituire la base per tentare di impedire un impiego incontrollato della AI in termini di disinformazione e propaganda politica; un impiego che rischia di generare un modello diffuso di *bubble democracy*, in cui i cittadini, anche in quanto elettori, sono confinati all'interno di sistemi di *social* sempre più chiusi e autoreferenziali, che li portano a ritenere le proprie idee le uniche plausibili, che impediscono un confronto reale e pluralista fra posizioni diverse [...] con un sostanziale svuotamento dall'interno

idee e dei valori».

41 R. Montaldo, *La tutela del pluralismo informativo nelle piattaforme online*, in questa *Rivista*, 1, 2020, 227. In generale, sull'argomento v. M. Delmastro-A. Nicita, *Big data*, Bologna, 2019, 95.

42 F. Severa, *La dissoluzione dello spazio pubblico. Il fattore "tecnologico" tra geodiritto e geopolitica*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 809.

43 A tal proposito, C. Bologna, *Libertà di espressione e riservatezza "nella rete"? Alcune osservazioni sul mercato delle idee nell'agorà digitale*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 71 osserva: «Il modello di un confronto trasparente (e consapevole) rischia di esser ancor più alterato dal c.d. *micro-targeting* politico, che applica meccanismi analoghi a quelli del *marketing online*, utilizzando i profili dei potenziali elettori per confezionare messaggi elettorali personalizzati». Anche L. Califano, *Brevi riflessioni su privacy e costituzionalismo al tempo dei big data*, in *Federalismi.it*, 2017, 9, 5 evidenzia che l'elettore «viene sempre più assimilato - almeno dagli attori del circuito politico-rappresentativo – in tutto e per tutto a un consumatore di cui anticipare gusti, preferenze e bisogni, con inevitabili conseguenze tanto sul concetto di cittadinanza quanto su quello di partecipazione politica». In merito agli effetti della profilazione per fini politici sulla formazione dell'opinione pubblica, v. P. Villaschi, *Profilazione online e manipolazione del consenso nella bubble democracy*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 262 ss.; L. Pasqui, *La costituzione economica tra opinione pubblica e bubble democracy*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 773 ss.; M. Betzu - G. Demuro, *Big data e i rischi per la democrazia rappresentativa*, in questa *Rivista*, 1, 2020, 222 ss.; G. D'ippolito, *Comunicazione politica online: dal messaggio politico sponsorizzato alla sponsorizzazione sui social network*, in questa *Rivista*, 2020, 1, 166, che parla efficacemente di «messaggio politico commercializzato»; D. Servetti, *Social network, deliberazione pubblica e legislazione elettorale di contorno*, in questa *Rivista*, 1, 2020, 194 ss.; L. Califano, *Autodeterminazione vs. eterodeterminazione dell'elettore: voto, privacy e social network*, in *Federalismi.it*, 16, 2019; B. Caravita, *Social network, formazione del consenso, istituzioni politiche: quale regolamentazione possibile?*, *in Federalismi.it,* 2, 2019. Con specifico riguardo al caso *Cambridge Analytica*; E. Assante, *Cosa ci può insegnare il caso Cambridge Analytica*, in *Federalismi.it*, 9, 2018.

44 Il binomio "sovranità del consumatore" – "sovranità del cittadino" è delineato da G. Pitruzzella, *La libertà di informazione nell'era di Internet* cit. 68; in tal senso, anche A. Ciancio, *Il pluralismo alla prova dei nuovi mezzi di comunicazione* cit., 37, che richiama le interessanti considerazioni di C. Sunstein, *#republic. La democrazia nell'epoca dei social media*, Bologna, 2017, 205 ss.

della sovranità popolare»[45].

In altre parole, la personalizzazione dell'informazione mediante l'elaborazione algoritmica dei dati ceduti dagli utenti è suscettibile di compromettere l'effettiva attuazione del principio democratico nella misura in cui finisce con l'impedire il confronto tra le diverse idee e opinioni presenti nella società, incidendo così sul "pluralismo interno" dell'informazione.

Per giunta, nella nuova economia delle piattaforme, anche il pluralismo "esterno" appare più che mai vulnerabile, stante la concentrazione del potere di mercato[46] in capo ai cd. *Over the top* (su tutti, Google, Facebook, Amazon), foriera di implicazioni[47] anche in punto di diritto antitrust[48].

Ed infatti, sebbene negli utenti non paia esservi sufficiente consapevolezza di ciò, in letteratura si va sempre più evidenziando come la Rete risulti affetta da una «congenita ambiguità»: «da un lato, c'è il massimo del decentramento[49] e di apertura della produzione di informazioni, ma, dall'altro, c'è una forte spinta alla concentrazione dei servizi che rendono effettivamente disponibile e utilizzabile questa informazione nelle mani di poche compagnie multinazionali»[50].

Rinviando l'esame dei profili connessi ai "poteri" delle piattaforme, non si può non rilevare come la personalizzazione dei contenuti informativi e la conseguente polarizzazione delle opinioni rappresentino il terreno di fioritura delle cd. *fake news*, che costituiscono il secondo fenomeno oggetto di analisi e sulle cui possibili derive anti-democratiche si è vivacemente dibattuto[51].

---

[45] C. Casonato, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal. Rivista di BioDiritto*, 2, 2019, 715.

[46] P. Costa, *Motori di ricerca e social media*, cit., 255.

[47] C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni,* in *Diritto pubblico comparato ed europeo*, maggio 2019, 109: «non può essere trascurato il rischio legato al fatto che la citata enorme quantità di dati raccolti sia gestita da un numero ridottissimo di società. Possono quasi contarsi sulle dita di una mano, infatti, le imprese che concentrano la gestione complessiva di tale mole di informazioni. Su queste basi, appare con evidenza l'emersione di un nuovo potere immenso, le cui caratteristiche lo rendono particolarmente sfuggente rispetto alle tradizionali forme di controllo e limitazione; un potere accresciuto dalla possibilità di trattare, proprio attraverso l'AI, una massa di dati incalcolabile e altrimenti concretamente ingestibile».

[48] In merito al rapporto tra diritto antitrust e pluralismo informativo v. G. Pitruzzella, *La libertà di informazione nell'era di Internet,* cit., 56; V. Pampanin, *Tutela del pluralismo informativo*, cit., 173; M. Cuniberti, *Costituzione e mezzi di comunicazione*, cit., 246; F. Donati, *Il principio del pluralismo delle fonti informative*, cit., 664-666. In merito ai problemi di diritto antitrust posti dalle piattaforme v., di recente, A. Licastro, *Platform economy: sei proposte di legge antitrust per fermare lo strapotere dei Big Tech*, in *orizzontididirittopubblico.com*.

[49] Anche G. Pitruzzella, *La libertà di informazione nell'era di Internet*, cit., 57, nell'individuare i mutamenti dell'informazione determinati dall'innovazione tecnologica, rileva come il sistema di produzione dell'informazione si sia «radicalmente decentralizzato», in quanto «chiunque può produrre informazioni nella rete, reagire all'informazione immessa da altri, proporre fatti, idee, critiche, nuovi punti di vista, foto e video. […] Siamo entrati pienamente in una nuova era dell'informazione che Yochai Benkler ha definito la network information economy».

[50] G. Pitruzzella, *La libertà di informazione nell'era di Internet*, cit., 60.

[51] Si vedano i numerosi contributi contenuti in *Federalismi.it* -24 aprile 2020, nonché C. Magnani, *Libertà d'informazione online e fake news: vera emergenza? Appunti sul contrasto alla disinformazione tra legislatori statali e politiche europee*, in *forumcostituzinoale.it*; V. Baldini, *Verità e libertà nell' espressione del pensiero... Prendendo spunto da casi concreti…*, in *dirittifondamentali.it*, 2, 2017; I. Spadaro, *Contrasto alle fake news e tutela della democrazia*, in *dirittofondamentali.it*, 1, 2019; A. Sciortino, *Fake news e post-verità nella società dell'algoritmo*, in *dirittifondamentali.*

Invero, sono ormai numerosi gli studi in cui si evidenziano i nessi intercorrenti tra disinformazione e "manipolazioni algoritmiche", il che – come si vedrà - spiega anche la scelta di rivolgere l'attenzione a siffatto profilo in sede di regolazione europea.
Se la diffusione di false informazioni rappresenta un fenomeno non certo nuovo, ciò che l'ha reso meritevole di attenzione e regolazione è il suo esponenziale sviluppo – sia in termini quantitativi che qualitativi – conseguente all'evoluzione tecnologica[52]. Si vedrà, infatti, come l'utilizzo di algoritmi sempre più sofisticati, nonché dell'IA generativa, abbiano incrementato non solo le modalità di diffusione – sempre maggiore - di *fake news*, ma anche le tecniche di produzione.
Tuttavia, la rilevanza del fenomeno non si esaurisce in questo.
La necessità di un intervento regolatorio è sorta in considerazione delle implicazioni suscettibili di derivarne, consistenti – si è detto - in un «pregiudizio pubblico»[53].
Ed infatti, mutuando la definizione proposta dall'*High level group on fake news and online disinformation* istituito dalla Commissione europea nel 2018 e poi accolta dal Codice per il contrasto alla disinformazione del 2018, la "disinformazione" è costituita da «informazioni false, inesatte o fuorvianti progettate, presentate e diffuse a scopo di lucro o per ingannare intenzionalmente il pubblico[54], e che possono arrecare un pregiudizio pubblico». Sicché, come chiarito dalla dottrina che ha concorso all'elaborazione della definizione, «la nozione di disinformazione, così intesa, include non tanto quelle condotte che l'ordinamento riconosce come inerentemente illecite (si pensi, per esempio,

it, 2, 2021; M. Bassini-G. E. Vigevani, *Primi appunti su fake news e dintorni*, in questa *Rivista*, 1, 2017; O. Pollicino, *Fake news, Internet and Metaphors* (*to be handled carefully*), ivi*,* 1, 2017; M. Cuniberti, *Il contrasto alla disinformazione in rete tra logiche del mercato e (vecchie e nuove) velleità di controllo,* ivi, 1, 2017; C. Pinelli, *"Postverità", verità e libertà di manifestazione del pensiero*, ivi, 1, 2017; F. Pizzetti, *Fake news e allarme sociale: responsabilità, non censura*, ivi, 1, 2017; C. Melzi d'Eril, *Fake news e responsabilità: paradigmi classici e tendenze incriminatrici*, ivi, 1, 2017; N. Zanon, *Fake news e diffusione dei social media: abbiamo bisogno di un'"Autorità Pubblica della Verità"?*, ivi, 1, 2018; M. Furno, *Bufale elettroniche, repressione penale e democrazia,* ivi, 1, 2018; V. Visco Comandini, *Le fake news sui social network: un'analisi economica*, ivi, 2, 2018; A. Mazziotti Di Celso, *Dal primo emendamento al bavaglio malese. Fake news, libertà di espressione e rovesciamento delle categorie politiche tradizionali*, ivi, 3, 2018; E. Lehner, *Fake news e democrazia*, ivi, 1, 2019; G. Marchetti, *Le fake news e il ruolo degli algoritmi*, ivi, 1, 2020; F. Sciacchitano*, Fake news e disinformazione online: misure internazionali*, ivi, 1, 2020. Tra i contributi recenti v. D. Vese, *Regulating fake news: the right to freedom of expression in the era of emergency*, in *P.A. Persona e Amministrazione*, 1, 2021; L. Del Corona, *I social media e la disinformazione scientifica: spunti per un cambiamento di rotta alla luce dell'esperienza statunitenze ed europea*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 473 ss.

[52] T.E. Frosini, *L'ordine giuridico del digitale*, in *CERIDAP*, 2, 2023, 57: «Questione differente è quella della disinformazione, che non è solo la notizia falsa ma più in generale un fenomeno degenerativo e disgregativo, che attenta alla libertà di informazione quale pilastro su cui si fonda il costituzionalismo. Certo, la disinformazione viaggia anche sulla televisione e sulla stampa ma assume forme più capziose e insidiose sulla rete, vuoi perchè non c'è nessun tipo di controllo, se non quello dell'utente che dovrebbe sapere distinguere ciò che è buono da ciò che è cattivo, vuoi perchè diventa "virale", potendo distribuirsi, in tempi rapidissimi, in numerosi siti internet in giro per il mondo, al punto da assumere una presunta ufficialità». In tal senso, v. anche C. Valditara, *Fake news: regolamentazione e rimedi*, in *Diritto dell'Informazione e dell'Informatica*, 2, 2021, 257 ss.

[53] Commissione europea, *A multi-dimensional approach to disinformation. Report of the independent High level group on fake news and online disinformation*, Bruxelles, 2018, 10.

[54] Ad esempio «per obiettivi politico-ideologici o di vantaggio economico e che mina la legittimità di un processo elettorale, rovina la reputazione di una grande società o crea un ambiente ostile boicottando il dibattito democratico, esacerbando la polarizzazione sociale per migliorare la propria immagine» (in termini, S. Sassi, *Disinformazione contro Costituzionalismo*, Napoli, 2021).

al caso della diffamazione o a quello della calunnia), ma, piuttosto, tutti quei casi ove la produzione e diffusione di un contenuto falso non integra, di per sé, una fattispecie illecita ma produce, comunque, un potenziale danno ai principi e valori democratici»[55].
Alla luce di questo, si è ulteriormente precisato come gli elementi costitutivi[56] della nozione di "disinformazione" - ovvero (i) la falsità verificabile (ii) la finalità perseguita (ottenere un vantaggio o causare un danno) (iii) l'attitudine a determinare un pregiudizio per la collettività (ad es. per la salute, per la sicurezza, per il corretto svolgimento di una competizione elettorale)[57] – valgano anche a distinguerla dal concetto di "misinformazione", che si caratterizza per l'assenza della volontà di diffondere il falso[58].
Come accennato, lo sviluppo tecnologico gioca un ruolo fondamentale sia nella produzione, che nella diffusione di disinformazione.
Quanto alle modalità di diffusione, si è già rilevato come le "bolle di filtraggio" determinate dall'utilizzo di sistemi di raccomandazione da parte delle piattaforme incrementino le possibilità di attecchimento delle *fake news*[59], atteso che, per effetto della *confirmation bias*, l'utente, isolato nella bolla ed esposto all'effetto amplificatore della *eco chamber*, tenderà ad accogliere come vere le informazioni coerenti con le proprie opinioni[60]. Per giunta, nella diffusione di disinformazione è frequente l'utilizzo di *social bot*, account falsi attraverso cui, in modo automatico o semiautomatico, vengono immesse e condivise in Rete *fake news*.
Per quanto concerne le modalità di produzione, studi recenti hanno rivolto l'attenzione ai risultati resi possibili dalle più sofisticate tecniche di AI, specie per gli effetti suscettibili di derivarne a danno delle dinamiche democratiche.
Si pensi, in modo paradigmatico, ai *deepfakes* o ai *large language models* (LLM): i primi, consistono in foto, video e audio creati grazie a software di intelligenza artificiale che utilizzano immagini e audio per modificare o riprodurre, in modo realistico, un volto,

[55] O. Pollicino-P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *Federalismi.it*, 12, 2024, 6.

[56] L. Lorello, *Il valore costituzionale della buona informazione*, in *Dirittifondamentali.it*, 3, 2022, 225.

[57] Sulla base di quest'elemento la dottrina ha distinto cinque categorie di *fake news*, ovvero le notizie false che: 1. ledono l'onore e la reputazione di una persona; 2. inoffensive se prese singolarmente, arrecano un danno se considerate nel loro complesso (ad es. nelle campagne elettorali per colpire l'avversario politico); 3. provocano un danno alla società (ad es. l'informazione falsamente scientifica che attribuisce ad una determinata terapia l'effetto di provocare gravi malattie); 4. generano un indebito vantaggio a uno o più soggetti determinati (ad es. quelle volte a mettere in buona luce soggetti determinati al fine di ottenere consenso per sé o per altri); 5. confutano infondatamente verità empiriche o scientifiche acquisite (ad es. cercano di confutare una verità scientifica o empirica ormai acquisita, facendo leva sull'ignoranza dei loro destinatari) (C. Valditara, *Fake news: regolamentazione* cit.).

[58] Può accadere che «fattispecie di disinformazione tendano a essere successivamente ricondivise e diffuse da soggetti differenti da quelli che le hanno originate: lo stesso contenuto, pertanto, rappresenta un caso di disinformazione con riferimento ai soggetti che lo hanno originato o disseminato nella consapevolezza della sua falsità e con la finalità di causare un danno o trarne beneficio, mentre costituisce un esempio di misinformazione con riferimento a quanti, appresa la notizia, abbiano contribuito a diffonderla nell'erronea convinzione della sua corrispondenza alla verità. Si tratta, come detto, di un aspetto tipico dell'ecosistema internet, ove i contenuti disinformativi hanno la capacità di propagarsi e assumere facilmente connotati virali» (in termini, O. Pollicino-P. Dunn, *Disinformazione* cit., 6).

[59] G. Marchetti, *Le fake news e il ruolo degli algoritmi*, in questa *Rivista*, 1, 2020, 31 ss.

[60] G. Pitruzzella, *La libertà di informazione*, cit., 10.

un corpo o una voce[61]; i *large language models*, afferenti anch'essi all'IA generativa, consentono di generare testi, scrivere articoli, rispondere a domande in modo automatico (ad es. ChatGPT), estraendo dal database le parole o frasi più coerenti con l'*input* ricevuto[62].

Mentre per questi ultimi il rischio più evidente risiede nella possibilità che l'*output* prodotto, pur presentato come autentico, sconti il difetto di veridicità, i *bias* e la discriminatorietà dei dati di addestramento[63], nel caso dei *deepfakes* a destare allarme è l'utilizzo volontario dei contenuti generati per l'inquinamento del dibattito politico, specie nell'odierno contesto geopolitico mondiale[64].-

Infine, il terzo fenomeno suscettibile di incidere sulla libertà di informazione si sostanzia in quella che viene icasticamente definita "censura privata"[65] e, pur intercettando anch'esso le problematiche sottese all'utilizzo dell'IA, investe maggiormente il piano dei limiti alla comprimibilità della libertà di espressione da parte delle piattaforme digitali.

Invero, con l'espressione "censura privata" la dottrina suole alludere alle distorsioni potenzialmente connesse all'attività di moderazione dei contenuti immessi in Rete (*content moderation*), per tale intendendosi l'attività di controllo cui, a seconda che si tratti di *social network* o di motori di ricerca e della *policy* adottata, possono fare seguito una

---

[61] Sul punto, v. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in questa *Rivista*, 1, 2023, 172. L'A. spiega che: «Le tecnologie di *deep learning* principalmente utilizzate per la creazione di *deepfakes* sono due. La prima è quella a cui i creatori di *deepfakes* ricorrono più spesso: essa prende il nome di *Generative Adversarial Networks* (meglio conosciuta con l'acronimo GAN), cioè il sistema con il quale è possibile realizzare le sostituzioni facciali. Il funzionamento della tecnologia GAN può essere riassunto nel modo che segue: un primo algoritmo è in grado di individuare i frammenti in cui i due soggetti (quello che prende il posto dell'originale e quello che viene sostituito con il volto altrui) hanno espressioni simili; a questo punto interviene un secondo algoritmo che svolge il successivo passaggio di posizionamento facciale, ovvero sovrappone concretamente i due volti in questione. Sostanzialmente, questi algoritmi di apprendimento automatico analizzano il materiale multimediale a disposizione e sono in grado di crearne uno altrettanto di qualità paragonabile. La seconda tecnologia prende invece il nome di *Autoencoders*: si tratta di un tipo di rete neurale che è in grado di estrarre informazioni sulle caratteristiche facciali apprese da immagini e utilizzarle per crearne delle altre con espressioni diverse».

[62] L'appartenenza degli LLM al campo dell'IA generativa è resa evidente dal fatto che «per intelligenza artificiale generativa si intende un campo dell'intelligenza artificiale (AI) che si concentra sulla creazione di sistemi in grado di generare dati, contenuti o *output* in modo automatizzato, spesso utilizzando tecniche basate su reti neurali profonde (*deep learning*), apprendimento automatico (*machine learning*) e modelli probabilistici in grado di raccogliere una conoscenza molto ampia, ricavandola da enormi quantità di dati, principalmente dal Web, e di produrre testi, immagini, suoni o video» (G. Fasano, *Le 'informazioni sintetizzate' generate dai large language models e le esigenze di tutela del diritto all'informazione: valori costituzionali e nuove regole*, in *Dirittifondamentali.it*, 1, 2024, 108).

[63] Ivi, 109.

[64] Molteplici esempi, particolarmente significativi, sono riportati da O. Pollicino-P. Dunn, *Disinformazione* cit., 11.

[65] M. Monti, *Privatizzazione della censura e Internet platforms: la libertà d'espressione e i nuovi censori dell'agorà digitale*, *in Rivista italiana di informatica e diritto*, 1, 2019, 35 ss. La letteratura sull'argomento è ampia. V., *ex multis*, M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati". Spunti di comparazione*, in *Rivista italiana di informatica e diritto*, 2, 2021, 45; G.L. Conti, *Manifestazione del pensiero attraverso la rete e trasformazione della libertà di espressione: c'è ancora da ballare per strada?*, in *Rivista AIC*, 4, 2018; R. Niro, *Piattaforme digitali e libertà di espressione fra autoregolamentazione e coregolamentazione: note ricostruttive*, in *Osservatorio sulle fonti*, 3, 2021.

pluralità di misure: la rimozione dei contenuti, la sospensione o disattivazione dell'account, la chiusura delle pagine che violano le condizioni d'uso del servizio, la penalizzazione nell'indicizzazione dei siti ritenuti inaffidabili[66].

Sul piano delle modalità con cui l'attività di moderazione viene svolta, è possibile distinguere quella operante *ex ante* e quella attivata *ex post*.

La prima si realizza attraverso sistemi di filtraggio basati su algoritmi capaci di rintracciare i contenuti vietati (perché illeciti o perché contrari ai termini di servizio della piattaforma).

La seconda può essere realizzata direttamente dalla piattaforma, solitamente per mezzo di algoritmi, o per effetto della segnalazione degli utenti, cui segue l'esame dei contenuti da parte dei moderatori[67].

Sotto il profilo dei contenuti filtrabili e/o rimuovibili, si distinguono quelli illeciti, la cui espunzione consegue all'applicazione di specifiche previsioni normative (ad es. l'oscuramento dei siti terroristici ai sensi della L. n. 43/2015) o all'esecuzione di provvedimenti giurisdizionali (ad es. la rimozione di post diffamatori)[68], e quelli ritenuti contrari ai termini di servizio della piattaforma[69], ad esempio perché disinformativi o perché integranti *hate speech*[70] pur senza costituire un illecito penalmente rilevante.

Come è intuibile, è soprattutto quest'ultima ipotesi a destare maggiori perplessità sul piano della compatibilità con l'art. 21 Cost[71].

---

[66] Per un'approfondita analisi, v. M. Monti, *La disinformazione online, la crisi del rapporto pubblico-esperti e il rischio della privatizzazione della censura nelle azioni dell'Unione Europea (Code of practice on disinformation*), in *Federalismi.it*, 11, 2020, 297.

[67] Con particolare riferimento a Facebook, v. M. Betzu, *Libertà di espressione e poteri privati nel cyberspazio*, in *Diritto Costituzionale*, 1, 2020, 122 ss.

[68] M. Monti, *Privatizzazione della censura* cit., 39-40, che distingue la «censura *de facto* autonomamente predisposta dai giganti della Rete» e «la censura *de jure* che conduce gli Stati ad avvalersi delle piattaforme digitali per regolamentare il discorso pubblico online in ossequio alle normative vigenti sui rispettivi territori nazionali», nonché, nell'ambito di quest'ultima, «la censura privata funzionale, ossia quella imposta dallo Stato a seguito di un controllo di natura giudiziale/amministrativo, e la censura privata sostanziale, ossia quella in cui il bilanciamento fra libertà di espressione e altri beni giuridici viene delegato dallo Stato direttamente alle piattaforme digitali».

[69] Con riguardo alle diverse *policies* in materia di contenuti politico-elettorali, v. P. Bonini, *L'autoregolamentazione dei principali Social Network. Una prima ricognizione delle regole sui contenuti politici*, in *Federalismi.it*, 11, 2020, 269 ss.

[70] Per un approfondimento, v. C. Confortini, *Diffamazione e discorso d'odio in internet* in *Persona e Mercato*, 4, 2023, 693; I. Anrò, *Online hate speech: la prospettiva dell'Unione europea tra regolamentazione della condotta dei prestatori di servizi intermediari e ricorso al diritto penale*, in *Osservatorio sulle fonti*, 1, 2023, 13; P. Dunn, *Carattere eccezionale dell'"hate speech" e nuove forme di responsabilità per contenuti di terzi nella giurisprudenza EDU. Nota a C.edu, Sanchez c. Francia, 15 maggio 2023* in *Osservatorio costituzionale*, 6, 2023, 238 ss.; id; *Moderazione automatizzata e discriminazione algoritmica: il caso dell'"hate speech"* in *Rivista italiana di informatica e diritto*, 1, 2022, 2 ss.; L. Califano, La libertà di manifestazione del pensiero ... in rete; nuove frontiere di esercizio di un diritto antico. "Fake news", "hate speech" e profili di responsabilità dei "social network" in *Federalismi.it*, 26, 2021, 1 ss.; M. Castellaneta, *Responsabilità del politico per commenti altrui su Facebook: conforme alla Convenzione europea la "tolleranza zero" nei casi di messaggi d'odio* in *Medialas. Rivista di diritto dei media*, 3, 2021, 211 ss.

[71] Secondo una diversa prospettiva, «tale sub-categoria farebbe insorgere limitate problematiche sotto il profilo costituzionalistico, trattandosi di un'attività di *content moderation* che le *Internet platforms* realizzano in assenza di indicazioni provenienti dall'autorità statuale. Il soggetto, infatti, darebbe esclusiva applicazione alle summenzionate regole vigenti all'interno della propria *community*, qualificabile alla stregua di un ordinamento privatistico. All'interno di quest'ultimo, dunque, il gestore attuerebbe

Ed invero, nel secondo caso, il bilanciamento della libertà di manifestazione del pensiero[72] - e della libertà di essere informati con gli altri diritti fondamentali potenzialmente coinvolti (ad es. il diritto ad una informazione corretta, o quanto meno non deliberatamente falsa; la reputazione; la salute pubblica), peraltro suscettibile di incidere sulla dignità della persona e sul principio di eguaglianza quale eguaglianza di trattamento degli utenti, è svolta da una soggetto privato in virtù del rapporto contrattuale intercorrente con l'utente del servizio, talvolta – ed è un aspetto non privo di significato nella riflessione sul ruolo assunto dalle piattaforme - anche su "delega" dello stesso legislatore, che attribuisce alla piattaforma il compito di svolgere valutazioni discrezionali in ordine alla ammissibilità dei contenuti[73]. Basti pensare al *Code of conduct on countering illegal hate speech online* (Codice di condotta per lottare contro la diffusione dell'incitamento all'odio online), con cui si è affidato alle piattaforme il compito di rimuovere i contenuti qualificabili coma atti di incitamento all'odio (*hate speech*)[74], o, ancora al Codice di condotta contro la disinformazione adottato nel 2018, poi modificato e rafforzato nel 2022 (*Strengthened Code of Practice on Disinformation*), con cui si è riconosciuto alle piattaforme digitali un ruolo centrale nel contrasto alla disinformazione[75].

Tralasciando per il momento l'analisi delle scelte regolatorie europee, nell'indagare il quadro dei rischi sottesi al fenomeno in discussione emergono le numerose criticità poste in evidenza dalla dottrina: (i) la selezione dei contenuti ammessi, specie se di rilievo politico, potrebbe essere condizionata da scelte ideologiche dell'impresa o da valutazioni di tipo economico[76]; (ii) la piattaforma potrebbe tendere ad accogliere tutte le richieste di rimozione avanzate dagli utenti per non incorrere nel rischio di successive

una censura "interna", cioè svolta in assenza di una previsione legislativa statuale e/o una decisione dell'autorità giudiziaria alla base che qualifichi il contenuto come illecito» (in termini, G. Vasino, *Censura "privata" e contrasto all'hate speech nell'era delle Internet Platforms*, in *Federalismi.it*, 4, 2023, 135).

72 Sulla possibilità di ricondurre l'*hate speech* e le *fake news* nell'ambito oggettivo dell'art. 21 Cost., v. M. Orofino, *Art. 21 Cost.: le ragioni per un intervento di manutenzione ordinaria*, cit., 87, il quale rileva come la questione si collochi e rianimi l'antico dibattito, su cui anche la giurisprudenza costituzionale ha preso posizione, relativo alla possibilità che le menzogne e le dichiarazioni offensive trovino copertura nell'art. 21 Cost.

73 Secondo G. Vasino, *Censura "privata"* cit., 136, si tratta dell'ipotesi più problematica, in quanto «un operatore economico, divent[a] titolare di un considerevole potere decisorio nonché del compito di operare delicati bilanciamenti fra interessi contrapposti».

74 Per un dettagliato approfondimento, v. G. Vasino, *Censura "privata",* cit., 136 ss.

75 M. Monti, *La disinformazione online* cit., 297 rileva come «questo trend di privatizzazione della censura fosse stato inaugurato dalla giurisprudenza della Corte di giustizia in relazione al diritto all'oblio. La sentenza *Google Spain* affidava infatti alle piattaforme il compito di vagliare e compiere il bilanciamento fra diritto all'informazione e diritto alla privacy, rimuovendo dai motori di ricerca i contenuti ritenuti non di interesse pubblico». Evidenzia, inoltre, come con il Codice si proponga «*alle Internet platforms* firmatarie, fra cui Google e Facebook, un'auto-regolamentazione eterodiretta che finisce per "responsabilizzare" questi soggetti mediali in relazione al paradigma della libertà di informazione. Questa forma di "responsabilizzazione" finisce però per assegnare alle *Internet platforms* un ruolo "paracostituzionale" in assenza di forme di regolamentazione, di cornici normative o di controllo da parte di soggetti pubblici (giudici o *authorities*), aprendo così la strada ad una «privatizzazione della censura», già avviata in questo settore e in altri campi limitrofi».

76 M. Monti, *Privatizzazione della censura* cit., 37.

contestazioni[77]; (iii) l'utilizzo dell'IA per la moderazione dei contenuti[78] porta con sé il rischio ormai noto dei cosiddetti *technical bias*, cosicché potrebbe derivarne «un effetto censorio di gran lunga superiore per alcune minoranze, dando luogo ad un secondario effetto discriminatorio a base razziale o fondato sull'orientamento sessuale»[79]; (iv) le forme di controllo privato sulle misure adottate dalle piattaforme – paradigmatico il caso dell'*Oversight Board* di Facebook[80] - costituiscono una tutela più apparente che reale per la libertà di espressione e per i suoi riflessi sul piano dell'informazione, atteso che anche le decisione assunte dal "controllore" risultano, in buona sostanza, condizionate dai parametri valutativi dalla piattaforma (il "controllato"), anziché da quelli desumibili dall'ordinamento giuridico[81].

Le suddette criticità, oltre a condizionare la libertà di espressione, hanno evidenti riflessi sul pluralismo informativo, nonché, potenzialmente, sul corretto funzionamento delle dinamiche democratiche; aspetto, quest'ultimo, rilevante non solo – lo si è visto - per l'incidenza della «buona informazione»[82] sull'attuazione del principio democratico, ma anche, più in generale, per l'entità assunta dai nuovi «poteri privati»[83], espressione con cui suole riferirsi ai «soggetti che agiscono nelle forme del diritto privato, ma che, per la loro posizione di forza economica e/o sociale, sono capaci di incidere sull'eser-

---

77 Ivi, 41.

78 Sulle modalità di utilizzo, v. G. Marchetti, *Le fake news e il ruolo degli Algoritmi*, in questa *Rivista*, 1, 2020, 29 ss.; O. Pollicino-P Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *Federalismi.it*, 12, 2024, 14.

79 G. Vasino, *Censura "privata"* cit., 146.

80 A. Gerosa, *La tutela della libertà di manifestazione del pensiero nella rete tra Independent Oversight Board e ruolo dei pubblici poteri. Commenti a margine della decisione n. 2021-001-FB-FBR,* in *Forum di Quaderni Costituzionali*, 2, 2021, 427 ss. In generale, v. A. Iannotti della Valle, *A Facebook court is born: towards the jurisdiction of the future?* in *European Journal of Privacy Law & Technologies*, 1, 2020, 9 ss.

81 M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *La Rivista Gruppo di Pisa*, 2, 2021, 176: «Le prime analisi delle poche decisioni sinora rese dal Comitato mostrano come il parametro utilizzato non sia mai l'ordinamento costituzionale dello Stato in cui è stata posta in essere la condotta, "ma solo standard giuridici internazionali, quali quelli delineati dalla Convenzione internazionale sui Diritti Civili e Politici o dalle varie raccomandazioni del Consiglio dell'ONU per i Diritti Umani o dal Relatore speciale per la promozione e protezione delle libertà di opinione ed espressione", che vengono utilizzati per operare in concreto il bilanciamento tra la libertà di espressione e gli altri valori di riferimento della piattaforma, come dignità e sicurezza. Anche nel famoso caso Trump, deciso il 5 maggio 2021, il Comitato individua quali parametri di riferimento sia la Convenzione internazionale sui diritti civili e politici, sia i Community Standards elaborati da Facebook, ponendoli sullo stesso piano. Al di là della correttezza o meno della decisione, che ha ritenuto legittima la sospensione dell'account di Trump, ma non appropriata se applicata senza la previsione di un termine finale, l'aspetto più significativo per il costituzionalista non può che essere la commistione tra i valori individuati da Facebook e gli standard internazionali dei diritti umani […]. Ed ecco che la vittima di questa operazione sono proprio le Costituzioni, rimpiazzate da un'autorità privata che si autolegittima».

82 L. Lorello, *Il valore costituzionale della buona informazione*, cit., 207 ss.

83 Sull'argomento, v. C. Pinelli, *Il costituzionalismo di fronte ai nuovi poteri privati*, in *Economia pubblica*, 1, 2022, 122 ss.; G. Malgieri-A. Davola, *Data powerful. Un'indagine sulla nozione di potere e il suo rapporto con la vulnerabilità nel mercato digitale*, in *Concorrenza e mercato*, 1, 2022, 67 ss.; L. Torchia, *Poteri pubblici e privati nel mondo digitale*, in *Il Mulino*, 1, 2024, 14 ss.; V. Cavani, *Nuovi poteri vecchi problemi*, in *Diritto pubblico comparato ed europeo*, 1, 2023, 223 ss.; L. Ammannati, *I signori nell'era dell'algoritmo*, in *Diritto pubblico*, 2, 2021, 381 ss.; M. Betzu, *I poteri privati nella società digitale: oligopoli e antitrust*, in *Diritto pubblico*, 3, 2021, 739 ss.; G. Vettori, *Sui poteri privati. Interazioni e contaminazioni*, in *Diritto pubblico*, 3, 2022, 829 ss.

cizio delle libertà fondamentali dei singoli»[84].

Sicché, anche sotto tale profilo, emergono i profili di contrasto con l'art. 1 Cost., che, come è stato rilevato, sintetizza in sé i due pilastri dello Stato democratico-pluralista: «da un lato la sovranità popolare, ovvero la remissione delle decisioni politiche al popolo, a maggioranza, attraverso il principio del governo rappresentativo, […] che implica la libera formazione del consenso attraverso una dialettica democratica. Dall'altro il *rule of law* costituzionale, ovvero la limitazione del potere, anche quello delle maggioranze politiche democratiche, attraverso il diritto, in nome della garanzia del pluralismo che trova la sua espressione nei diritti fondamentali delle persone»[85].

Nello scenario sin qui descritto, entrambi risultano sottoposti a tensione, per effetto, da un lato, del condizionamento esercitato dalle Big Tech sul libero confronto delle idee attraverso le risorse tecnologiche[86], dall'altro, della "resistenza" alla regolazione pubblica[87], cui fa da contraltare la tendenza dei nuovi poteri privati all'autoregolazione e, così, l'incidenza sempre più pervasiva e incontrollata sulle libertà fondamentali[88].

Ed invero, proprio nelle riflessioni svolte dalla dottrina costituzionalistica sulla cd. "privatizzazione della censura", sembra di poter cogliere una diffusa preoccupazione per ciò che di più profondo tale fenomeno parrebbe rivelare, ovvero il rischio che l'autoregolazione elevi soggetti privati «al rango di soggetti costituzionali, protagonisti

---

[84] O. Grandinetti, *Le piattaforme digitali* cit., 179. F. Paruzzo, *I sovrani della rete*, Torino, 2022, 110: «Le piattaforme digitali finiscono per manifestare un potere di fatto che estende la propria azione ben oltre la sfera del mercato e che incide su dinamiche connesse alla sfera pubblica e ai diritti fondamentali»; A. Pisaneschi, *Reti sociali ed elezioni: il fallimento del mercato e il quadro regolatorio europeo*, in *Liber amicorum per Pasquale Costanzo*, Consulta online, 3 febbraio 2020, 3: «Il rischio – o finanche ormai la certezza – è che in nome della libertà di informazione si sia creato un mercato oligopolistico privo di regole, con la presenza di un numero limitato di soggetti che condiziona il dibattito politico e l'accesso all'informazione».

[85] T. Groppi, *Alle frontiere dello Stato Costituzionale, innovazione tecnologica e intelligenza Artificiale*, in *Consulta Online*, 3, 2020, 678. A. Iannuzzi, *Le fonti del diritto dell'Unione Europea per la disciplina della società digitale*, in F. Pizzetti (a cura di), *La regolazione europea della società digitale*, Torino, 2024, 15 evidenzia un duplice ordine di aspetti: innanzitutto, il tema della sovranità e del potere nell'era digitale rappresenta uno dei problemi più rilevanti del costituzionalismo del XXI secolo; in secondo luogo, è in atto un processo parallelo ed ulteriore rispetto all'affermazione del nuovo "potere sovrano" di carattere tecnologico, ovvero il declino della sovranità degli Stati-nazione.

[86] A tal proposito, parla di "sovranità digitale" A. Simoncini, *Sovranità e potere nell'era digitale*, in T. E. Frosini-O. Pollicino-E. Apa-M. Bassini (a cura di), *Diritti e libertà in internet*, Firenze, 2017, 25, che ne individua come caratteristiche fondanti: la potenza di calcolo e l'automazione. Tra i contributi recenti, L. Durst, *Diritti e predittività. Spunti introduttivi in tema di sistemi decisori automatizzati*, in F. Fabrizzi-L. Durst (a cura di), *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era digitale*, Napoli, 2024, 17 evidenzia come «la natura prevalentemente privata dei soggetti produttori e finanziatori delle nuove tecnologie dell'informazione in senso lato presenta infatti come contropartita, a fronte delle opportunità di sviluppo della conoscenza e del mercato digitale, il rischio di trasformare un fenomeno, originariamente di tipo meramente tecnologico ed economico, in una nuova forma di potere, capace di incidere significativamente nelle scelte delle persone, non solo grazie alla possibilità di conoscerne comportamenti e desideri, ma anche di influenzare decisioni e azioni, tanto nella sfera privata quanto pubblica, senza che gli utenti ne abbiano la dovuta consapevolezza».

[87] T. Groppi, *Alle frontiere dello Stato Costituzionale, innovazione tecnologica e intelligenza Artificiale*, cit., 679-680 evidenzia la "crisi della *rule of law*" per effetto dell'emersione dei nuovi poteri globali.

[88] Come notato da M. Bassini, *Internet e libertà di espressione*, cit., 76: «Se la comparsa sulla scena di poteri, o quantomeno di attori privati non pare potersi dire appannaggio esclusivo della rete, ma un più generale e diffuso portato della globalizzazione, ciò che rende particolarmente meritevole di attenzione l'ascesa di questi operatori nel terreno specifico di Internet è la loro capacità di incidere in misura dirimente sul livello di tutela di alcune libertà fondamentali, *in primis*, la libertà di manifestazione del pensiero».

indiscussi di "costituzioni sociali" dalla dubbia natura».[89]

Si tratta di aspetti emersi in modo assai chiaro nei casi balzati alle cronache d'oltreoceano (ad es. nel caso Facebook/Trump)[90] e nazionali (ad es. Facebook/Casa Pound e Facebook-Forza Nuova), su cui si avrà modo di tornare nelle pagine seguenti.

Per ora basti osservare come le note vicende giurisdizionali in cui questi ultimi sono sfociati, connotati dall' «alta tensione assiologica […] tra plurimi beni di rango costituzionale»[91], se, da una parte, hanno fatto emergere gli anticorpi del nostro ordinamento rispetto alle possibili distorsioni innanzi paventate[92], al contempo lasciano intravedere «questioni più ampie», tra cui quella «del ruolo giocato dalle piattaforme on-line, dei limiti che queste incontrano nella conformazione dei diritti fondamentali o, viceversa, della problematica configurabilità, in capo ad esse, di una funzione di custodia dei contenuti della democrazia costituzionale»[93], questioni per la cui soluzione – come si vedrà - la garanzia della tutela giurisdizionale dei singoli soggetti lesi, pur imprescindibile, non può bastare[94].

Ed infatti, rinviando sul punto a quanto verrà detto trattando del quadro regolatorio europeo, emerge sin d'ora la condivisibilità dei rilievi critici proposti dalla dottrina, volti in modo pressoché unanime ad evidenziare l'irrinunciabilità di una regolazione pubblica[95]; regolazione che, a seconda delle impostazioni prospettate, potrebbe assumere diversi gradi di intensità e intervenire *ex ante* - con un controllo preventivo (pubblico) sui contenuti disponibili in Rete[96] o limitando l'attività di moderazione delle

---

89 M. Betzu, *Libertà di espressione e poteri privati nel cyberspazio*, in *Diritto Costituzionale*, 2020, 1, 122 ss. Sui rischi connessi a forme di *enforcement* dei diritti fondamentali basate sul ruolo dei prestatori di servizi, v. M. Bassini, *Internet e libertà di espressione* cit., 405 ss.

90 M. Manetti, *Facebook, Trump e la fedeltà alla Costituzione*, in *Forum di Quaderni Costituzionali*, 1, 2021; O. Pollicino-G. De Gregorio-M. Bassini, *Trump's Indefinite Ban: Shifting the Facebook Oversight Board away from the First Amendment Doctrine*, in *VerfBlog*, 5 novembre 2021, in verfassungsblog.de/fob-trump-2/.

91 C. Caruso, *I custodi di silicio. Protezione della democrazia e libertà di espressione nell'era dei social network*, in *Liber amicorum per Pasquale Costanzo*, 17 aprile 2020, 3 (in *Consulta online*).

92 In dottrina, v. A. Venanzoni, *Pluralismo politico e valore di spazio di dibattito pubblico della piattaforma social Facebook: la vicenda CasaPound*, in *Diritto di Internet*, 12 dicembre 2019; I. M. Lo Presti, *CasaPound, Forza Nuova e facebook. Considerazioni a margine delle recenti ordinanze cautelari e questioni aperte circa la relazione tra partiti politici e social network*, in *Forum di Quaderni costituzionali,* 2, 2020, 924 ss.; B. Mazzolai, *La censura su piattaforme digitali private: il caso "Casa Pound c. Facebook"*, in *Il Diritto dell'informazione e dell'informatica*, 36(1), 2020, 109 ss.; S. Piva, *Facebook è un servizio pubblico? La controversia su CasaPound risolleva la quaestio dell'inquadramento giuridico dei social network*, in *Dirittifondamentali.it*, 2, 2020.

93 C. Caruso, *I custodi di silicio* cit., 3-4.

94 A. Gerosa, *La tutela della libertà di manifestazione del pensiero* cit., 439 rileva, in modo condivisibile, come la strada del processo civile possa risultare eccessivamente lunga e costosa per l'utente medio e così indurlo a rinunciare alla tutela. V. anche Conte, *Disinformazione digitale, fiduciarietà informativa, rimedi contrattuali*, in *Annali SISDC*, 10, 2023, 26, che pone l'accento sull'importanza della tutela privatistica in forma collettiva e richiama, in questo senso, A. Gentili, *Fine del diritto all'informazione?*, in M. D'Auria (a cura di), *I problemi dell'informazione nel diritto civile, oggi. Studi in onore di Vincenzo Cuffaro*, Roma, 2022, 66, che sottolinea l'utilità dell'azione *ex* artt. 139 ss. del codice del consumo.

95 E. Bruti Liberati, *Poteri privati e nuova regolazione pubblica*, in *Diritto pubblico*, 1, 2023, 285 ss.

96 G. Pitruzzella, *La libertà di informazione nell'era di Internet,* in questa *Rivista*, 1, 2018, 19 ss.; O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione nell'era di Internet*, ivi, 1, 2018, 48 ss., G. De Gregorio, *The marketplace of ideas nell'era della post-verità: quali responsabilità per gli attori pubblici e privati online?*, ivi, 1, 2017, 91 ss., M. Monti, *Fake news e social network: la verità ai tempi di Facebook*, ivi, 1, 2017, 79 ss.

piattaforme ai soli contenuti contrari all'ordinamento giuridico[97] - o *ex post*, mediante l'introduzione di un controllo pubblico – da rimettere, secondo parte della dottrina, ad un'autorità amministrativa appositamente costituita - sulle misure limitative adottate dalle piattaforme[98].

3. Dall'auto-regolazione alla co-regolazione europea

Il fenomeno da ultimo descritto, invero, si è sviluppato per effetto della scelta ordinamentale - negli USA[99] e, in un primo tempo, anche in Europa - di optare per la *self-regulation* dei prestatori di servizi digitali.

Tale modello di regolazione, che costituisce esercizio di autonomia privata, viene in rilievo ogniqualvolta le regole di comportamento (codici di condotta, *standard* di produzione, regole interne di funzionamento delle piattaforme) siano definite da soggetti privati; segnatamente, «ricorre quando un gruppo di soggetti e/o una formazione sociale esponenziale di tale gruppo fissano autonomamente regole che li riguardano», solitamente attraverso la fissazione di *standard* che i privati si impegnano ad osservare[100].

In particolare, è possibile distinguere tre tipologie di *self-regulation*: la prima opera su base completamente volontaria e generalmente è priva di effetti giuridicamente vincolanti; la seconda – l'autoregolazione delegata – si ha quando la legge sollecita i privati all'adozione di regole, di cui si limita a definire i principi base, e non prevede conseguenze giuridiche in caso di violazione, rimettendo la soluzione di eventuali controversie agli organi interni (è il caso dei codici di condotta previsti dagli artt. 40 e 41 GDPR);

---

97 O. Grandinetti**, Le piattaforme digitali** *come. "poteri privati" e la censura online*, in *Rivista italiana di informatica e diritto*, 2, 2022, 181: «Sul piano costituzionale dovrebbe perciò escludersi che la censura privata delle piattaforme possa prevalere sulla libertà di manifestazione dei singoli salvo che essa non si risolva nel reprimere espressioni già considerate illecite dall'ordinamento giuridico (compresa ovviamente la Costituzione). È pur vero che spesso sono proprio gli Stati a "delegare" alle piattaforme quell'attività di censura che essi non effettuano per ragioni legate ad asserite difficoltà tecniche o economiche. Ma anche "uno Stato che miri ad essere minimo" dovrebbe considerare "irrinunziabile" il compito di decidere cosa possa essere pubblicamente diffuso in rete».

98 A. Gerosa, *La tutela della libertà* cit., 439, che richiama, sul punto, G. Pitruzzella, *Quel filtro necessario per le notizie false sul web*, in *Corriere della sera*, 2 gennaio 2017 e, in precedenza, G. Pitruzzella, *Italy antitrust chief urges EU to help beat fake news*, in *Financial Time*s, 30 dicembre 2016.

99 Per un approfondimento, v. R. Niro, *Piattaforme digitali e libertà di espressione fra autoregolamentazione e coregolazione: note ricostruttive*, in *Osservatorio sulle fonti*, 3, 2021, 1380 ss.

100 M. Ramajoli, *Self regulation*, *soft regulation* e *hard regulation nei mercati finanziari*, in *Rivista della Regolazione dei mercati*, 2, 2016, 54. L'A. rileva ulteriormente: «secondo il parere del Comitato economico e sociale europeo sul tema "Autoregolamentazione e coregolamentazione nel quadro legislativo dell'UE", del 4 settembre 2015, con il termine autoregolamentazione (o, meglio, autoregolazione), "si designa genericamente, quando ci si riferisce al comportamento economico, l'adozione da parte degli attori economici di certe regole di condotta nelle relazioni reciproche oppure nei confronti di terzi sul mercato e nella società, regole il cui rispetto è frutto di un accordo tra gli stessi attori, senza meccanismi coercitivi esterni" (punto 3.2). L'autoregolazione è adottata spesso con procedimenti non rispettosi del principio di trasparenza e di pubblicità e l'inottemperanza nei suoi riguardi non è giuridicamente sanzionata. La non sanzionabilità sul piano giuridico dei comportamenti devianti è però accompagnata da altre "misure di reazione" che possono essere prese dall'aggregazione privata autrice della *self regulation* nei confronti di coloro che, dopo avere aderito alla stessa, la violino. Queste misure sono essenzialmente sanzioni di tipo reputazionale oppure fattuale, come l'esclusione dalla lista degli aderenti all'aggregazione privata».

la terza – *enforced self-regulation* – si caratterizza per un triplice ordine di profili: (i) il ruolo rivestito dal regolatore pubblico, che verifica la congruità delle regole definite dalle imprese rispetto all'interesse pubblico perseguito, (ii) il controllo pubblico sull'indipendenza e sull'efficienza dell'organo di controllo interno all'impresa, (iii) la previsione legislativa di conseguenze giuridiche in caso di violazione delle regole definite dai privati e ratificate dal regolatore pubblico[101].

Per quanto più rileva ai fini del nostro discorso, l'auto-regolazione si realizza attraverso la definizione di *policies*[102] alla cui stregua le piattaforme individuano i contenuti contrari alle condizioni d'uso del servizio applicate agli utenti (ad esempio perché disinformativi o contenenti i cd. "discorsi d'odio"), per poi incidere o sulla visibilità di tali contenuti attraverso la loro retrocessione, demonetizzazione o rimozione, oppure, in modo ancor più pervasivo, sugli utenti che tali contenuti abbiano diffuso, attraverso la sospensione o rimozione del loro account.

Sulla base di quanto rilevato dalla dottrina che ampiamente si è occupata dell'argomento, nell'ordinamento statunitense[103] la convinta adesione al modello della *self-regulation* si è fondata sulla concezione di Internet quale «nuovo mercato delle idee»; sottende, cioè, la convinzione che la «capacità auto-correttiva del "mercato"» possa «fare emergere, attraverso una *free competition* di idee e opinioni (anche quelle false), la verità» senza la necessità di alcun intervento pubblico[104].

Viceversa, nell'ordinamento europeo si è realizzata una progressiva trasformazione del modello regolatorio, dapprima improntato al paradigma della *self-regulation*[105], ma

---

[101] F. Di Porto-N. Rangone, *Strategie regolatorie e qualità della regolazione*, in C. Cambini-A. Manganelli-G. Napolitano-A. Nicita (a cura di), *Economia e diritto della regolazione*, Bologna, 2024, 94-95.

[102] S. Piva, *Libertà di informazione e piattaforme digitali. Questioni aperte nei paesi liberal-democratici e considerazioni sulle "misure di guerra" nella Federazione russa*, in *Costituzionalismo.it*, 2, 2022, 200.

[103] Per l'analisi delle più rilevanti posizioni espresse dagli studiosi della regolazione di Internet nel contesto statunitense, v. M. Bassini, *Internet e libertà di espressione*, cit., 21 ss.

[104] O. Pollicino, *Asimmetrie valoriali transatlantiche tra self-regulation, hard law e co-regolazione (ovvero sul se e sul come regolamentare le strategie contro la disinformazione on line)*, in *Osservatorio sulle fonti,* 2, 2023, 231.

[105] Per l'analisi di siffatta evoluzione, v. S. Sassi, *L'Unione Europea e la lotta alla disinformazione online*, in *Federalismi.it*, 15, 2023, 189 ss., che distingue le misure "di prima generazione" (dal 2015) e quelle di "seconda generazione" (dal 2020); E. Longo, *Libertà di informazione e lotta alla disinformazione nel Digital Services Act*, in *Giornale di diritto amministrativo*, 6, 2023, 739. Tra gli atti più rilevanti, si rammentano: Risoluzione del Parlamento europeo del 15 giugno 2017 sulle piattaforme on-line e il mercato unico digitale (2016/2276(INI)), n. 36; Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee of the Regions, "Commission Work Program 2018. An agenda for more united, stronger and more democratic Europe", Strasbourg 24.10.2017, COM(2017)650 final; Relazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, sull'attuazione della comunicazione "Contrastare la disinformazione on-line: un approccio europeo", Bruxelles, 5.12.2018, COM(2018)794 *final*; Consiglio dell'Unione europea, Sforzi complementari per rafforzare la resilienza e contrastare le minacce ibride, Bruxelles, 10.12.2019, 14972/19; Comunicazioni della Commissione europea e dell'Alto rappresentante dell'Unione per gli affari esteri e la politica di sicurezza al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni, Tackling COVID-19 disinformation – Getting the facts right, Bruxelles, 10.6.2020, JOIN(2020)8 *final*; Comunicazione della Commissione europea al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, sul piano d'azione per la democrazia europea, Bruxelles, 3.12.2020, COM(2020) 790 *final*.

successivamente evolutosi[106] in quella che la dottrina unanimemente tende a qualificare come co-regolazione[107].

Ed invero, nel contesto europeo, l'inadeguatezza dell'auto-regolazione è emersa ben presto.

Fra le molteplici ragioni evidenziate – l'inidoneità ad assicurare un regime effettivo di responsabilità[108], l'insufficienza delle misure adottate dalle piattaforme (anche per la riluttanza ad adottare quelle riduttive del numero degli utenti), gli esiti talvolta irragionevoli prodotti dagli strumenti algoritmici di controllo[109] – ne è stata segnalata una in particolare: «il rischio che siano proprio le piattaforme, dopo aver invocato la libertà di espressione come giustificazione per l'assenza di eteroregolazione, ad assumere misure che possono incidere significativamente sulle libertà fondamentali, senza che siano previsti rimedi o correttivi idonei»[110].

Si è così palesata la difficoltà di importare dall'ordinamento statunitense il paradigma del *free marketplace of ideas*[111], difficoltà riconducibile essenzialmente alle diverse radici culturali e costituzionali dell'ordinamento europeo: il costituzionalismo americano è dominato dal Primo emendamento, dalla libertà di espressione e, conseguentemente, non può che favorire la fiducia nella capacità autocorrettiva del "mercato delle idee" e nello strumento tecnologico attraverso cui detto mercato si realizza; il costituzionalismo europeo, al contrario, attribuisce pari importanza alla libertà di espressione e agli

---

[106] Tale evoluzione si pone in linea con l'obiettivo di realizzare quella che è stata definita la "sovranità digitale" europea rispetto alle grandi aziende tecnologiche (v. Commissione europea, Discorso sullo Stato dell'Unione 2020. Costruiamo il mondo in cui vogliamo vivere: un'Unione vitale in un mondo fragile, 15; Commissione europea, Una strategia europea per i dati, doc. COM(2020) 66 final, 19 febbraio 2020, 5; Commissione europea, Bussola per il digitale 2030: il modello europeo per il decennio digitale, doc. COM(2021) 118 final, 9 marzo 2021, 1; Commissione europea, Tutela dei diritti fondamentali nell'era digitale – Relazione 2021 sull'applicazione della Carta dei diritti fondamentali dell'Unione europea, doc. COM(2021) 819 final, 10 dicembre 2021; Dichiarazione europea sui diritti e principi digitali per il decennio digitale proclamata dal Parlamento europeo, dal Consiglio e dalla Commissione europea in data 15 dicembre 2022, pubblicata in *GUUE* C 23, del 23 gennaio 2023, 1). Il fenomeno è noto come "*Brussels effect*" (v. A. Bendiek-I. Stuerzer, *The Brussels Effect, European Regulatory Power and Political Capital: Evidence for Mutually Reinforcing Internal and External Dimensions of the Brussels Effect from the European Digital Policy Debate*, in *Digital Society*, 1, 2023, 1 ss.). Sul punto, v. anche G. Fasano, *Le 'informazioni sintetizzate' generate dai large language models*, cit., 122-123.

[107] «Per co-regolamentazione (o, anche qui, per co-regolazione), secondo il Parere del Comitato economico e sociale europeo, s'intende "una forma di regolamentazione delle parti interessate (stakeholder) che è promossa, orientata, guidata o controllata da una terza parte (sia essa un organismo ufficiale o un'autorità di regolamentazione indipendente) di norma dotata di poteri di esame, di controllo e, in alcuni casi, sanzionatori" (punto 3.4). La *co-regulation* è altrimenti definibile come *audited self regulation* (autoregolazione monitorata), termine che mette maggiormente in evidenza la circostanza per cui se, da una parte, le regole di disciplina sono poste in essere da soggetti privati o da loro organismi associativi, dall'altra, esse sono assoggettate a un controllo indiretto da parte di un'autorità pubblica» (in termini, M. Ramajoli, *Self regulation*, *soft regulation* cit., 55).

[108] S. Del Gatto, *Il Digital Services Act: un'introduzione*, in *Giornale di diritto amministrativo*, 7, 2023, 727.

[109] L. Torchia, *I poteri di vigilanza, controllo e sanzionatori nella regolazione europea della trasformazione digitale*, in *Riv. Trim. dir. pubbl.*, 2022, 4, 1101 ss.; L. Ammanati, *Regolatori e supervisori nell' era digitale: ripensare la regolazione*, in *Giurisprudenza Costituzionale*, 3, 2023, 1453 ss.

[110] L. Torchia, *I poteri di vigilanza,* cit.

[111] G. De Gregorio, *Il diritto delle piattaforme digitali: un'analisi comparata dell'approccio statunitense ed europeo al governo della libertà di espressione*, in *DPCE Online*, 2022, Speciale, 1455 ss.

altri diritti fondamentali[112], tra cui – per quanto più rileva ai fini del discorso - quello ad una «buona informazione»[113].

Ed infatti, nella attuale regolazione europea sembra proprio di poter cogliere l'intento di un ragionevole bilanciamento tra il diritto ad una informazione di qualità (veritiera e plurale), in astratto tutelabile dalle piattaforme con l'auto-regolazione e l'attività di *content moderation*, e la libertà di informare, fondata nella libertà di espressione e imprescindibile garanzia del pluralismo informativo.

Siffatto bilanciamento è perseguito chiaramente dal Regolamento UE 2022/2065 (*Digital Services Act*)[114], che, pur attribuendo ai prestatori di servizi digitali un ruolo centrale nel contrasto alla disinformazione, pone taluni limiti all'attività di moderazione dei contenuti, garantendo – lo si vedrà meglio in seguito (v. par. 4) – la trasparenza del "procedimento decisionale" e forme di tutela anche extragiudiziaria avverso le misure limitative della libertà di espressione.

Invero, come rilevato in modo uniforme dalla dottrina, l'opzione accolta dal legislatore europeo si connota per le caratteristiche tipiche della co-regolazione, modello regolatorio consistente nella definizione delle regole da parte del regolatore pubblico in dialogo con gli operatori economici e preordinato ad un duplice ordine di obiettivi:

---

[112] O. Pollicino, *Asimmetrie valoriali transatlantiche tra self-regulation, hard law e co-regolazione (ovvero sul se e sul come regolamentare le strategie contro la disinformazione on line)*, in *Osservatorio sulle fonti,* 2, 2023, 231-232.

[113] L. Lorello, *Il valore costituzionale della buona informazione* cit., 207 ss.

[114] Il *Digital Services Act* si inserisce nell'ambito della Strategia europea per il mercato unico digitale insieme al Regolamento (UE) 2022/1925 del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (*Digital Markets Act* - DMA). Sotto il profilo funzionale, «obiettivo condiviso delle due discipline, in linea con la scelta dell'art. 114 TFUE come base giuridica, è quello di definire un quadro di regole armonizzato e uniforme in tutta l'Unione europea al fine di superare la frammentazione normativa esistente che ostacola il corretto funzionamento del mercato interno, rappresenta un vulnus per la certezza del diritto, genera costi per le imprese e riduce il benessere dei consumatori. Mentre il *Digital Markets Act* è diretto a prevenire le conseguenze negative del potere di mercato in mano alle grandi piattaforme digitali, i c.d. *gatekeepers*, il *Digital Services Act* disciplina obblighi e responsabilità dei fornitori dei servizi intermediari nel fornire l'accesso a beni, servizi o contenuti, inclusi i marketplace online, con l'obiettivo di contribuire al corretto funzionamento del mercato interno dei servizi intermediari attraverso l'introduzione di norme armonizzate che garantiscano un ambiente online sicuro, prevedibile e affidabile» (in termini, S. Del Gatto *Il Digital Services Act: un'introduzione*, in Giornale di diritto amministrativo, 2023, 6, 725). Sotto il profilo dell'ambito territoriale di applicazione, i regolamenti «hanno in comune, peraltro, un criterio territoriale di applicazione definito in termini particolarmente ampi, perché la nuova disciplina regolamentare si applica a qualsiasi soggetto che operi sul territorio dell'Unione, e a qualsiasi servizio reso o sistema di intelligenza artificiale commercializzato o utilizzato all'interno del mercato unico, indipendentemente dallo stabilimento del soggetto o dalla origine del servizio o dalla costruzione del prodotto fuori dall'Unione. Si è cercato così di affrontare l'evidente asimmetria fra i confini del regolatore, che comunque non vanno oltre i confini dell'Unione europea, e la dimensione globale e almeno in parte aterritoriale dei principali soggetti ai quali la regolazione è indirizzata» (L. Torchia, *I poteri di vigilanza* cit.). In dottrina, con riguardo al DSA, v. F. Casolari, *Il Digital Services Act e la costituzionalizzazione dello spazio digitale europeo*, in *Giurisprudenza italiana*, 2, 2024, 462; G. Finocchiaro, *"Digital Services Act" - Responsabilità delle piattaforme. Responsabilità delle piattaforme e tutela dei consumatori*, in *Giornale di diritto amministrativo*, 6, 2023, 730 ss.; G. Sgueo, *"Digital Services Act - Governance". L'architettura istituzionale del "Digital Services Act"*, in *Giornale di diritto amministrativo*, 6, 2023, 746 ss.; S. Scola, *"Digital Services Act": occasioni mancate e prospettive future nella recente proposta di regolamento europeo per il mercato unico dei servizi digitali*, in *Contratto e impresa. Europa*, 1, 2022, 127 ss.; G. Giordano, *La responsabilità degli intermediari digitali nell'architettura del "Digital Services Act": è necessario che tutto cambi affinché tutto rimanga com'è?*, in *Comparazione e diritto civile*, 1, 2023, 193 ss.

«prevenire l'eccesso di regolazione ed evitare al tempo stesso un deficit di intervento regolatorio rispetto a pratiche che si possano manifestare a distanza o che evolvono nel tempo»[115].

Anticipando aspetti su cui si avrà modo di ritornare, è possibile sin d'ora rilevare come, nel contrasto alla disinformazione, il carattere dialogico della regolazione emerga, ad esempio, laddove si prevede che siano le piattaforme e i motori di ricerca di grandi dimensioni a (i) valutare i possibili "rischi sistemici"[116] connessi alla progettazione o al funzionamento del loro servizio e dei suoi sistemi, compresi i sistemi algoritmici (art. 34)[117] e, conseguentemente, ad (ii) adottare «misure di attenuazione ragionevoli, proporzionate ed efficaci, adattate ai rischi sistemici specifici individuati a norma dell'articolo 34, prestando particolare attenzione agli effetti di tali misure sui diritti fondamentali» (art. 35).

Dunque, sia i "rischi sistemici" (art. 34 par. 1) – ovvero quelli suscettibili di incidere su interessi di indubbio rilievo costituzionale (la libertà di espressione e di informazione, il pluralismo dei media, il dibattito pubblico, i processi elettorali e la sicurezza pubblica[118]) - sia i "fattori di rischio" (art. 34 par. 2) - ovvero le cause dei possibili rischi (tra cui anche quelle costituite dai sistemi di moderazione dei contenuti, dalla diffusione di contenuti illegali o contrari alle condizioni d'uso del servizio)[119] - sia le "misure di attenuazione" di cui all'art. 35 (ad es. la sperimentazione e l'adeguamento dei sistemi algoritmici), sono configurati come "cataloghi aperti", integrabili dalle piattaforme[120],

---

115 F. Di Porto-N. Rangone, *Strategie regolatorie e qualità della regolazione*, cit., 95-96.

116 In dottrina si è posto in luce come l'approccio regolatorio basato sul rischio costituisca il paradigma cui l'UE è ricorsa nella strategia europea sul digitale. In particolare, I.P. Di Ciommo, *Riserva di umanità, prevedibilità, rischio: categorie giuridiche e innovazione digitale*, in F. Fabrizzi-L. Durst (a cura di), *Controllo e predittività*, cit., 60 rileva come il *risk-based approach*, che ha ispirato il GDPR, il DSA e, da ultimo, l'AI Act, «sia di fatto funzionale a bilanciare opposte esigenze: assecondare l'innovazione ed il progresso tutelando, al contempo, i diritti degli individui». Con specifico riferimento al DSA, E. Longo, *La disciplina del "rischio digitale"*, in F. Pizzetti, *La regolazione europea*, cit., 75 evidenzia come il "rischio" costituisca, al contempo, giustificazione e oggetto della regolazione e richiama l'attenzione su due specifici aspetti: «il DSA non fornisce una definizione di cosa si intenda con questa formula ["rischi sistemici"] pur riferendosi a essa in modo pervasivo. Il DSA è pieno di meccanismi di regolazione del rischio che riguardano l'identificazione, la gestione e la mitigazione, l'informazione e la trasparenza, la vigilanza esterna e interna, nonché la revisione e la rendicontazione dei rischi. Perciò, al di là del semplice testo, sarà necessario che nella prassi e nella costruzione dell'*enforcement* del Regolamento si creino le basi per una nuova individuazione e approfondimento dei caratteri propri dei "rischi digitali". La seconda riflessione tocca il cuore della sfida regolatoria europea alle grandi piattaforme. La strategia giustificata sul rischio rappresenta una scelta rivolta a creare una meta-regolazione non impositoria ma neanche troppo libertaria […]. L'approccio basato sul rischio si accompagna infatti a meccanismi di co-regolazione più che a semplici impostazione di *command and control* o di pura *self-regulation*». Sulla rilevanza del rischio quale fondamento dell'impianto degli obblighi di *due diligence*, v. M. Orofino, *Il Digital Service Act tra continuità (solo apparente) ed innovazione*, in F. Pizzetti, *La regolazione europea*, cit., 174.

117 E. Birritteri, *Contrasto alla disinformazione* cit., 73: «Si tratta di una norma chiave che si pone l'obiettivo di sensibilizzare le piattaforme sull'esigenza di farsi carico degli interessi di tutti gli *stakeholder* che possono in qualche misura essere influenzati dalla loro attività, non potendo le esigenze di business e di profitto essere perseguite a discapito di tali diritti individuali e beni collettivi. Ciò secondo un approccio sistematico e sfruttando la capacità organizzativa e di gestione di modelli di compliance e metodologie di analisi del rischio che simili grandi corporation certamente possiedono».

118 In ordine al rapporto con la disinformazione, v. considerando 83, 84 e 88.

119 V. considerando 84.

120 A. Palumbo-J. Piemonte, *Delega di funzioni regolamentari e lotta ai rischi sistemici causati della disinformazione*

seppure, come si vedrà, sotto la supervisione pubblica.

Al dialogo regolatorio è improntato anche il "meccanismo di risposta alle crisi" (art. 36)[121], in forza del quale, quando circostanze eccezionali comportano una grave minaccia per la sicurezza pubblica o per la salute pubblica nell'Unione o in sue significative parti[122], le piattaforme o i motori di ricerca di dimensioni molto grandi sono tenuti – in seguito alla decisione assunta dalla Commissione e su raccomandazione del Comitato europeo per i servizi digitali – a: (i) valutare il modo in cui il funzionamento e l'uso dei loro servizi contribuiscano, o possano contribuire, all'insorgere della minaccia; (ii) individuare e applicare misure specifiche, efficaci e proporzionate; (iii) relazionare alla Commissione in ordine a tali profili.

Dunque, anche in questo caso, l'individuazione della soluzione più efficace viene rimessa alla discrezionalità delle piattaforme.

Significativa, in tal senso, la previsione di cui all'art. 36 par. 5: «la scelta delle misure specifiche da adottare a norma del paragrafo 1, lettera b), e del paragrafo 7, secondo comma, spetta al fornitore o ai fornitori destinatari della decisione della Commissione»[123]; tuttavia, «la Commissione può avviare, di propria iniziativa o su richiesta del fornitore, un dialogo con quest'ultimo per stabilire se, alla luce delle circostanze specifiche del fornitore, le misure previste […] siano efficaci e proporzionate ai fini del conseguimento degli obiettivi perseguiti» (par. 7).

Nella stessa ottica si colloca la disciplina dei "protocolli di crisi" adottabili in caso di «circostanze straordinarie che incidono sulla sicurezza pubblica o sulla salute pubblica» (art. 48), atteso che la loro elaborazione e applicazione può solo essere incoraggiata dalla Commissione, su raccomandazione del Comitato europeo per i servizi digitali.

Infine, afferiscono al modello della co-regolazione i codici di condotta[124] (artt. 45-46-

nel Digital Services Act: quali rischi per la libertà di espressione?, in questa *Rivista*, 3, 2023, 126-127 rilevano un contrasto con il principio di legalità quale limite alle restrizioni dei diritti fondamentali e, in particolare, per quanto qui rileva, della libertà di espressione: «Vi sono due principali criticità sollevate dall'art. 35 per il rispetto del principio di legalità. Da un lato, la definizione dei rischi sistemici e le condizioni per l'adozione delle misure di attenuazione dei rischi e, dall'altro, la portata delle possibili restrizioni che ne conseguono. Entrambi questi aspetti influiscono sulla prevedibilità dell'interferenza con la libertà di espressione. In primo luogo, i rischi sistemici che giustificano l'adozione di misure di attenuazione non sono sufficientemente definiti all'interno del *DSA*, e anzi la loro individuazione è lasciata a *VLOPs* e *VLOSEs*. […]In secondo luogo, l'art. 35 non definisce quale portata possano avere le misure di attenuazione sulla libertà di espressione».

[121] V. considerando 91.

[122] Si pensi agli scenari richiamati da L. Ciliberti, *Free flow of information – Il contrasto alla disinformazione in tempi di guerra*, in questa *Rivista*, 2, 2022, 349 ss.

[123] Secondo l'interpretazione offerta dalla dottrina, dovrebbe ritenersi che: «La Commissione possa imporre, nella propria decisione, soltanto l'adozione di un certo e ampio *genus* di misure (ad es., l'adeguamento dei sistemi di raccomandazione delle notizie, oppure delle condizioni generali o delle procedure di moderazione dei contenuti), dovendo essere lasciate al libero apprezzamento del soggetto regolato, in ultima istanza, la costruzione e l'attuazione specifica della *policy*, della misura di dettaglio da adottare, la scelta su come mettere a terra concretamente le modifiche delle proprie regole autonormate, senza che i poteri di ingerenza dell'organo pubblico europeo possano spingersi fino a obbligare gli attori privati ad adottare misure predeterminate, escludendo qualsiasi margine di scelta su come implementare e integrare il tipo di provvedimenti richiesti all'interno del proprio contesto operativo interno» (E. Birritteri , *Contrasto alla disinformazione* cit., 80).

[124] I codici di condotta previsti dal DSA sono stati anticipati dal Codice di condotta contro la disinformazione adottato a livello europeo nel 2018, poi modificato e rafforzato nel 2022 (*Strengthened*

47), che ne costituiscono paradigmatico esempio[125] in quanto adottati su base volontaria[126] per contribuire alla corretta applicazione del Regolamento[127].

Come accennato, il coinvolgimento degli operatori economici nella regolazione non è estraneo all'intervento pubblico.

In particolare, per quanto concerne l'adozione delle misure di attenuazione dei rischi sistemici *ex* art. 35, non è irrilevante che la Commissione, in cooperazione con i coordinatori dei servizi digitali e in seguito a consultazioni pubbliche, possa emanare orientamenti con l'obiettivo di definire le migliori pratiche e raccomandare eventuali misure (art. 35 par. 3), nonché collaborare alla definizione di quelle contenute nei codici di condotta (art. 45 par. 2).

Anche questi ultimi, infatti, presuppongono una *governance* pubblica[128], che vede come protagonisti la Commissione e il Comitato europeo per i servizi digitali con la funzione

Code of Practice on Disinformation 2022, in https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation).

O. Pollicino, *Asimmetrie valoriali transatlantiche*, cit., 233 rileva che: «Il Codice di condotta del 2018 contro la disinformazione, che pure rappresentava un unicum a livello mondiale quale modello di impegno volontario, da parte del potere digitale privato, ad adottare tutta una serie di misure che contenessero il fenomeno, è stato deludente quanto a vaghezza degli obblighi assunti da parte delle stesse piattaforme e l'assenza quasi completa di criteri per la verificabilità e la misurabilità degli impegni. In particolare, la versione del 2018 non prevedeva le condizioni fondamentali per rendere il codice uno strumento efficace al contrasto contro la disinformazione, in particolare considerata l'assenza di obbiettivi e linee guida definiti dalla Commissione e strumenti di misurazione delle misure adottate dai firmatari. Si è quindi deciso, in sostanza, a partire dall'autunno del 2021, di riscrivere, sotto il coordinamento di chi scrive ed alla luce delle linee guida della Commissione che erano nel frattempo state adottate, un nuovo codice che potesse colmare le lacune del precedente ed essere uno strumento assai più efficace per contrastare un fenomeno che, intanto, era diventato di una gravità assoluta per gli effetti di inquinamento e di polarizzazione del discorso pubblico, che sempre più caratterizza le "digital agorà" ospitate dalle grandi piattaforme». In merito al Codice per il contrasto alla disinformazione, v. M. Monti, *Lo "strengthened Code of Practice on Disinformation": un'altra pietra della nuova fortezza digitale europea?* in questa *Rivista*, 2, 2022, 317 ss.; G. Pagano, *Il Code of Practice on Disinformation. Note sulla natura giuridica di un atto misto di autoregolazione*, in *Federalismi.it*, 11, 2019, 1 ss.

[125] In generale, sui codici di condotta v.: L. Ammannati-F. Costantino, *Intelligenza artificiale e regolazione dei mercati digitali. Modelli di regolazione e di regolatori*, in A. Pajno-F. Donati-A. Perrucci (a cura), *Intelligenza artificiale e diritto: una rivoluzione?. Vol. 1 Diritti fondamentali, dati personali e regolazione*, Bologna, 2022, 571 ss.; N. Maccabiani, *Co-regolamentazione, nuove tecnologie e diritti fondamentali: questioni di forma e di sostanza,* in *Osservatorio sulle fonti*, 3, 2022.

[126] Considerando 103: «[…] L'attuazione dei codici di condotta dovrebbe essere misurabile e soggetta a controllo pubblico, tuttavia ciò non dovrebbe pregiudicare il carattere volontario di tali codici e la libertà delle parti interessate di decidere se aderirvi. In determinate circostanze è importante che le piattaforme online di dimensioni molto grandi cooperino all'elaborazione di specifici codici di condotta e vi aderiscano. Il presente regolamento non osta a che altri prestatori di servizi, attenendosi agli stessi codici di condotta, aderiscano alle stesse norme in materia di dovere di diligenza, adottino le migliori pratiche e traggano beneficio dagli orientamenti emanati dalla Commissione e dal comitato».

[127] M. Orofino, *Il Digital Service Act tra continuità (solo apparente) ed innovazione*, cit., 164: «Nel DSA, i Codici di condotta mantengono la loro rilevanza specifica come strumento essenziale per garantire l'applicazione corretta del Regolamento, come indicato negli artt. 45-47 e nei relativi considerando […] forniscono linee guida fondamentali al fine di coadiuvare i prestatori di servizi di intermediazione sia per la valutazione dei rischi che per la definizione di soluzione tecniche ed organizzative necessarie per attuare tali misure. Inoltre, i Codici di condotta sono chiamati a svolgere una funzione specifica per le piattaforme e i motori di ricerca di grandi dimensioni nelle situazioni in cui emerga un nuovo e significativ rischio sistemico. In tali circostanze, essi consentono un dibattito aperto, che coinvolge la società civile, per affrontare efficacemente il rischio identificato».

[128] S. Del Gatto, *Il Digital Services Act*, cit., 727.

di: incoraggiare e agevolare l'adozione dei codici; garantire che questi definiscano chiaramente i loro obiettivi e contengano indicatori chiave per la misurazione dei risultati; monitorare e valutare periodicamente il conseguimento degli obiettivi tenendo conto degli indicatori chiave di prestazione; incoraggiare l'eventuale riesame e adattamento periodico, nonché, in caso di inottemperanza sistematica, invitare i firmatari ad adottare le misure necessarie (art. 45).

Parimenti, nel meccanismo di risposta alle crisi attivato dalle piattaforme, la Commissione monitora l'applicazione delle misure adottate e riferisce periodicamente - almeno una volta al mese - al Comitato (art. 36 par. 7), nonché al Parlamento europeo e al Consiglio una volta all'anno e - in ogni caso - tre mesi dopo la fine della crisi (art. 36 par. 11). Inoltre, se ritiene che le misure previste o attuate non siano efficaci o proporzionate, può, previa consultazione del Comitato, chiedere al fornitore della piattaforma di riesaminarle (art. 36, par. 7).

Alla luce del quadro sinteticamente tracciato e anticipando le conclusioni di seguito svolte, è possibile già osservare come la scelta di delineare un dialogo regolatorio tra Istituzioni pubbliche e operatori dell'economia digitale risulti ragionevole[129] e, specie per gli aspetti involgenti la libertà di informazione, utile, se non addirittura necessaria. Invero, nei settori connotati da rapido sviluppo tecnologico, la regolazione presuppone un'*expertise* di livello talmente elevato che, per risultare efficace, non può ragionevolmente prescindere dalle competenze tecniche di cui gli stessi soggetti privati – i regolati – dispongono[130].

Se questo è vero in generale, nella prospettiva offerta dal presente scritto – la tutela della libertà di informazione - i vantaggi della co-regolazione risultano ancor più evidenti, non foss'altro perché, come è stato evidenziato, gli operatori dispongono di un'enorme mole di dati e, dunque, di un vantaggio informativo su cui calibrare la regolazione; inoltre, possono assicurare una più efficace attuazione delle regole grazie ai sistemi algoritmici di cui dispongono e alla loro costante innovazione[131].

---

[129] L. Ammanati, *Regolatori e supervisori nell' era digitale: ripensare la regolazione*, in *Giur. cost*., 3, 2023, 1453 ss. In tal senso, anche M.E. Bartoloni, *La regolazione privata nel sistema costituzionale dell'Unione Europea. Riflessioni sulla disciplina relativa al settore dell'innovazione,* in *osservatoriosullefonti.it,* 2021, 3, 1334: «La regolazione privata [...] di recente si sta affermando soprattutto nella disciplina delle nuove tecnologie. Poiché la disciplina di settori relativi all'innovazione tecnologica esige un grado di *expertise* e competenze tecniche possedute essenzialmente da quegli stessi soggetti privati che devono essere regolati, si ritiene che una regolazione efficacie non possa prescindere dal loro coinvolgimento».

[130] M.E. Bartoloni, *La regolazione privata nel sistema costituzionale dell'Unione europea,* cit., 1334 rileva come le fonti europee tendano a coinvolgere i privati nella disciplina di settori "tecnologicamente sensibili". A. Iannuzzi, *Le fonti del diritto dell'Unione europea*, cit., 21: «il rimedio per non incorrere in una legislazione soggetta ad un rapida obsolescenza, e anche per non soffocare l'innovazione tecnologica, non può essere altro che il ricorso ad una legislazione per principi».

[131] L. Ammanati, *Regolatori e supervisori nell' era* digitale, cit. 1453. Come evidenziato da C. Pinelli, *L'evoluzione della normativa dell'Unione Europea*, in C. Pinelli-U. Ruffolo, *I diritti delle piattaforme*, Torino, 2023, 29: «La presa d'atto dei fallimenti dell'autoregolamentazione si accompagna alle preoccupazioni per le reazioni di rigetto che un meccanismo regolatorio troppo invasivo potrebbe provocare nelle piattaforme di maggiori dimensioni con una sua conseguente elusione, tanto più plausibile per via di uno spessore tecnico dei sistemi di comunicazione in rete non padroneggiabile dall'esterno, dunque neanche da Stati o dall'Unione Europea. La prudenza del meccanismo regolatorio potrebbe essere derivata dalla consapevolezza di rischi del genere e dalla necessità, in alternativa, di responsabilizzare direttamente le piattaforme coinvolgendole nell'individuazione da parte loro di congegni in grado di garantire la messa

Sino a qui, al fine di mettere in evidenza il carattere "cooperativo" della regolazione, l'analisi del DSA è stata svolta dall'angolo prospettico degli spazi riservati al regolatore pubblico e alle imprese digitali, che parrebbero configurarsi come "regolatori privati"[132] e, al tempo stesso, destinatari della regolazione pubblica.

Mutando prospettiva e focalizzando l'attenzione, non sulle modalità della regolazione, ma sulla finalità perseguita dal legislatore europeo[133], con specifico riguardo alla tutela della libertà di informazione emerge come l'assetto regolatorio delineato dal DSA persegua due principali obiettivi: il contrasto alle "manipolazioni algoritmiche" a tutela del diritto a una informazione veritiera, consapevole e plurale; l'introduzione di limiti all'attività di moderazione dei contenuti, a tutela della libertà di espressione e dei suoi corollari (la libertà di informare e il pluralismo informativo).

Siffatti obiettivi, coerenti con quelle che, nella prima parte dell'analisi, sono state individuate come principali minacce per la libertà di informazione, si traducono in direttrici della regolazione.

Invero, le regole delineate per gli operatori della Rete sono di intensità via via maggiore a seconda della tipologia di operatore economico - prestatori di servizi intermediari[134] (Capo III - Sez. I), prestatori di servizi di memorizzazione di informazioni (*hosting*), comprese le piattaforme online (Capo III - Sez. II), fornitori di piattaforme online[135] (Capo III – Sez. III), piattaforme di dimensioni molto grandi (cd. *Very Large Online Platform*–VLOP (ad es. Facebook, Twitter)) o motori di ricerca[136] di dimensioni molto grandi[137] (cd. *Very Large Online Search Engines*-VLOSEs) (Capo III - Sez. V) – ma, per quanto

in opera del meccanismo stesso, e prima ancora di farli conoscere».

[132] *Ibid.*

[133] «Garantire un ambiente online sicuro, prevedibile e affidabile, in cui i diritti fondamentali sanciti dalla Carta siano efficacemente tutelati e l'innovazione sia agevolata, contrastando la diffusione di contenuti illegali online e i rischi per la società che la diffusione della disinformazione o di altri contenuti può generare» (considerando 9).

[134] Ai sensi dell'art. 3, lett. g) del Regolamento, per "servizio intermediario" si intende «uno dei seguenti servizi della società dell'informazione: un servizio di semplice trasporto (cosiddetto "*mere conduit*"), consistente nel trasmettere, su una rete di comunicazione, informazioni fornite da un destinatario del servizio o nel fornire accesso a una rete di comunicazione; un servizio di memorizzazione temporanea (cosiddetto "*caching*"), consistente nel trasmettere, su una rete di comunicazione, informazioni fornite dal destinatario del servizio, che comporta la memorizzazione automatica, intermedia e temporanea di tali informazioni effettuata al solo scopo di rendere più efficiente il successivo inoltro delle informazioni ad altri destinatari su loro richiesta; un servizio di memorizzazione di informazioni (cosiddetto "*hosting*"), consistente nel memorizzare informazioni fornite da un destinatario del servizio su richiesta dello stesso».

[135] Ai sensi dell'art. 3, lett. i), per "piattaforma online" si intende «un servizio di memorizzazione di informazioni che, su richiesta di un destinatario del servizio, memorizza e diffonde informazioni al pubblico, tranne qualora tale attività sia una funzione minore e puramente accessoria di un altro servizio o funzionalità minore del servizio principale e, per ragioni oggettive e tecniche, non possa essere utilizzata senza tale altro servizio e a condizione che l'integrazione di tale funzione o funzionalità nell'altro servizio non sia un mezzo per eludere l'applicabilità del presente regolamento».

[136] Ai sensi dell'art. 3, lett. i) del Regolamento, per "motore di ricerca online" si intende «un servizio intermediario che consente all'utente di formulare domande al fine di effettuare ricerche, in linea di principio, su tutti i siti web, o su tutti i siti web in una lingua particolare, sulla base di un'interrogazione su qualsiasi tema sotto forma di parola chiave, richiesta vocale, frase o di altro *input*, e che restituisce i risultati in qualsiasi formato in cui possono essere trovate le informazioni relative al contenuto richiesto».

[137] Ai sensi dell'art. 33 del Regolamento, sono considerati "piattaforme e motori di ricerca di dimensioni

qui rileva, gli obiettivi cui tendono sono essenzialmente: (i) la salvaguardia della libertà di espressione e di informazione attraverso l'introduzione di limiti alla *content moderation* e (ii) il contrasto ai possibili effetti lesivi derivanti dall'utilizzo di sistemi algoritmici e di AI, capaci – lo si è visto – di comprimere il diritto all'autodeterminazione informativa[138] e di veicolare disinformazione[139].

Pur essendo il quadro regolatorio delineato dal DSA molto ampio ed articolato, nelle pagine seguenti si approfondiranno, in particolare, i suddetti aspetti, per poi trarre talune considerazioni conclusive in ordine alle implicazioni suscettibili di derivarne.

4. I limiti alla *content moderation*

Sul fronte dei limiti alla *conten moderation* - ovvero all'attività di controllo svolta dalle piattaforme sui contenuti immessi in Rete, cui può conseguire l'adozione di una pluralità di misure limitative dell'utente (ad es. la rimozione dei contenuti, la sospensione o disattivazione dell'account, la chiusura delle pagine che violano le condizioni d'uso del servizio, ecc.) - il DSA[140] ha previsto regole di *due diligence* in capo ai prestatori di servizi, con specifiche prescrizioni per le piattaforme online e per le piattaforme e i motori di ricerca di dimensioni molto grandi.

Come si noterà procedendo con l'analisi, anche questo versante della regolazione intercetta le criticità connesse all'utilizzo dell'AI, che, tuttavia, viene in rilievo non come possibile veicolo di disinformazione o di compressione del pluralismo informativo, bensì come strumento di individuazione dei contenuti illegali o comunque contrastanti

molto grandi" quelli con un numero medio mensile di destinatari attivi del servizio nell'Unione pari o superiore a 45 milioni, designati come tali dalla Commissione (art. 33, par. 4). Per le VLOPs e VLOSEs il considerando 79 stabilisce che data la "modalità di progettazione dei loro servizi", che è «generalmente ottimizzata a vantaggio dei loro modelli di business spesso basati sulla pubblicità e può destare preoccupazioni per la società», sono «necessarie una regolamentazione e un'esecuzione efficaci al fine di individuare e attenuare adeguatamente i rischi e i danni sociali ed economici che possono verificarsi».

138 A. Papa, *La problematica tutela del diritto all'autodeterminazione informativa nella big data society*, in AA.VV., *Liber amicorum per Pasquale Costanzo*, in *Consulta online*, 17 aprile 2020.

139 Molto chiaro, in tal senso, è il quadro descritto da E. Biritteri, *Contrasto alla disinformazione* cit., 3: «Ad esempio, come noto, i sistemi di raccomandazione tendono a riproporre all'utente contenuti sempre più in linea con la propria precedente attività in rete, con la conseguenza di innescare un continuo bombardamento nei suoi riguardi di contenuti falsi che lo hanno già in precedenza interessato e che rischiano così di divenire rapidamente virali in rete con tutto ciò che di negativo può derivarne – o di post potenzialmente molto pericolosi per il suo benessere psicofisico (si pensi a utenti che tendono ad essere attratti, per uno stato depressivo, da informazioni relative ad atti di autolesionismo). Si può far riferimento, altresì, alle tecniche di manipolazione intenzionale del servizio (tra cui l'interazione artificiosa tra più account per aumentare in modo fraudolento la visibilità di certe notizie, o l'uso agli stessi fini di bot automatici e profili fake) spesso utilizzati in campagne coordinate di disinformazione».

140 L'art. 34 DSA annovera tra i fattori di rischio suscettibili di incidere, ad esempio, sul dibattito civico e sui processi elettorali anche i sistemi di moderazione dei contenuti; conseguentemente, tra le misure di attenuazione dei rischi, l'art. 3 indica «l'adeguamento delle procedure di moderazione dei contenuti, compresa la velocità e la qualità del trattamento delle segnalazioni concernenti tipi specifici di contenuti illegali e, se del caso, la rapida rimozione dei contenuti oggetto della notifica o la disabilitazione dell'accesso agli stessi, in particolare in relazione all'incitamento illegale all'odio e alla violenza online, nonché l'adeguamento di tutti i processi decisionali pertinenti e delle risorse dedicate alla moderazione dei contenuti».

con le condizioni d'uso del servizio applicate all'utente.

Volendo schematizzare l'assetto delle regole definite dal DSA con riguardo alla *content moderation*[141], è possibile enucleare tre principali linee di intervento: (i) quella orientata a garantire la trasparenza delle attività di moderazione, (ii) quella volta a plasmare il procedimento finalizzato all'adozione di eventuali misure restrittive, infine, (iii) quella concernente le forme di tutela avverso le limitazioni imposte all'utente.

Nell'ambito della prima linea di intervento viene in rilievo l'obbligo di inserire nelle condizioni contrattuali[142] informazioni riguardanti le politiche, le procedure, le misure e gli strumenti utilizzati per la moderazione dei contenuti, compresi il processo decisionale algoritmico e la verifica umana (art. 14, par. 1)[143], nonché quello di pubblicare,

[141] L'art. 3, lett. t) del Regolamento definisce la moderazione dei contenuti come «attività, automatizzate o meno, svolte dai prestatori di servizi intermediari con il fine, in particolare, di individuare, identificare e contrastare contenuti illegali e informazioni incompatibili con e condizioni generali, forniti dai destinatari del servizio, comprese le misure adottate che incidono sulla disponibilità, sulla visibilità e sull'accessibilità di tali contenuti illegali o informazioni, quali la loro retrocessione, demonetizzazione o rimozione o la disabilitazione dell'accesso agli stessi, o che incidono sulla capacità dei destinatari del servizio di fornire tali informazioni, quali la cessazione o la sospensione dell'account di un destinatario del servizio».

[142] Sul rapporto intercorrente tra il DSA e il Codice del consumo, l'AGCM ha concluso: «è lo stesso DSA (cfr. articolo 2, ma anche il considerando 10) a fare espressamente salvo il diritto europeo in materia di tutela dei consumatori. Peraltro, la "Risoluzione del Parlamento europeo del 12 dicembre 2023 sulla progettazione di servizi online che crea dipendenza e sulla tutela dei consumatori nel mercato unico dell'UE", nel prendere atto delle diverse previsioni normative già in vigore, tra cui il DSA (espressamente richiamato nei "visti"), ritiene in ogni caso possibile vietare tali condotte nell'ambito del vigente quadro normativo di tutela del consumatore (punto 6)». V. il provvedimento 21 maggio 2024, n. PS12566, con cui AGCM ha sanzionato Meta per aver posto in essere una pratica commerciale scorretta in violazione dell'art. 20 del Codice del consumo, segnatamente, per aver omesso « i) con riguardo alla piattaforma FB, di indicare le modalità (automatizzata o manuale) con cui è stata assunta la decisione di sospendere l'account, ossia di interrompere i propri servizi; ii) con riguardo a entrambi i *social network*, di fornire indicazioni della possibilità di contestare la decisione di sospendere l'account, oltre che con ricorso interno, anche adendo un organo di risoluzione extragiudiziale delle controversie o ricorrendo a un giudice per contestare la decisione del Professionista e aver previsto un termine relativamente breve (di 30 giorni) per contestare tramite ricorso interno diretto a Meta la decisione del Professionista». Sul punto, l'Autorità ha rilevato: «l'omessa indicazione di informazioni utili a contestare l'interruzione dei servizi FB e IG non può essere ritenuta conforme alla diligenza professionale attesa da un operatore come Meta. Per il consumatore non è indifferente, ai fini di un'eventuale contestazione della decisione assunta dal Professionista, conoscere le modalità con cui tale decisione è stata adottata. Anche l'indicazione dei mezzi alternativi al ricorso interno (organo giurisdizionale o stragiudiziale) è rilevante, potendo il consumatore ritenere, in difetto della stessa, che quest'ultimo sia l'unico mezzo a disposizione per contestare la decisione di Meta. Il consumatore, infatti, potrebbe considerare che *social network* come IG e FB, essendo di dimensione "sovranazionale", siano in qualche modo insindacabili dal giudice nazionale o da soggetti diversi da Meta, quali gli organi stragiudiziali di risoluzione delle controversie (*ADR)*. Peraltro, l'interruzione dei servizi offerti da Meta, in concreto, impedisce all'utente di svolgere attività sociali o professionali a esso collegate e considerabili "essenziali" nella società contemporanea (c.d. "*de-platforming*"): il Professionista diligente avrebbe quindi senz'altro dovuto fornire ai soggetti privati del servizio la più ampia informativa, anche circa le modalità di contestazione ed eventuale risoluzione delle problematiche riscontrate, al fine di "reintegrare" nel *social network* l'utente attinto dalla misura sospensiva».

[143] L'attività di moderazione dei contenuti potrebbe comportare il trattamento dei dati personali, sebbene – come chiarito dal Garante europeo per la protezione dei dati personali con l'*Opinion 1/2021 on the proposal for a Digital Services Act* - «In conformità con i requisiti di minimizzazione dei dati e protezione dei dati fin dalla progettazione e per impostazione predefinita, la moderazione dei contenuti non dovrebbe, per quanto possibile, comportare alcun trattamento di dati personali. […]. Laddove il trattamento di dati personali sia necessario, come per il meccanismo di reclamo, tali dati

almeno una volta all'anno, relazioni chiare e facilmente comprensibili sulle attività di moderazione dei contenuti svolte (art. 15, par. 1)[144].

Nell'ottica della trasparenza si colloca anche l'obbligo (per i prestatori di servizi di memorizzazione di informazioni) di fornire una motivazione chiara e puntuale per le restrizioni imposte, specificando quantomeno: il contenuto e le ragioni della restrizione[145] sotto il profilo della base giuridica o delle condizioni contrattuali violate; le informazioni relative agli strumenti automatizzati usati per adottare la decisione o per individuare i contenuti stigmatizzati; i mezzi di ricorso a disposizione (art. 17).

Sul secondo versante – quello procedimentale - il DSA delinea una sorta di procedimento su segnalazione di terzi, prevedendo (con specifico riferimento ai prestatori di servizi di memorizzazione di informazioni (comprese le piattaforme online))[146] – un triplice ordine di obblighi: (i) la predisposizione di meccanismi volti consentire a qualsiasi persona o ente[147] di notificare la presenza di contenuti illegali; (ii) l'assunzione della decisione «in modo tempestivo, diligente, non arbitrario e obiettivo»; (iii) la comunicazione della decisione assunta senza indebito ritardo, fornendo informazioni sull'eventuale utilizzo di strumenti automatizzati, nonché sulle possibilità di ricorso disponibili (art. 16).

Ancora, con riguardo al "procedimento" per la moderazione dei contenuti, a tutti i prestatori di servizi intermediari viene imposto di agire «in modo diligente, obiettivo e proporzionato nell'applicare e far rispettare le restrizioni [all'uso dei servizi], tenendo debitamente conto dei diritti e degli interessi legittimi di tutte le parti coinvolte, compresi i diritti fondamentali dei destinatari del servizio, quali la libertà di espressione, la libertà e il pluralismo dei media, e gli altri diritti e libertà fondamentali sanciti dalla Carta [dei diritti fondamentali dell'UE]» (art. 14, par. 4).

Infine, sotto il terzo profilo investito dalla regolazione delle attività di *content moderation* - quello delle tutele attivabili avverso le restrizioni imposte all'utente – sono individuati due strumenti di tutela alternativi a quella giurisdizionale: il sistema interno di gestione

dovrebbero riguardare solo i dati necessari per questo scopo specifico, applicando nel contempo tutti gli altri principi del regolamento (UE) 2016/67926». Inoltre, in caso di trattamento dei dati personali, gli obblighi informativi previsti dall'art. 14 DSA sono da ritenersi complementari e integrativi di quelli *ex* artt. 12-14 GDPR e, come ulteriormente chiarito dal Garante, le misure adottate dovrebbero essere «il più possibile mirate e progettate in conformità a principi quali la minimizzazione dei dati e per impedire, per impostazione predefinita, sia la raccolta che la divulgazione di dati personali, in conformità all'articolo 25 del regolamento (UE) 2016/679».

144 Il contenuto delle relazioni viene specificamente dettagliato dall'art. 15. Ulteriori obblighi informativi, ad integrazione di quelli già previsti dall'art. 15, sono contemplati per i fornitori di piattaforme online dall'art. 24, e, per i fornitori di piattaforme online di dimensioni molto grandi o di motori di ricerca online di dimensioni molto grandi, dall'art. 42.

145 Ovvero: eventuali restrizioni alla visibilità di informazioni specifiche fornite dal destinatario del servizio, comprese la rimozione di contenuti, la disabilitazione dell'accesso ai contenuti o la retrocessione dei contenuti; la sospensione, la cessazione o altra limitazione dei pagamenti in denaro; la sospensione o la cessazione totale o parziale della prestazione del servizio; la sospensione o la chiusura dell'account del destinatario del servizio.

146 Sez. II ("Disposizioni aggiuntive applicabili ai prestatori di servizi di memorizzazione di informazioni, comprese le piattaforme online").

147 L'art. 22 detta una specifica disciplina per i "segnalatori attendibili".

dei reclami[148] e la contestazione dinanzi ad un organismo di risoluzione extragiudiziale delle controversie, dotato dei requisiti di cui all'art. 21 DSA e suscettibile di certificazione da parte del Coordinatore dei servizi digitali dello Stato membro in cui l'organismo è stabilito (in Italia, l'AGCOM ai sensi del d.l. n. 123/2023).

Delineato il quadro normativo concernente i limiti all'attività di moderazione dei contenuti, è possibili svolgere talune considerazioni in ordine all'adeguatezza - sotto tale profilo - delle scelte regolatorie operate in sede europea.

Pur lambendosi tematiche di indubbio rilievo pubblicistico, quali il ruolo dei cd. nuovi "poteri privati" e la loro compatibilità con l'ordinamento costituzionale, si resterà nel solco tracciato dalle riflessioni introduttive, con cui si è tentata una ricognizione dei rischi sottesi alla cd. "privatizzazione della censura", rischi involgenti non solo la libertà di espressione e il pluralismo informativo, ma, più in generale, il corretto funzionamento delle dinamiche democratiche.

Si è già avuto modo di osservare come l'opzione della co-regolazione, sicuramente preferibile a quella dell'auto-regolazione, già sperimentata e rivelatasi insufficiente, presenti indubbi vantaggi anche rispetto all'etero-regolazione, con cui si rinuncerebbe a sfruttare l'*expertise* delle piattaforme in un settore connotato, non solo, dall'incessante sviluppo tecnologico e dunque dal rischio della rapida obsolescenza regolatoria, ma anche da dimensioni talmente vaste da far sembrare persino illusoria l'aspirazione ad un controllo integralmente pubblico.

In particolare, sul piano dei limiti alla *content moderation*, la flessibilità insita nella co-regolazione emerge con tutta evidenza sotto il profilo che potremmo definire "sostanziale". Invero, l'individuazione dei contenuti suscettibili di rimozione è rimessa completamente alla discrezionalità delle imprese, anziché, come auspicato da una parte della dottrina, essere limitata ai soli contenuti contrari al diritto UE e degli Stati membri (contenuti illeciti)[149].

Siffatto profilo desta già un primo interrogativo, concernente la compatibilità della scelta regolatoria con il nostro quadro costituzionale.

Segnatamente, occorre chiedersi se e in che misura la nostra Costituzione consenta agli operatori economici l'introduzione di limitazioni più o meno intense alle libertà garantite dall'art. 21 Cost..

La risposta può essere rinvenuta nell'art. 41 Cost.[150], atteso che non solo l'attività di

[148] Più nel dettaglio, il DSA prevede che i fornitori di piattaforme online forniscano ai destinatari del servizio, comprese le persone o gli enti che hanno presentato una segnalazione, l'accesso a un sistema interno di gestione dei reclami efficace, volto a consentire la presentazione - per via elettronica e gratuitamente - di reclami contro le limitazioni imposte a causa della diffusione di contenuti illegali o incompatibili con le condizioni generali, nonché avverso le decisioni assunte dal fornitore della piattaforma sulle segnalazioni ricevute. L'accesso al sistema interno di gestione dei reclami deve essere garantito per almeno sei mesi (decorrenti dal giorno in cui il destinatario è stato informato della decisione) e il reclamo presentato attraverso il sistema interno di gestione, che non può essere deciso esclusivamente con strumenti automatizzati, deve essere gestito «in modo tempestivo, non discriminatorio, diligente e non arbitrario»; se fondato, deve concludersi con il tempestivo annullamento della decisione contestata (art. 20).

[149] O. Grandinetti, *Le piattaforme digitali*, cit., 184.

[150] U. Ruffolo, *Piattaforme e content moderation: "censura privata" o soft law governabile dall'autonomia negoziale (contrattuale, autodisciplinare, coregolamentare)? L'efficacia "orizzontale" di precetti costituzionali quali l'art. 21 ed il limite dell'ordine pubblico (e della "meritevolezza" dell'interesse contrattuale)*, in C. Pinelli-U. Ruffolo, *I diritti nelle*

moderazione dei contenuti viene posta in essere nell'esercizio di un'attività di impresa, ma anche la finalità perseguita, pur potendo sottendere scelte ideologiche[151], tende a collocarsi nell'ambito di una precisa strategia imprenditoriale[152], volta alla creazione di uno spazio virtuale accogliente e incentivante la partecipazione degli utenti.
In altri termini, a venire in rilievo sono i limiti posti dall'art. 41 Cost.[153], tra cui quello dell'"utilità sociale" - «principio-valvola» in grado di assicurare l'adeguamento dell'ordinamento al continuo evolversi della vita politica e sociale[154] - «consente una protezione dei diritti fondamentali in una fase per così dire collettiva della loro esistenza, quando cioè sono messi in pericolo non tanto in quanto riferiti a un singolo individuo, ma in un orizzonte più ampio, nel contesto di una collettività più o meno ampia e definita di persone […]»[155]; invero, mediante l'utilità sociale, quale «tramite tra la sfera individuale e quella collettiva dei diritti», trovano attuazione il principio della solidarietà sociale e quello della ragionevolezza, che costituiscono «il metro per decidere in merito a come

piattaforme, Torino, 2023, 43.

[151] Tale profilo interseca quello relativo alla possibilità di equiparare il controllo delle piattaforme sui contenuti alle scelte di posizionamento "etico" dell'editore; in tal senso, v. U. Ruffolo, *Piattaforme e content moderation*, cit., 57. Tuttavia, da una diversa prospettiva, si potrebbe ritenere che la selezione dei contenuti operata dalle piattaforme produca implicazioni ben maggiori, atteso che – come è stato detto - «il fenomeno, nel mondo digitale, assume dimensioni incomparabili e sembra presentare differenze anche qualitative, dal momento che mentre i media tradizionali, per quanto ideologicamente orientati, si confrontano comunque tra di loro in una sorta di *agorà* mediatica, sicché ogni testata è per forza di cose tenuta a dare almeno conto dell'esistenza dell'altro e a rapportarvisi, nel mondo digitale la condivisione dei contenuti avviene all'interno di cerchie, più o meno ristrette, che si formano e si ridefiniscono in continuazione, in cui i rapporti personali tendono a sovrapporsi alle affinità culturali e ideologiche e dove il confronto con posizioni diverse è per lo più precluso in partenza, se non nella forma dello scontro violento, del discorso d'odio, della scomposta invettiva» (G.E. Vigevani, *I media di servizio pubblico nell'età della rete. Verso un nuovo fondamento costituzionale, tra autonomia e pluralismo*, Torino, 2018, 23).

[152] G. Monti, *Privatizzazione della censura*, cit., 37.

[153] Secondo U. Ruffolo, *Piattaforme e content moderation* cit., non solo l'art. 41 Cost., ma anche l'art. 21 Cost. quale norma costituzionale direttamente applicabile (con "effetti orizzontali"), è in grado di condizionare l'autonomia contrattuale, precludendo le limitazioni irragionevoli della libertà di espressione. Si segnala anche la diversa impostazione che, qualificando i *social networks* come "formazioni sociali" *ex* art. 2 Cost., desume da tale inquadramento il fondamento dei "poteri" delle piattaforme e, conseguentemente, rinviene nella tutela dei diritti inviolabile il limite entro cui detti poteri possono essere esercitati. V. in tal senso, M.R. Allegri, *Ubi social ibi ius. Fondamenti costituzionali dei social network e profili giuridici della responsabilità dei provider*, Milano, 2018, 35: «la nozione di formazione sociale, infatti, non presuppone affatto che tutti i suoi membri siano posti sullo stesso piano. Al contrario, quanto più una formazione sociale è dotata di un'organizzazione stabile, tanto più si avverte l'esigenza di suddividere le funzioni al suo interno, prevedendo un potere di comando che detta le norme di comportamento legate ai fini associativi ed eventualmente faccia valere sanzioni nei confronti di chi viola la disciplina di gruppo; quindi non tutte le parti godono di analoghe sfere di libertà: talune si trovano in posizione dominante e altre in posizione più debole. Il *social network provider*, dunque, è membro costitutivo – anzi, indispensabile – della formazione sociale, ed è tenuto al pari degli altri membri al rispetto dei diritti individuali di tutti gli aderenti». Sulla qualificazione in termini di "formazioni sociali", v. P. Passaglia, *Internet nella Costituzione italiana: considerazioni introduttive*, in M. Nisticò - P. Passaglia (a cura di), *Internet e Costituzione*, Torino, 2014, 37 ss.; in senso critico, v. M. Cuniberti, *Poteri e libertà nella rete*, in *Medialws. Rivista di diritto dei media*, 3, 18, 44 ss., che valorizza le potenzialità di tutela insite nell'art. 41 Cost.

[154] A. Baldassarre, *Iniziativa economica privata (libertà di),* in *Enciclopedia del diritto*, vol. XXI, Milano, 1971, 604.

[155] L. Delli Priscoli, *Il limite dell'utilità sociale nelle liberalizzazioni*, in *Giurisprudenza commerciale*, 2, 2014, 359-360.

effettuare il necessario bilanciamento di valori»[156].

La possibilità che il bilanciamento tra diritti fondamentali abbia un'incidenza sull'autonomia contrattuale - di cui le condizioni d'uso del servizio stabilite dalle piattaforme sono evidentemente esplicazione è spiegabile alla luce della immediata applicabilità delle disposizioni costituzionali[157] ai rapporti interprivati[158], aspetto – quest'ultimo – su cui si avrà modo di tornare anche in sede conclusiva.

Applicando siffatte coordinate alla questione che ci occupa, si è ammesso che le imprese digitali possano «modulare negozialmente, nei rapporti con gli utenti, sia le tipologie di contenuti che le stesse si obbligano ad ospitare, sia quelle che si riservano la facoltà di vietare, rifiutare o rimuovere. Non ne consegue, tuttavia, la validità di qualsiasi clausola negoziale […]. La previsione costituzionale, ed il conseguente e cogente principio generale di protezione e rispetto della libertà di pensiero e comunicazione, possono dunque costituire, di volta in volta ed in relazione alle concrete circostanze del caso […] limite alla legittimità e validità delle clausole contrattuali limitanti»[159].

Ed infatti, l'incidenza del regolamento contrattuale sulla libertà di manifestazione del pensiero, e viceversa, ha trovato concreto riscontro in talune recenti applicazioni giurisprudenziali.

Invero, sebbene sia emerso un orientamento – minoritario[160] e poco condiviso in dot-

---

156 *Ibid.*

157 La questione intercetta il tema della "efficacia orizzontale" delle norme costituzionali, spesso richiamato dalla dottrina al fine di spiegare le differenze intercorrenti con l'approccio accolto negli USA, laddove la diretta applicabilità delle previsioni costituzionali ai soggetti privati presuppone che sussistano le condizioni per equipararli a soggetti pubblici: «Oltreoceano […] di norma, è esclusa l'efficacia delle previsioni costituzionali nei rapporti interprivati, salvo che un soggetto privato rivesta nei fatti un ruolo assimilabile a quello svolto da un soggetto pubblico (*state actor*) esercitando nei confronti di altri privati poteri assimilabili a quelli pubblici (si tratta delle nota e risalente *state action doctrine*). Peraltro, le rigorose condizioni richieste per superare il test per la qualificazione come *state actor* rendono difficile trarre da questa sola qualifica (potere privato) conseguenze giuridiche nell'ordinamento statunitense. In Europa ed in Italia, la tradizione dell'efficacia diretta delle disposizioni costituzionali (nota come *Drittwirkung*) invece consente di ricollegare a quella qualifica alcune conseguenze giuridiche ed apre, quantomeno per la giurisprudenza maggioritaria all'introduzione del rispetto delle libertà costituzionali tra i parametri per valutare i limiti all'autonomia contrattuale delle parti, anche relativamente alle CGC delle piattaforme digitali» (in termini, O. Grandinetti, *Le piattaforme digitali*, cit., 179-180). Sulle implicazioni che ne derivano con riguardo alle diverse modalità di tutela della libertà di espressione, v. altresì, v. R. Niro, *Piattaforme digitali e libertà di espressione fra autoregolamentazione e coregolazione: note ricostruttive*, cit., ss.; M. Bassini, *Libertà di espressione*, cit., 48 ss.; O. Pollicino, *L'efficacia orizzontale dei diritti fondamentali previsti dalla Carta. La giurisprudenza della Corte di giustizia in materia di digital privacy come osservatorio privilegiato*, in questa *Rivista*, 3, 2018, 138 ss.

158 P. Perlingeri, *Il diritto civile nella legalità costituzionale secondo il sistema italo-comunitario delle fonti*, Napoli, 2006, 535 ss.; id, *Profili del diritto civile*, Napoli, 1994, 18.

159 U. Ruffolo, *Piattaforme e content moderation*, cit, 55.

160 In modo emblematico, Trib. Roma, sez. spec. imprese, ord. 12 dicembre 2019: «È infatti evidente il rilievo preminente assunto dal servizio di Facebook (o di altri social network ad esso collegati) con riferimento all'attuazione di principi cardine essenziali dell'ordinamento come quello del pluralismo dei partiti politici (art. 49 Cost.), al punto che il soggetto che non è presente su Facebook è di fatto escluso (o fortemente limitato) dal dibattito politico italiano […]. Ne deriva che il rapporto tra Facebook e l'utente che intenda registrarsi al servizio (o con l'utente già abilitato al servizio come nel caso in esame) non è assimilabile al rapporto tra due soggetti privati qualsiasi in quanto una delle parti, appunto Facebook, ricopre una speciale posizione: tale speciale posizione comporta che Facebook, nella contrattazione con gli utenti, debba strettamente attenersi al rispetto dei principi costituzionali e ordinamentali».

trina[161] – incline a ridimensionare la rilevanza del rapporto contrattuale piattaforma-utente, la giurisprudenza prevalente[162], in modo condivisibile, tende a qualificare la disattivazione dell'*account* come esercizio del diritto di recesso per giusta causa conseguente alla violazione delle condizioni contrattuali da parte dell'utente[163].

Significative - nel solco di altre dello stesso tenore - le pronunce secondo cui «la società resistente [non] può seriamente essere paragonata ad un soggetto pubblico nel fornire un servizio, pur di indubbia rilevanza sociale e socialmente diffuso, comunque prettamente privatistico»[164]; «occorre, infatti, pur sempre fare riferimento allo scenario contrattuale di fronte al quale ci si trova: ovvero l'esercizio di un diritto di recesso, peraltro pattiziamente convenuto – o meglio unilateralmente predisposto ed accettato per adesione – all'interno di un contratto atipico e corrispettivo di fornitura di servizi che, quanto alle persone fisiche, assume anche la specificità di contratto di consumo»[165].

---

[161] V. G.E. Vigevani, *Dal "caso Casapound" del 2019 alla "sentenza Casapound" del 2022: piattaforme digitali, libertà d'espressione e odio on line nella giurisprudenza italiana*, in questa *Rivista*, 2, 2023, 147, che richiama, sul punto, P. Villaschi, *Facebook come la RAI?: note a margine dell'ordinanza del Tribunale di Roma del 12.12.2019 sul caso CasaPound c. Facebook*, in *Osservatorio Costituzionale*, 2, 2020, 441: «L'inquadramento delle piattaforme come "fori pubblici" o comunque soggetti esercenti un servizio pubblico appare problematico, trattandosi di imprese che perseguono finalità commerciali e che non hanno, *ex lege*, obblighi di servizio pubblico e di pluralismo interno. […] Come sottolinea con chiarezza Pietro Villaschi, una simile linea argomentativa sconta "un vizio di fondo, in quanto, in assenza, come si diceva, di una solida ricostruzione dottrinale e giurisprudenziale, va di fatto surrettiziamente a configurare come pubblici soggetti che tali non sono, né offrono un servizio pubblico (come, invece, per fare un esempio legato sempre al mondo dei media è un ente come la RAI)». Inoltre – chiarisce l'A. – «resta la distinzione tra diritto di parola e diritto al mezzo, che non ha esplicito riconoscimento nella giurisprudenza costituzionale. Secondo la Corte, infatti, "che tutti abbiano diritto di manifestare il proprio pensiero con ogni mezzo, non può significare che tutti debbano avere, in fatto, la materiale disponibilità di tutti i possibili mezzi di diffusione, ma vuol dire, più realisticamente, che a tutti la legge deve garantire la giuridica possibilità di usarne o di accedervi. Tale principio sembra doversi estendere anche alla rete e, in particolare, ai social network, nonostante l'evidente diversità rispetto agli strumenti di comunicazione precedenti, pena la trasformazione per via giurisprudenziale di un diritto di libertà delle piattaforme in una funzione legislativamente attribuita al solo servizio pubblico, quella di assicurare il "pluralismo interno"».

[162] Ivi, 148 rileva anche un terzo orientamento, che, pur aderendo all'inquadramento contrattuale del rapporto utente-piattaforma, rinviene non solo il diritto, ma anche l'obbligo delle piattaforme di rimuovere contenuti discriminatori, incitanti all'odio o disinformativi, anche quando non integranti illecito sul piano penale. V. Trib. Roma, sez. dir. della persona, 5 dicembre 2022, n. 17909; Trib. Roma, sez. dir. della persona, ord. 23 febbraio 2020: «I contenuti, che inizialmente erano stati rimossi e poi a fronte della reiterata violazione hanno comportato la disattivazione degli account dei singoli ricorrenti e delle pagine da loro amministrate tutte ricollegabili a Forza Nuova, sono illeciti da numerosi punti di vista. Non solo violano le condizioni contrattuali, ma sono illeciti in base a tutto il complesso sistema normativo di cui si è detto all'inizio […] Facebook non solo poteva risolvere il contratto grazie alle clausole contrattuali accettate al momento della sua conclusione, ma aveva il dovere legale di rimuovere i contenuti, una volta venutone a conoscenza, rischiando altrimenti di incorrere in responsabilità (si veda la sentenza della CGUE sopra citata e la direttiva CE in materia), dovere imposto anche dal codice di condotta sottoscritto con la Commissione Europea».

[163] V. ivi, 152, che richiama in tal senso: Trib. Siena, sez. un. civ., ord. 19 gennaio 2020; Trib. Trieste, sez. civ., ord. 27 novembre 2020; Trib. Varese, sez. I civ., ord. 2 agosto 2022; Trib. Milano, 19 luglio 2021.

[164] Trib. Siena, sez. un. civ., ord. 19 gennaio 2020.

[165] Trib. Trieste, sez. civ., ord. 27 novembre 2020, che si rivela di particolare interesse anche laddove afferma che, in caso di recesso ingiustificato (integrante inadempimento contrattuale), potrebbero sussistere i presupposti della responsabilità extracontrattuale per la lesione dei diritti costituzionali dell'utente, essendo «evidente che esercitando il proprio diritto contrattuale di fruire dei servizi di FACEBOOK l'utente dia altresì sfogo a diritti primari, quali l'identità personale, la libertà di espressione e di pensiero, quella di associazione, ed altri. Questi diritti in larga misura trascendono la

Nondimeno, come innanzi chiarito, l'inquadramento privatistico non apre all'arbitrio assoluto e incondizionato delle piattaforme, atteso che il controllo giurisdizionale sulla validità delle condizioni contrattuali implica un vaglio - alla stregua dei parametri costituzionali (artt. 2, 3, 21, 41 co. 2 Cost.) - sulla ragionevolezza dei limiti alla libertà di espressione.

Anche sotto tale profilo si rinvengono conferme nei recenti arresti giurisprudenziali[166], alla cui stregua: (i) l'autonomia privata è soggetta, oltreché al rispetto delle norme imperative *ex* art. 1418 c.c., ai "limiti imposti dalla legge" ai sensi dell'art. 1322 co. 1 c.c., da intendersi come «ordinamento giuridico nella sua complessità, comprensivo delle norme di rango costituzionale e sovranazionale»; (ii) nell'esercizio dell'autonomia privata non è possibile limitare in modo ingiustificato un diritto costituzionale della controparte, nella fattispecie la libertà di manifestazione del pensiero; (iii) la previsioni di limitazioni alla libertà di espressione non comportano uno squilibrio contrattuale dei diritti e degli obblighi in danno dell'utente, e dunque, la vessatorietà delle clausole contrattuali, nella misura in cui tali limitazioni non si risolvono in una lesione dell'art. 21 Cost.; (iv) con riferimento agli *standard* della *community* adottati da Facebook, le limitazioni alla libertà di espressione sono poste a tutela di altri diritti costituzionalmente rilevanti, pertanto è da escludersi la vessatorietà delle relative clausole contrattuali.

Sebbene la tutela giurisdizionale possa costituire un argine alla discrezionalità esercitata dalle piattaforme in sede di moderazione dei contenuti, va però rilevato come, per tal via, si rischi di condizionare l'effettiva garanzia delle libertà costituzionali all'attivazione della tutela giurisdizionale da parte dell'utente sottoposto all'applicazione di "misure restrittive" e, dunque, al controllo *ex post* (se attivato)[167].

Sicché, sarebbe stata forse auspicabile la definizione di una più puntuale cornice rego-

specifica dinamica contrattuale, integrano - affiancandosi ad esso - l'oggetto dell'ordinaria prestazione contrattuale, intesa quale messa a disposizione del servizio offerto agli utenti in corrispettivo della cessione di dati personali»; pertanto, «dal medesimo fatto lesivo (recesso) possono derivare lesioni non suscettibili di trovare piena tutela sul piano contrattuale dei rimedi all'inadempimento: ne consegue che la disciplina extracontrattuale, a protezione di diritti primari tutelati di riflesso rispetto a quello inerente alla semplice posizione contrattuale, potrebbe trovare applicazione, qualora non si prospetti un utile e pieno ristoro dei danni attraverso l'azione contrattuale. Data quindi per pacifica la possibilità di lesione di un assetto complesso di interessi giuridicamente protetti, ecco allora che, qualora la violazione configuri al contempo anche la lesione di un diritto assoluto, e sussistano dunque le condizioni per chiedere entrambe le tutele in quanto vi sia un comportamento colposo o doloso che leda situazioni giuridiche soggettive assolute, potrebbe trovare autonomo spazio l'azione aquiliana, siccome idonea a garantire una tutela specifica della posizione del soggetto bisognoso di tutela, in considerazione della specificità dei diritti coinvolti». Sull'argomento, v. S. Thobani, *L'esclusione da Facebook tra lesione della libertà di espressione e diniego di accesso al mercato*, in *Pers. Merc.*, 2021, 426 ss.

166 Tribunale di Varese, sez. I civile, ord. 2 agosto 2022, con nota di M. Gozzi, *Internet e "standard della community" nella regolamentazione delle piattaforme online: sulla non vessatorietà delle clausole in materia di disinformazione*, in questa *Rivista*, 3, 2022, 308 ss.

167 Si considerino, inoltre, le implicazioni in punto di certezza del diritto, segnalate da M. Bassini, *Internet e libertà di espressione*, cit., 95 ss.: «Se il bilanciamento riposa interamente nelle mani dei giudici, infatti, il rischio di imprevedibilità dei rimedi di volta in volta adottati si fa largo, accrescendo il margine di discrezionalità e comprimendo simmetricamente quello di certezza del diritto. Autorevole dottrina va da tempo ammonendo sul rischio che l'attività giurisdizionale divenga il vero e proprio fulcro di un'opera di bilanciamento che dovrebbe invece riposare nelle mani del legislatore e che le corti dovrebbero, semmai, limitarsi a validare, declinandola in base alle specificità del caso concreto».

latoria da parte del legislatore europeo, che, come già rilevato dalla dottrina[168], con il *Digital Services Act* ha rinunciato ad indirizzare la definizione delle *policies* adottate dalle piattaforme[169], privilegiando la regolazione della *content moderation* sul versante procedurale (artt. 16-17-18-20)[170].

Ne consegue che la proporzionalità degli *standard* definiti per l'individuazione dei contenuti stigmatizzabili e delle restrizioni applicabili, pur suscettibile di controllo giurisdizionale - o extragiudiziario (attraverso il sistema interno di gestione dei reclami o la contestazione dinanzi ad un organismo di risoluzione extragiudiziale) - non è perseguita attraverso la definizione di un set di regole, o quantomeno di principi, operanti *ex ante*.

Nondimeno, il deficit di "legalità sostanziale" potrebbe risultare in parte compensato dal coinvolgimento degli utenti nella definizione degli *standard* applicati dalle piattaforme, nell'ottica di un apporto collaborativo funzionale alla più efficace tutela dei diritti fondamentali.

D'altra parte, i vantaggi della dimensione partecipativa sono stati già sperimentati con l'elaborazione del Codice per il contrasto alla disinformazione del 2022, tra i cui aspetti di maggior pregio si è evidenziata proprio «la varietà e diversità di provenienza dei firmatari che non sono più soltanto le grandi piattaforme, ma comprendono anche esponenti della società civile, della comunità dei *fact-checkers* e delle imprese pubblicitarie, il che evidentemente ha portato spesso ad un aspro ma molto sano contraddittorio»[171].

Un varco in tal senso è aperto dall'art. 45 del Regolamento, che contempla il coinvolgimento delle organizzazioni della società civile e di tutte le altre parti interessate nella elaborazione dei codici di condotta contenenti misure di attenuazione dei "rischi sistemici significativi"[172], tra cui, per quanto qui rileva, quelli incidenti sulla libertà di

---

168 M. Betzu, *Poteri pubblici e poteri private*, cit., 180; E. Birritteri, *Contrasto alla disinformazione*, cit., 57; I. De Vivo, *il potere d'opinione delle piattaforme online: quale ruolo del "regulatory turn" europeo nell'oligopolio informativo digitale?, Federalismi.it*, 2, 2024, 59; O. Pollicino, *The quadrangular shape of the geometry of digital power(s) and the move towards a procedural digital constitutionalism*, in *European Law Journal*, 29(1-2), 2023.

169 R. Niro, *Piattaforme digitali e libertà di espressione fra autoregolamentazione e coregolazione*, cit., 1390; G. De Minico, *Fundamental Rights European digital regulation and algorithmic chal-lenge*, in *Media Laws. Rivista di diritto dei media*, 1, 2021.

170 In senso critico E. Birritteri, *Contrasto alla disinformazione* cit., 57 osserva: «Da un lato, invero, introdurre specifiche procedure di dettaglio sul piano della moderazione dei contenuti e dei reclami, valide per qualsiasi prestatore di servizi intermediari a prescindere dallo specifico mercato di riferimento, dal tipo di attività, dalla dimensione, secondo un modello *one size fits all*, sarebbe stato molto rischioso e, forse, controproducente, con il rischio di imporre oneri eccessivamente gravosi e non necessari […]. Dall'altro lato, però, pur senza legittimare inutili irrigidimenti burocratici, sarebbe stato a nostro avviso utile aggiungere qualche specificazione in più in merito ai diritti di garanzia minimali dell'utente sul piano delle misure che la piattaforma può disciplinare e adottare incidendo sui suoi diritti fondamentali (su tutti, dalla nostra prospettiva, la libertà di espressione)». Secondo O. Pollicino, *Asimmetrie*, cit., 241: «richiedendo che la moderazione dei contenuti da parte dei *social network* aderisca allo standard "pubblico" si rischierebbe di svuotare completamente la promessa di libertà insita nell'avvento di queste piattaforme, mettendo a rischio il loro possibile uso come strumento di protesta o di contrasto alla propaganda o alla censura pubblica. La mano dei *social* sarebbe così guidata dai poteri degli Stati, magari orientati anche da finalità virtuose, ma forse in alcuni casi animati da propositi meno commendevoli, specialmente nell'ambito di contesti di democrazie immature o di natura bellica come quelli tristemente noti».

171 O. Pollicino, *Asimmetrie valoriali*, cit., 234.

172 Sull'argomento v. N. Maccabiani, *Co-regolamentazione, nuove tecnologie e diritti fondamentali: questioni di*

espressione e di informazione (ad es. le misure riguardanti le condizioni generali applicate dalle piattaforme e i sistemi di moderazione dei contenuti).
Restando sul piano delle garanzie operanti *ex ante*, un'ultima notazione deve essere svolta con riguardo all'implementazione del principio di trasparenza, la cui importanza si rinviene negli svariati richiami operati dal Regolamento.
A tal proposito, non si può trascurare come la finalità garantista sottesa all'applicazione del principio rischi di essere vanificata allorché, come sovente accade, le piattaforme facciano ricorso a sistemi di AI nell'attività di selezione dei contenuti ammessi.
Ed invero, sebbene il Regolamento imponga di rendere noto agli utenti - con un linguaggio chiaro, semplice e comprensibile - il processo decisionale algoritmico utilizzato nell'attività di moderazione, si tratta di una prescrizione insuscettibile di effettiva attuazione ogniqualvolta il filtraggio dei contenuti, ad esempio di quelli disinformativi, avvenga mediante sistemi di AI difficilmente comprensibili per l'utente medio.
Infine, anche sul versante delle forme di tutela *ex post*, sembra di poter cogliere talune criticità.
Segnatamente, con riferimento al sistema interno di gestione dei reclami, emergono le perplessità già espresse dalla dottrina con riguardo a forme di "giustizia privata" di cui l'*Oversight Board* di Facebook[173] - seppure connotato da peculiari caratteristiche di autoproclamata indipendenza - è paradigmatico esempio.
Invero, ferma restando la possibilità di attivare la tutela giurisdizionale, siffatti strumenti alternativi di risoluzione delle controversie piattaforma-utente - pur utili alla deflazione del contenzioso giurisdizionale, nonché ad una rapida e meno dispendiosa risoluzione della controversia - possono costituire un'effettiva forma di tutela dei diritti fondamentali solo a condizione che siano codificate garanzie di "giusto procedimento"[174] e requisiti di indipendenza, che invece difettano nelle previsioni del DSA.
In compenso, un più elevato livello di garanzie può essere colto nella disciplina degli organismi di risoluzione extragiudiziale delle controversie di cui all'art. 21, per i quali, pur non essendo introdotta una puntuale regolazione del procedimento, si contemplano requisiti di imparzialità e indipendenza (anche sul piano finanziario) e - aspetto ancor più significativo - si rinvia al diritto degli Stati membri sotto il profilo delle regole procedurali applicabili.

5. Il contrasto alle "manipolazioni algoritmiche"

Con riguardo al secondo versante della regolazione, il legislatore europeo muove da una chiara consapevolezza: «Quando ai destinatari del servizio vengono presentate inserzioni pubblicitarie basate su tecniche di *targeting* ottimizzate per soddisfare i loro interessi e potenzialmente attirare le loro vulnerabilità, ciò può avere effetti negativi particolarmente gravi. In alcuni casi, le tecniche di manipolazione possono avere un

forma e di sostanza, in *Osservatorio sulle fonti*, 3, 2022, 79.

173 A. Iannotti della Valle, *La giurisdizione privata nel mondo digitale al tempo della crisi della sovranità: il modello dell'Oversight Board di Facebook*, in *Federalismi.it*, 26, 2021, 166.

174 Per una critica in tal senso, v. E. Birritteri, *Contrasto alla disinformazione* cit., 65.

impatto negativo su interi gruppi e amplificare i danni per la società, ad esempio contribuendo a campagne di disinformazione o discriminando determinati gruppi. Le piattaforme online sono ambienti particolarmente sensibili per tali pratiche e presentano un rischio per la società […]»[175].

Tale premessa spiega perché numerose previsioni del DSA siano volte a contrastare l'impatto delle "manipolazioni algoritmiche", prevedendo, ad esempio, che la diffusione di *fake news*[176] sia uno degli ambiti di intervento dei codici di condotta *ex* art. 45 o, ancora, che ai fornitori di piattaforme online sia preclusa la possibilità di progettare, organizzare o gestire le interfacce online «in modo tale da ingannare o manipolare i destinatari dei loro servizi o da materialmente falsare o compromettere altrimenti la capacità dei destinatari dei loro servizi di prendere decisioni libere e informate» (art. 25)[177]. Coerentemente con la necessità di avvalersi dell'*expertise* dei regolati, la norma individua con chiarezza il fine perseguito dal divieto - riassumibile nella tutela della libertà di informazione e di autodeterminazione – ma rinuncia a definire nel dettaglio il novero delle pratiche vietate, inevitabilmente condizionato dall'incessante sviluppo tecnologico.

Stante l'impossibilità di un capillare controllo pubblico *ex ante*, il medesimo approccio trova accoglimento laddove si rimette dalla discrezionalità delle piattaforme e dei motori di ricerca (solo quelli di dimensioni molto grandi) la valutazione relativa all'impatto dei sistemi algoritmici sui diritti fondamentali, in particolare – per quanto qui rileva - sulla libertà di espressione e di informazione, sul pluralismo informativo, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica, vulnerabile con la diffusione di *fake news* ("valutazione dei rischi" *ex* art. 34)[178].

Un maggior grado di dettaglio si rinviene nell'individuazione delle possibili "misure di attenuazione dei rischi" (art. 35), tra cui, oltre alla sperimentazione e all'adeguamento dei sistemi algoritmici, spicca, per il contrasto al *deepfake*, «il ricorso a un contrassegno ben visibile per fare in modo che un elemento di un'informazione, sia esso un'immagine, un contenuto audio o video, generati o manipolati, che assomigli notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che a una persona appaia falsamente autentico o veritiero, sia distinguibile quando è presentato sulle interfacce online e, inoltre, la fornitura di una funzionalità di facile utilizzo che consenta ai destinatari del servizio di indicare tale informazione».

Nondimeno, nell'ottica della co-regolazione, resta salva la discrezionalità delle piattaforme sia nella scelta tra le misure indicate, il cui elenco è meramente esemplificativo, sia nella possibilità di individuarne di ulteriori, purché «ragionevoli, proporzionate ed efficaci, adattate ai rischi sistemici specifici individuati […] prestando particolare atten-

[175] V. considerando 69.

[176] V. considerando 104.

[177] Ai sensi dell'art. 25, par. 3: «La Commissione può emanare orientamenti sull'applicazione del paragrafo 1 con riguardo a pratiche specifiche, in particolare: a) attribuire maggiore rilevanza visiva ad alcune scelte quando si richiede al destinatario del servizio di prendere una decisione; b) chiedere ripetutamente che un destinatario del servizio effettui una scelta laddove tale scelta sia già stata fatta, specialmente presentando pop-up che interferiscano con l'esperienza dell'utente; c) rendere la procedura di disdetta di un servizio più difficile della sottoscrizione dello stesso». V. considerando 67.

[178] V. considerando 84 e considerando 88.

zione agli effetti di tali misure sui diritti fondamentali»[179].

Un'ultima osservazione deve essere svolta con riguardo ai sistemi di raccomandazione, per tali intendendosi i sistemi interamente o parzialmente automatizzati che una piattaforma utilizza al fine di suggerire informazioni specifiche tramite la propria interfaccia online o mettere in ordine di priorità determinate informazioni o determinare in altro modo l'ordine o l'importanza delle informazioni visualizzate[180].

Segnatamente, è interessante rilevare come, rispetto ai sistemi di raccomandazione, la tutela dei diritti degli utenti venga garantita principalmente – ma non solo - attraverso l'implementazione del principio di trasparenza, in forza del quale le piattaforme sono tenute a indicare - in modo chiaro e intellegibile - nelle condizioni generali di contratto: (i) i principali parametri utilizzati, affinché se ne possano desumere il motivo per cui alcune informazioni vengono suggerite in via prioritara, i criteri più significativi con cui vengono determinate le informazioni suggerite e le ragioni dell'importanza dei parametri scelti[181], nonché (ii) la presenza di opzioni che consentano di modificare o influenzare i parametri utilizzati[182] (in tal caso, dovrà anche essere resa disponibile una funzionalità utile a selezionare e modificare facilmente l'opzione preferita)[183].

Inoltre, le piattaforme e i motori di ricerca di dimensioni molto grandi devono assicurare che almeno un'opzione per ciascuno dei loro sistemi di raccomandazione non sia basata sulla profilazione, per tale intendendosi, ai sensi dell'art. 4, punto 4) del GDPR, «qualsiasi forma di trattamento automatizzato di dati personali[184] consistente nell'u-

---

179 V. art. 35 ("Attenuazione dei rischi").

180 Art. 3, lett. s). A tal proposito, v. il considerando 70: «Un elemento essenziale dell'attività di una piattaforma online consiste nel modo in cui le informazioni sono messe in ordine di priorità e presentate nella sua interfaccia online per facilitare e ottimizzare l'accesso alle stesse da parte dei destinatari del servizio. Ciò avviene, ad esempio, suggerendo, classificando e mettendo in ordine di priorità le informazioni in base ad algoritmi, distinguendole attraverso testo o altre rappresentazioni visive oppure selezionando in altro modo le informazioni fornite dai destinatari. Tali sistemi di raccomandazione possono avere un impatto significativo sulla capacità dei destinatari di recuperare e interagire con le informazioni online, anche per facilitare la ricerca di informazioni pertinenti per i destinatari del servizio e contribuire a migliorare l'esperienza dell'utente. Essi svolgono inoltre un ruolo importante nell'amplificazione di determinati messaggi, nella diffusione virale delle informazioni e nella sollecitazione del comportamento online. Di conseguenza, le piattaforme online dovrebbero provvedere in modo coerente affinché i destinatari del loro servizio siano adeguatamente informati del modo in cui i sistemi di raccomandazione incidono sulle modalità di visualizzazione delle informazioni e possono influenzare il modo in cui le informazioni sono presentate loro. Esse dovrebbero indicare chiaramente i parametri di tali sistemi di raccomandazione in modo facilmente comprensibile per far sì che i destinatari del servizio comprendano la modalità con cui le informazioni loro presentate vengono messe in ordine di priorità. Tali parametri dovrebbero includere almeno i criteri più importanti per determinare le informazioni suggerite al destinatario del servizio e i motivi della rispettiva importanza, anche nel caso in cui le informazioni».

181 Art. 27 ("Trasparenza dei sistemi di raccomandazione"), par. 2.

182 Art. 27, par. 1.

183 Art. 27, par. 3.

184 Alla stregua delle "Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679", elaborate dal Gruppo di lavoro Articolo 29, «La profilazione è costituita da tre elementi: deve essere una forma di trattamento automatizzato; deve essere effettuata su dati personali; il suo obiettivo deve essere quello di valutare aspetti personali relativi a una persona fisica. L'articolo 4, punto 4, fa riferimento a "qualsiasi forma di trattamento automatizzato" e non al trattamento "unicamente" automatizzato (di cui all'articolo 22). La profilazione deve implicare una qualche forma di trattamento automatizzato, sebbene il coinvolgimento umano non

tilizzo di tali dati per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti»[185].

L'utilizzo dell'AI per l'adeguamento dei servizi digitali alla preferenze degli utenti[186] fa sorgere la necessità di coordinare la disciplina del *Digital Services Act* con l'*Artificial Intelligence Act* (Regolamento europeo UE 1689/2024)[187], che, non a caso, definisce il proprio perimetro applicativo escludendone l'incompatibilità con le previsioni del DSA (art. 2 par. 5) e prevedendo che gli obblighi introdotti per gli operatori che immettono sul mercato o utilizzano sistemi di IA integrino quelli imposti dal DSA, nonché il

comporti necessariamente l'esclusione dell'attività dalla definizione. La profilazione è una procedura che può implicare una serie di deduzioni statistiche. Spesso viene impiegata per effettuare previsioni su persone usando dati provenienti da varie fonti per dedurre qualcosa su una persona in base alle qualità di altre persone che sembrano statisticamente simili. Il regolamento afferma che la profilazione è il trattamento automatizzato di dati personali per valutare determinati aspetti personali, in particolare per analizzare o prevedere aspetti riguardanti persone fisiche. L'uso del verbo "valutare" suggerisce che la profilazione implichi una qualche forma di valutazione o giudizio in merito a una persona. […]In generale, la profilazione consiste nella raccolta di informazioni su una persona (o un gruppo di persone) e nella valutazione delle loro caratteristiche o dei loro modelli di comportamento al fine di includerli in una determinata categoria o gruppo, in particolare per analizzare e/o fare previsioni, ad esempio, in merito a: capacità di eseguire un compito, interessi, o comportamento probabile. […] Il processo decisionale automatizzato ha una portata diversa da quella della profilazione, a cui può sovrapporsi parzialmente o da cui può derivare. Il processo decisionale esclusivamente automatizzato consiste nella capacità di prendere decisioni impiegando mezzi tecnologici senza coinvolgimento umano. […]Le decisioni automatizzate possono essere prese ricorrendo o meno alla profilazione, la quale a sua volta può essere svolta senza che vengano prese decisioni automatizzate. Tuttavia, la profilazione e il processo decisionale automatizzato non sono necessariamente attività separate. Qualcosa che inizia come un semplice processo decisionale automatizzato potrebbe diventare un processo basato sulla profilazione, a seconda delle modalità di utilizzo dei dati».

[185] Ai sensi dell'art. 22 GDPR ("Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione"), **«**1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona.
2. Il paragrafo 1 non si applica nel caso in cui la decisione:
a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento;
b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato;
c) si basi sul consenso esplicito dell'interessato.
3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione.
4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato».

[186] «Ad esempio, i sistemi di IA possono essere utilizzati per fornire motori di ricerca online, in particolare nella misura in cui un sistema di IA, come un *chatbot* online, effettua ricerche, in linea di principio, su tutti i siti web, incorpora i risultati nelle sue conoscenze esistenti e si avvale delle conoscenze aggiornate per generare un unico *output* che combina diverse fonti di informazione» (considerando 119).

[187] Che definisce "sistema di IA" «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali» (art. 3).

quadro di gestione dei rischi cui quest'ultimo fa soggiacere le piattaforme e i motori di ricerca di dimensioni molto grandi[188].

Nondimeno, dalle previsioni del Regolamento IA emergono non solo i profili di sovrapposizione e complementarietà al DSA sotto il profilo dell'ambito applicativo, ma anche la chiara consapevolezza – altrettanto evidente nel DSA - che la diffusione di contenuti generati o manipolati artificialmente possa alimentare il fenomeno della disinformazione[189], sino ad incidere sul corretto svolgimento dei processi democratici; sicché, il rapporto tra i due corpi normativi viene delineato anche in termini di reciproca funzionalità.

È quanto pare di poter cogliere laddove si dispone che l'obbligo, previsto per i fornitori e i *deployer*[190] di taluni sistemi di IA, di rilevare e rendere noto se gli *output* di tali sistemi siano generati o manipolati artificialmente, è particolarmente importante per l'efficace attuazione del DSA, specie per l'adempimento da parte delle piattaforme e dei motori di ricerca di dimensioni molto grandi dell'obbligo di individuare e attenuare i "rischi sistemici" conseguenti alla diffusione di contenuti generati o manipolati artificialmente, soprattutto il rischio di impatti negativi - effettivi o prevedibili - sui processi democratici, sul dibattito pubblico e sui processi elettorali[191].

Infine, la necessità di una lettura sistematica e coordinata dei due regolamenti è stata evidenziata dalla dottrina secondo cui «non sarebbe coerente che i rischi inaccettabili ai sensi dell'*Artificial Intelligence Act* fossero accettabili per il *Digital Services Act* o qualificabili come rischi sistemici», atteso che – lo si è visto - «questi ultimi, ai sensi del *Digital Services Act* non sono vietati ma, piuttosto, considerati inevitabili, anche se da attenuare ai sensi dell'art. 35»; pertanto - nonostante l'assenza di un divieto espresso nel DSA - dovrebbe ritenersi che ai prestatori di servizi digitali sia precluso l'utilizzo di sistemi

[188] Il Regolamento IA «disciplina i sistemi di IA e i modelli di IA imponendo determinati requisiti e obblighi agli operatori del mercato pertinenti che li immettono sul mercato, li mettono in servizio o li utilizzano nell'Unione, integrando in tal modo gli obblighi per i prestatori di servizi intermediari che incorporano tali sistemi o modelli nei loro servizi disciplinati dal regolamento (UE) 2022/2065. Nella misura in cui sono integrati in piattaforme online di dimensioni molto grandi designate o motori di ricerca online di dimensioni molto grandi designati, tali sistemi o modelli sono soggetti al quadro di gestione dei rischi di cui al regolamento (UE) 2022/2065. […] In tale quadro, le piattaforme di dimensioni molto grandi e i motori di ricerca di dimensioni molto grandi sono tenuti a valutare i potenziali rischi sistemici derivanti dalla progettazione, dal funzionamento e dall'utilizzo dei propri servizi - compresi quelli derivanti dalle modalità di progettazione dei sistemi algoritmici impiegati nel servizio o da potenziali usi impropri – e ad adottare misure di attenuazione adeguate per la tutela dei diritti fondamentali» (considerando 118).

[189] In tal senso, è significativo il considerando 110: «I modelli di IA per finalità generali potrebbero comportare rischi sistemici che includono, tra l'altro, qualsiasi effetto negativo effettivo o ragionevolmente prevedibile in relazione a incidenti gravi, perturbazioni di settori critici e serie conseguenze per la salute e la sicurezza pubbliche; eventuali effetti negativi, effettivi o ragionevolmente prevedibili, sui processi democratici e sulla sicurezza pubblica ed economica; la diffusione di contenuti illegali, mendaci o discriminatori. […] In particolare, gli approcci internazionali hanno finora rilevato la necessità di prestare attenzione – tra i vari rischi – all'agevolazione della disinformazione».

[190] Per "*deployer*" si intende «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale» (art. 3 n. 4).

[191] V. considerando 136.

di intelligenza artificiale vietati dal Regolamento IA[192], ad esempio quelli «aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo» (art. 5 Reg. UE 1689/2024).

6. Conclusioni

Sino a qui l'analisi è stata condotta nella prospettiva offerta dall'oggetto dell'indagine; nondimeno i fenomeni suscettibili di incidere sulla libertà di informazione involgono anche altri diritti fondamentali, cui è opportuno riservare almeno qualche breve considerazione conclusiva.

Tra i diritti implicati spicca quello alla protezione dei dati personali, che da sempre coinvolto nell'evoluzione tecnologica quale ragione di nuove esigenze di tutela, non può non essere condizionato dall'affermazione di modelli di *business* basati sull'utilizzo dell'IA per l'elaborazione e lo sfruttamento dei dati degli utenti.

Ed invero, se in un primo tempo lo sviluppo tecnologico ha segnato il passaggio dal diritto a non subire interferenze esterne (cd. *right to be let alone*) al diritto di controllo sui propri dati, l'affermarsi dell'economia *data driven* rende oggi necessaria una nuova e più ampia accezione del diritto in parola, eccedente la sfera individuale, posto che – come è stato detto - «le scelte rese possibili grazie ai dati acquisiti individualmente, avvengono non solo sul singolo, ma sull'intero gruppo di cui fa parte […] la lesione dei diritti non si limita al singolo, ma mina la vita collettiva e i diritti di tutti nel lungo periodo»[193].

Sicché, il "nuovo" diritto alla protezione dei dati personali potrebbe essere inteso anche come diritto a non subire le "manipolazioni algoritmiche" rese possibili dall'elaborazione dei dati[194] attraverso i sistemi di IA.

Siffatta configurazione fornirebbe ulteriori garanzie di tutela ai diritti fondamentali – *in prims* il diritto all'identità personale, di cui a breve si tratterà – e, al contempo, rivestirebbe un ruolo determinante nella salvaguardia dei valori democratici sottesi al no-

[192] S. Tommasi, *Digital Services Act e Artificial Antelligence Act: tentativi di futuro da armonizzare*, in *Persona e Mercato*, 2, 2023, 285.

[193] F. Faini, *Data society. Governo dei dati e tutela del diritto nell'era digitale*, Milano, 2019, 410 ss.

[194] Si rammenta che il GDPR delinea un'ampia definizione di "dato personale" laddove afferma che: «si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale» (art. 4). Sicché, è da condividere l'osservazione secondo cui «date le capacità analitiche in continuo sviluppo, è probabile che quasi tutti i tipi di dati possono essere elaborati insieme ad altri dati ed essere qualificati come dati personali. Infatti, i dati personali possono essere resi anonimi, diventando così dati non personali; al contrario, i dati non personali possono essere integrati con altri set di dati più complessi, consentendo così l'identificazione indiretta delle persone» (in termini, F. Colaprisco, *Data Governance Act. Condivisione e "altrusimo" dei dati*, in *Focus "Servizi e piattaforme digitali"* AISDUE, 3, 2021, 64).

stro ordinamento costituzionale, atteso che – lo si è visto - le prospettive aperte dallo sviluppo dell'economia *data driven* sono suscettibili di pregiudicare anche la libertà di informazione, la partecipazione alla vita collettiva, l'evoluzione della società nella sua interezza, dimostrando, oggi più che in passato, quanto la tutela del diritto alla protezione dei dati personali possa condizionare quella di altri diritti fondamentali eccedenti la sfera individuale[195].

Ancora, dallo scenario delineato affiora la necessità di riconsiderare il contenuto del diritto all'identità personale[196], che, coniato dalla giurisprudenza agli inizi degli anni '70, sembra oggi arricchirsi di nuovi significati.

Invero, secondo l'impostazione ormai consolidata, il diritto in parola consiste ne «l'interesse di ogni persona a non vedere travisato o alterato all'esterno il proprio patrimonio intellettuale, politico, sociale, religioso, professionale, a causa dell'attribuzione di idee, opinioni o comportamenti differenti da quelli che l'interessato ritenga propri e abbia manifestato nella vita di relazione»[197]; in altri termini, siffatto diritto «assicura la fedele rappresentazione alla propria proiezione sociale»[198].

Muovendo da tale assunto, non v'è motivo di dubitare che la lesione possa aver luogo anche nell'ambiente digitale, laddove – lo si è visto - l'identità di ciascuno è definita attraverso l'elaborazione algoritmica dei dati immessi in Rete (*digital person*[199]).

Ed infatti, guardando al momento costitutivo dell'identità personale, è del tutto evidente come la profilazione basata su modelli di analisi sempre più pervasivi delinei «un'identità non consapevole»[200], un'«identità catturata»[201] all'insaputa dell'utente del web, tracciata sulla base dei dati più utili e funzionali alla definizione del profilo.

---

195 D'altre parte, in dottrina è già emerso che «il diritto alla riservatezza presenta un tratto identificante che altri diritti fondamentali non hanno: ha una sua propria autonomia, concettuale e positiva, grazie alla quale reclama tutela in sé e per sé; è, però, anche un diritto presupposto, diciamo pure un *implied Right*, la sua tutela risultando, per questo verso, indiretta, attraverso quella offerta ad altri diritti. È vero, però, anche l'inverso, vale a dire che tutelando la riservatezza si ha, di riflesso, anche quella di altri diritti, ad essa funzionalmente connessi. […] Per un certo verso, questo è vero per ogni diritto, sol che si ammetta che ciascuno di essi fa "sistema" coi restanti […]. È pur vero, tuttavia, che il nesso che lega i diritti non si presenta allo stesso modo o con la stessa intensità nel passaggio dall'uno all'altro dei diritti stessi. Per la riservatezza, invece, ciò si riscontra in una misura tale da rendersi, a conti fatti, indistinguibile ciò che specificamente la riguarda da ciò che è invece proprio di altri diritti cui la stessa risulta strettamente legata» (A. Ruggeri, *Dignità dell'uomo, diritto alla riservatezza, strumenti di tutela (prime notazioni*, in *Consulta OnLine*, 371-372).

196 In merito al fondamento costituzionale del diritto all'identità personale, S. Fois, *Questioni sul fondamento costituzionale del diritto all' "identità personale"*, in G. Alpa-M. Bessone-L. Boneschi-G. Caiazza (a cura di), *L'informazione e i diritti della persona*, Napoli, 1983, 155 ss.; V. Zeno-Zencovich, *Identità personale*, in *Dig. disc. priv*., IX, Torino, 1995; G. Finocchiaro, *Identità personale* (*diritto alla*), in *Dig. disc. priv.*, Agg., Torino, 2010; id*, Il diritto all'identità personale su Internet*, in *Dir. inf*., 3, 2012, 383 ss.

197 G. Pino, *Il diritto all'identità personale ieri e oggi. Informazioni, mercato, dati personali*, in R. Panetta (a cura di), *Libera circolazione e protezione dei dati personali*, Milano, 2006, 260. In giurisprudenza v. Cass. civ., sez. I, 22 giugno 1985 n. 3769, in *Foro it*., I, 1985, 2211; Corte Cost., 3 febbraio 1994 n. 13, in *Foro it*., I 1994, 668.

198 Cass. civ., sez. I, 22 giugno 1985 n. 3769.

199 G. Alpa, *L'identità digitale e la tutela della persona. Spunti di riflessione*, in *Contratto e impresa*, 3, 2017, 725-726.

200 M. Bianca, *La filter bubble e il problema dell'identità digitale*, in questa *Rivista*, 2, 2019, 12.

201 S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2012, 305.

Emerge, così, oltre all'eterodeterminazione, già di per sé problematica, anche l'inevitabile parziarietà di un'identità conformata dalle finalità perseguite con la profilazione; come è stato detto, «le classificazioni dei dati e soprattutto la loro connessione ricostruisce un'identità che in parte combacia con quella reale e in parte la deforma, la ingigantisce o la deprime, a seconda degli angoli visuali […] in cui la persona è stata scomposta»[202].

Anche volgendo l'attenzione al carattere dinamico-evolutivo dell'identità personale, la lesione del diritto può essere colta con chiarezza se solo si considera che – come rilevato nelle premesse - la necessità di rendere più sicura la predizione dei comportamenti e delle scelte degli utenti induce le piattaforme a determinarli; si tratta di «interventi pensati per aumentare la certezza che le cose vengano fatte: suggeriscono, spingono, dirigono, manipolano e modificano comportamenti verso direzioni specifiche, per mezzo di azioni impercettibili come inserire una determinata frase nel nostro feed di Facebook […]»[203].

E così, in una dinamica di continua e sempre maggiore compenetrazione tra realtà e spazio virtuale, l'identità digitale arbitrariamente delineata dagli algoritmi viene sfruttata per plasmare e manipolare l'identità che si manifesta nelle scelte, nei comportamenti e nelle preferenze quotidianamente espresse anche al di fuori del web, con l'ulteriore evidente lesione della libertà di autodeterminazione[204].

Per giunta, l'inquadramento degli utenti in classificazioni incapaci di adeguarsi alla loro

---

[202] G. Alpa, *L'identità digitale,* cit., 725-726. V. anche G. Resta*, Identità personale e identità digitale*, in *Dir. inf*., 2, 2007, 511 ss.: «Le tecniche di raccolta dei dati e profilazione individuale, rese possibili dalle nuove tecnologie, determinano il rischio che l'io venga frammentato, a sua insaputa, in una molteplicità di banche dati, offrendo così una raffigurazione parziale e potenzialmente pregiudizievole della persona, la quale verrebbe così ridotta alla mera sommatoria delle sue proiezioni elettroniche. Il diritto all'identità, di riflesso, assume nuove connotazioni, in quanto implica non più soltanto la "corretta rappresentazione in ciascun contesto", ma presuppone una "rappresentazione integrale della persona" e per di più una "rappresentazione non affidata solo agli strumenti automatizzati"». In merito all'incidenza sul diritto all'identità personale, è utile richiamare le parole di S. Rodotà, *Persona, riservatezza, identità. Prime note sistematiche sulla protezione dei dati personali*, in *Riv. crit. dir. priv*., 1997, 605, secondo cui: «L'unità della persona viene spezzata. Al suo posto troviamo tante "persone elettroniche", tante persone create dal mercato quanti sono gli interessi che spingono alla raccolta delle informazioni. Stiamo diventando "astrazioni nel cyberspazio"; siamo di fronte ad un individuo "moltiplicato", non per la sua scelta di assumere molteplici identità, ma per ridurlo alla misura delle dimensioni di mercato». L'impatto sull'identità personale è rilevata anche da O. Sesso Sarti, *Profilazione e trattamento dei dati personali*, L. Califano-C. Colapietro (a cura di), *Innovazione tecnologica* e *valore* della *persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, 2017, 574, nonché da G. Ziccardi, *Profilazione dell'individuo, Big Data, e metadati: comprendere le tecnologie attuali per comprendere i contenuti d'odio online*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 51, che, richiamandosi al pensiero di Rodotà, osserva: «L'essere umano è, oggi, in competizione con le macchine che trattano i suoi dati. Macchine che riescono a recuperare informazioni, e a elaborare dati, che sono in grado di disegnare, attorno all'individuo, un profilo (o "corpo elettronico", come scriverebbe Stefano Rodotà) ancora più preciso di come l'essere umano conosca se stesso, e capace persino di prevedere e, perché no, di orientare i comportamenti. A causa di una costante produzione di dati che riguardano l'individuo ci si trova, ha sempre sostenuto Rodotà, a operare nella società digitale con una sorta di doppio corpo. Vi è, infatti, la possibilità di controllo non solo del lato fisico della persona ma anche di quell'insieme di dati che, giorno dopo giorno, si popolano, aumentano e costituiscono, in un certo senso, una seconda persona: un lato-ombra perfettamente visibile e, soprattutto, altrettanto importante».

[203] S. Zuboff, *Il capitalismo della sorveglianza*, Roma, 2019, 126.

[204] S. Rodotà, *Tecnologia e diritti*, Bologna, 1995, 115.

più intima evoluzione personale[205], oltre a vulnerare il pieno e libero sviluppo dell'identità[206], può persino esporli a trattamenti discriminatori[207] ed anche determinarne l'esclusione sociale[208]; come osservato dall'Agenzia europea per i diritti fondamentali, «*discrimination is a crucial topic when it comes to the use of AI, because the very purpose of machine learning algorithms is to categorise, classify and separate*»[209].

A conferma di ciò, basti pensare che il *data mining*, quale processo automatizzato di scoperta di schemi utili a disvelare relazioni statistiche nell'ambito del set di dati addotti a fondamento del calcolo[210] (dati di addestramento"), è ontologicamente una forma di discriminazione statistica, in quanto dati di addestramento distorti (ad esempio perché riflettono un pregiudizio o per la sottorappresentazione o sovrarappresentazione di un campione di popolazione) inevitabilmente conducono a modelli discriminatori[211]. Sicché – lo si è visto – anche l'utilizzo dell'IA per la moderazione dei contenuti disponibili in Rete, potendo riprodurre i cosiddetti *technical bias*, è suscettibile di produrre discriminazioni a danno delle minoranze.

Dunque, i fenomeni individuati nelle premesse introduttive quali rischi specifici per la libertà di informazione e per i principi costituzionali ad essa correlati, *in primis* il principio democratico, rivelano anche una tensione profonda, sebbene forse meno evidente, con il principio (super)costituzionale[212] della dignità umana, implicante - al pari del

---

[205] M. Bianca, La *filter bubble*, cit., 11, a proposito della «connotazione dinamica» dell'identità personale, parla dell' «interesse all'attualizzazione nel tempo della propria identità».

[206] R. De Meo, *Autodeterminazione e consenso*, cit., 587 ss.

[207] M. Falcone, *Big data e pubbliche amministrazioni* cit. In tal senso, v. anche E. Pellecchia, *Privacy, decisioni automatizzate e algoritmi*, in E. Tosi (a cura di), *Privacy digitale*, Milano, 2019, 422: «Profilazione e decisioni automatizzate possono segregare le persone in specifiche categorie riducendo la loro possibilità di scelta, possono consolidare gli stereotipi, scoraggiare azioni rivelatrici di condotte "divergenti" (es. partecipare a gruppi di discussione su droghe, alcolismo, malattie mentali, sesso, e altri argomenti), produrre discriminazioni inattese (talvolta fondate su caratteristiche non modificabili)».

[208] M. Infantino, *La responsabilità per danni algoritmici: prospettive europeo-continentali*, in *Resp. civ. e prev.*, 5, 2019, 1762 ss.: «Si pensi alle conseguenze negative, sul piano economico, reputazionale e/o emotivo, che possono derivare a una persona dalla circolazione e dal trattamento algoritmico delle informazioni che la riguardano. Un algoritmo le cui istruzioni e/o conclusioni siano discriminatorie nei confronti di certi gruppi può tradursi nell'impossibilità, per chi appartiene o è ritenuto appartenere a quei gruppi, di accedere a certe utilità o servizi, che si tratti dell'ammissione a un colloquio di lavoro, dell'erogazione di un mutuo o dell'accesso a un'assicurazione».

[209] Report *Artificial intelligence and fundamental rights*, 68, reperibile in https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf. In dottrina v. E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale* in *Liber amicorum per Pasquale Costanzo*, *Consulta OnLine*, 2020; P. Zuddas, *Intelligenza artificiale e discriminazioni* in *Liber amicorum per Pasquale Costanzo*, *Consulta OnLine*, 2020; L. Giacomelli, *Big brother is "gendering" you. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale*?, in *BioLaw Journal. Rivista di BioDiritto*, 2, 2019, 269 ss.

[210] E. Pellecchia, *Privacy, decisioni automatizzate e algoritmi* cit., 422: «L'insieme accumulato di relazioni scoperte viene comunemente chiamato "modello" e questi modelli possono essere utilizzati per automatizzare il processo di classificazione di entità o attività di interesse, stimare il valore di variabili non osservate o prevedere risultati futuri».

[211] Sui rischi di discriminazione insiti nei modelli di IA, *ex multis* E. Mantovani, *Intelligenza artificiale e discriminazione: quali prospettive? il modello inglese del data trust*, in *La Rivista Gruppo di Pisa*, 2021, Quaderno n. 3. Fascicolo speciale monografico, 370 ss.

[212] C. Drigo, *La dignità umana quale valore (super)costituzionale*, in L. Mezzetti (a cura di), *Principi costituzionali*, Torino, 2011, 266.

principio personalista[213] di cui è corollario - il divieto di strumentalizzazioni aventi a oggetto la persona[214].

In considerazione di questo e, al contempo, dando per presupposto che il principio personalista sia «storicamente condizionato e necessariamente flessibile»[215], capace di soddisfare esigenze di tutela non ipotizzabili al tempo della Costituente[216], può ben comprendersi quanto il principio in parola, nonché le potenzialità espansive della dignità, «stella polare nella ricerca di nuovi diritti» ma anche «completamento e sostegno dei vecchi diritti»[217], possano rivelarsi utili per la tutela dei diritti fondamentali incisi dallo sviluppo dell'economia digitale.

Ponendosi in questa prospettiva, infatti, può scongiurarsi il rischio che gli spazi lasciati scoperti dalla regolazione europea si traducano in ostacoli alla piena tutela della persona.

Invero, non v'è dubbio che il *Digital Services Act* costituisca un deciso passo avanti per la tutela dei diritti fondamentali, non foss'altro perché con esso si è risposto al bisogno, ormai impellente, di una regolazione pubblicistica volta a contenere l'incidenza dei nuovi – per capacità tecnologica e forza economica - "poteri privati" sui diritti della persona.

Nondimeno, come evidenziato con specifico riguardo ai fenomeni analizzati, il DSA tende a privilegiare la regolazione dei profili procedurali, individuando nella "procedimentalizzazione" e nella trasparenza - intesa come conoscibilità del procedimento decisionale, degli strumenti utilizzati (algoritmi e modelli di IA), nonché dei parametri applicati e, a monte, definiti dai prestatori di servizi (ad es. per l'impostazione dei sistemi di raccomandazione o per l'attività di moderazione dei contenuti disponibili in Rete) - le essenziali garanzie di tutela dei diritti degli utenti. Viceversa – lo si è evinto con immediatezza nell'analisi dei limiti alla *content moderation* – all'esito del bilanciamento tra le istanze – e, per certi versi, l'oggettiva necessità - di auto-regolazione degli operatori economici e l'esigenza di etero-regolazione per la tutela dei diritti, si è riservato alle prime ampio accoglimento sul piano sostanziale, ovvero nella definizione di "ciò che è consentito" agli utenti.

---

213 P. Perlingieri, *Principio personalista, dignità umana e rapporti civili*, in *Annali SISDiC*, 5, 2020, 2.

214 Il «principio personalistico che anima la nostra Costituzione, la quale vede nella persona umana un valore etico in sé, vieta ogni strumentalizzazione della medesima per alcun fine eteronomo ed assorbente» (Cass. civ., sez. I, 16 ottobre 2007, n. 21748, in *Consulta OnLine*).

215 P. Perlingieri, *Principio personalista, dignità umana*, cit., 5. L'A. osserva, infatti, che «il principio personalista nella nostra epoca ha anche – ma non soltanto – il ruolo di porre limiti alla ferrea logica economica, all'aggressività dei mercati e ancor più al terribile incontrollato sviluppo tecnologico che nell'intelligenza artificiale sembra destinato ad avere il suo apice».

216 A. Vedaschi, *Il principio personalista*, in L. Mezzetti (a cura di), *Diritti e doveri*, Torino, 2013, 223; L. Chieffi, *Dignità umana e sviluppi del principio personalista. Brevi note introduttive,* in *Rassegna di diritto pubblico europeo*, 1, 2013, 6: «Grazie ad una progressiva trasformazione del contenuto valoriale della dignità umana, resa possibile dalla duttilità e dinamicità esegetica della normativa costituzionale di riferimento, l'interprete potrà giungere ad una progressiva dilatazione della portata garantistica dei diritti, così da incrementare il grado di tutela in presenza delle profonde innovazioni introdotte nella società nei più svariati settori tecnologici».

217 A. Ruggeri, *La dignità dell'uomo e il diritto di avere diritti (profili problematici e ricostruttivi)*, in *Consulta OnLine*, 398.

Per quanto il principio di trasparenza detenga un'innegabile valenza garantista[218] e, come concordemente ritenuto[219], il "procedimento", di per sé, assurga a garanzia dei diritti nel rapporto potere-libertà, è intuibile che tale approccio possa aprire il varco a un deficit di tutela per tutti i diritti esposti al rischio di lesione conseguente ai due fenomeni esaminati: le "manipolazioni" rese possibili dall'utilizzo di modelli di IA sempre più pervasivi e l'attività di moderazione dei contenuti svolta dalle piattaforme, anche mediante l'IA.

Nondimeno, come anticipato, non è inevitabile che questo accada.

Sul primo fronte - quello dell'incidenza dell'intelligenza artificiale sui diritti fondamentali[220] - un ruolo decisivo può essere svolto dall'applicazione del Regolamento IA.

Invero, sulla scorta di una prima osservazione delle potenzialità insite nelle previsioni dell'*AI Act*, sembra di rilevare che quest'ultimo, più del *Digital Services Act*, possa rivelarsi determinante nel contrasto alle esternalità negative prodotte dallo sfruttamento dell'IA[221].

Nonostante le criticità segnalate da una parte della dottrina[222], induce a tale conclusione un fattore di ordine generale inerente all'impianto regolatorio.

Segnatamente, sulla base di un modello *top-down* molto più marcato di quello seguito

---

[218] Sulla funzione di garanzia sottesa alla trasparenza e all'obbligo di motivazione, v., G. Arena, *Il segreto amministrativo, Profili storici e sistematici*, Milano, 1983; P. Barile, *Democrazia e segreto*, in *Quaderni cost*ituzionali, 1987, 29 ss.; F. Merloni (a cura di), *La trasparenza amministrativa*, Milano, 2008; M. Occhiena, *Pubblicità e trasparenza*, in M. Renna-F. Saitta (a cura di), *Studi sui principi del diritto amministrativo*, Milano, 2012, 141 ss.; F. Manganaro, *L'evoluzione del principio di trasparenza*, in Aa.Vv.., *Scritti in memoria di Roberto Marrama*, Napoli, 2012; F. Lombardi, *La trasparenza tradita*, Napoli, 2021; G. Corso, *Motivazione degli atti amministrativi e legittimazione del potere negli scritti di Antonio Romano Tassone*, in *Diritto amministrativo*, 2014, 470 ss.; G. Arena, *Le diverse finalità della trasparenza amministrativa*, in F. Merloni (a cura di), *La trasparenza amministrativa*, cit., 29 ss.; A. Cassatella, *Il dovere di motivazione nell'attività amministrativa*, Padova, 2013, 249 ss.; G. Mannucci, *Uno, nessuno, centomila. Le motivazioni del provvedimento amministrativo*, in *Diritto pubblico*, 2012, 837 ss. Con particolare riferimento alla questione della "trasparenza algoritmica", v., *ex multis*, A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, in *BioLaw Journal – Rivista di Biodiritto*, 1, 2019, 77 ss; A. Simoncini - S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 2019, 87 ss.; E. Pellecchia, *Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation*, in *Le nuove leggi civili commentate*, 2018, 1210 ss.; P. Forte, *Diritto amministrativo e data science. Appunti di intelligenza amministrativa artificiale (AAI)*, in *P.A.-Persona e Amministrazione*, 1, 2020, 259 ss.; E. Spiller, *Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata*, in *Biolaw Journal– Rivista di Biodiritto*, 2, 2021, 419 ss.

[219] Numerosi spunti di riflessione possono essere tratti dagli studi su procedimento e potere pubblico; senza pretesa di completezza, v. F. Benvenuti, *Funzione amministrativa, procedimento, processo*, in *Rivista trimestrale di diritto pubblico*, 1952, 126 ss.; G. Pastori (a cura di), *La procedura amministrativa*, Milano, 1964; M. Nigro, *Procedimento amministrativo e tutela giurisdizionale contro la pubblica amministrazione (Il problema di una legge generale sul procedimento amministrativo)*, in *Rivista proc. civ.*, 1980, 252 ss.

[220] Sul tema, v. A. Pajno-F. Donati-A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? . Vol. I Diritti fondamentali, dati personali e regolazione*, Bologna, 2022.

[221] In generale, sul Regolamento IA v. F.M. Mancioppi, *La regolamentazione dell'intelligenza artificiale come opzione per la salvaguardia dei valori fondamentali dell'UE*, in *federalismi.it*, 7, 2024, 112 ss.; G. Lemme, *La proposta di regolamento europeo sulla Intelligenza artificiale e la gestione dei rischi: una battaglia che puó essere vinta?*, in *Rivista Trimestrale di Diritto dell'Economia*, 2, 2024, 259 ss.; G. Pesce, *L'intelligenza artificiale alla prova del diritto europeo: verso il diritto della paura?*, in *Amministrativ@mente*, 1, 2024, 14 ss.

[222] G. Smorto, *Distribuzione del rischio e tutela dei diritti nel regolamento europeo sull'intelligenza artificiale. Una riflessione critica*, in *Foro italiano*, 4, 2024, 208 ss.; C. Novelli, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *federalismi.it*, 2, 2024, 95 ss.

dal DSA[223], nel Regolamento IA l'individuazione delle categorie di rischio, l'ascrizione alle stesse dei diversi sistemi di IA, nonché la disciplina dei differenti livelli di rischio, non è rimessa alla valutazione dei destinatari del Regolamento, bensì è operata da quest'ultimo e con l'attribuzione alla Commissione di ampi poteri di valutazione *ex ante*; sicché, secondo parte della dottrina, tale approccio potrebbe persino suggerire che «le logiche che permeano l'AIA e che ne informeranno l'attuazione abbiano carattere pseudo-emergenziale»[224].

Sul secondo fronte, quello dei limiti al fenomeno icasticamente definito "censura privata", ancora una volta potrà rivelarsi fondamentale – gli orientamenti della giurisprudenza di merito surrichiamata lo hanno già dimostrato - l'efficacia orizzontale delle norme costituzionali (cd. *Drittiwirkung*[225]), ovvero – per quanto qui rileva - la loro diretta ap-

[223] Sul punto, P. Dunn-G. De Gregorio, *AI Act, rischio e costituzionalismo digitale*, in *Medialaws.eu*, 22 aprile 2022, raffrontando i due Regolamenti e, ponendoli in comparazione con il GDPR, rilevano: «Se nel GDPR vi è dunque una delegazione completa, secondo un modello *bottom-up*, dei doveri di valutazione e mitigazione, il DSA si discosta da tale sistema, individuando i criteri oggettivi di classificazione dei *provider*. Tuttavia, questo passaggio da una logica *bottom-up* a una logica *top-down* non è ancora completo: soprattutto nel caso delle piattaforme online di dimensioni molto grandi, infatti, un ampio margine di discrezionalità è comunque previsto per la mitigazione di rischi sistemici connessi alle loro attività. […] Nell'AI Act, il passaggio da un modello *bottom-up* a un modello *top-down* è più marcato […]. La prospettiva adottata dall'AI Act è, in effetti, per certi versi opposta a quella del GDPR. Se nel GDPR la valutazione del rischio e la predisposizione di misure atte a tutelare i diritti individuali alla riservatezza e protezione dei dati erano attività delegate direttamente al titolare e al responsabile del trattamento dati, nel caso dell'AI Act la prospettiva è rovesciata: è il regolamento stesso a operare tale attività. In effetti, se è vero che è presente, con riferimento ai sistemi di IA ad alto rischio, la previsione dell'obbligo di istituire, attuare, documentare e mantenere un sistema di gestione dei rischi, è altresì vero che nell'ecosistema del Regolamento tale norma sembra avere un carattere per lo più residuale […]. Nonostante ciò, sembra tuttavia potersi individuare quanto meno un elemento caratterizzante sia il GDPR, sia il DSA, sia l'AI Act. In effetti, tutti e tre gli atti normativi mirano a realizzare, attraverso il concetto di "rischio", un bilanciamento tra gli interessi in gioco: da un lato, l'interesse, di matrice economica, all'innovazione e allo sviluppo di un mercato unico digitale competitivo sul piano internazionale; dall'altro lato, l'interesse, sovente opposto, alla tutela dei valori democratici e dei diritti e delle libertà fondamentali degli individui. Il rischio funge, in altre, parole, da *proxy* per un'attività, quella del bilanciamento, strettamente connessa a una dimensione costituzionale.[…] In altre parole, sebbene le modalità siano diverse, e diversa sia la declinazione del *risk-based approach*, il fine pare essere, in ultima analisi, univoco: la tutela dei valori fondanti il costituzionalismo digitale europeo».

[224] F. Ferri, *Il giorno dopo la rivoluzione: prospettive di attuazione del regolamento sull'intelligenza artificiale e poteri della Commissione europea*, in *Quaderni AISDUE*, 2, 2024, 18 fonda tale conclusione proprio sul rilievo secondo cui: «Se l'AIA condivide con i regolamenti appena richiamati – *in primis* con il DSA – l'obiettivo di proteggere la base valoriale dell'Unione e il fatto di essere espressione tangibile del costituzionalismo digitale europeo, innova rispetto ad essi anche e soprattutto perché, come argomentato in dottrina, si fonda su un «*top-down risk approach*», evincibile già nella fase di valutazione. […] Ne deriva che l'Unione, specialmente attraverso la Commissione, si riserva di fissare a monte le regole basilari, di attuarle in un'ottica quanto più unitaria e, se del caso, di rivederle ricorrendo a uno strumentario che offra garanzie di flessibilità operativa».

[225] Sull'argomento, v. E. Navarretta, *Costituzione, Europa e diritto privato. Effettività e* Drittwirkung *ripensando la complessità giuridica*, Torino, 2018; P. Femia (a cura di), *Drittwirkung: principi costituzionali e rapporti tra privati*, Napoli, 2018, che svolge un'approfondita disamina della dottrina tedesca. L'argomento è di particolare interesse, in quanto, a differenza dell'ordinamento italiano, l'art. 1 *Grundgesetz* (GG) afferma espressamente che «i […] diritti fondamentali vincolano la legislazione, il potere esecutivo e la giurisdizione come diritti direttamente applicabili». Per l'esame di talune recenti applicazioni, v. F. Episcopo, *L'efficacia orizzontale dei diritti fondamentali al vaglio della Corte Federale Tedesca. Brevi note a margine di alcune recenti sentenze del Bundesverfassungsgericht*, in *giustiziainsime.it*-28 maggio 2020, laddove si rileva che: «Con il *leading case* "Luth" il BVerfG ha affermato che i diritti fondamentali sono inerenti all'"ordine oggettivo di valori costituzionali che devono essere rispettati in tutti i settori del diritto" e,

plicabilità[226] al rapporto (contrattuale) prestatore-utente quali parametri alla cui stregua vagliare la ragionevolezza e proporzionalità delle limitazioni adottate dalle piattaforme in applicazione delle condizioni contrattuali; il che, nella prospettiva pubblicistica, si traduce in garanzia di coerenza del bilanciamento tra libertà di iniziativa economica *ex* art. 41 Cost. e diritti fondamentali dell'utente con i principi personalista, di solidarietà e di eguaglianza *ex* artt. 2 e 3 Cost.[227].

Ed invero, sebbene in dottrina si siano delineati orientamenti di segno diverso[228], che sarebbe impossibile approfondire in questa sede, i fenomeni esaminati mettono in luce

pertanto, possono avere effetti orizzontali indiretti (c.d. *mittelbare Drittwirkung*), che i giudici nazionali devono garantire anche quando chiamati a risolvere controversie tra privati, attraverso l'interpretazione costituzionalmente orientata delle norme di legge e, in particolare, delle clausole generali. In questo senso, i diritti costituzionali possono incidere anche sull'autonomia privata, ma solo in via mediata, cioè tramite l'incidenza delle norme costituzionali sulle fonti eteronome che tale autonomia disciplinano. […] Le due decisioni in commento aprono a una importante innovazione in tema di efficacia orizzontale dei diritti fondamentali. "Stadium Ban" riconosce l'efficacia orizzontale del principio di parità di trattamento nel diritto privato e, cosa non di poco conto, lo fa senza passare per la mediazione delle clausole generali, semplicemente dando rilievo ad alcune situazioni di fatto – formulate in termini chiaramente non tassativi – che giustificano la "responsabilizzazione dei privati" rispetto al dettato dell'art. 3(1) GG. "Hotel Ban", oltre che confermare la soluzione sopra adottata, solleva questione dell'efficacia orizzontale dell'art. 3(3) GG, sebbene non la risolva apertamente. Limitandosi alle considerazioni strettamente necessarie per risolvere il caso di specie, la Corte afferma infatti che, in ogni caso, tale diritto non può creare direttamente dei doveri in capo ai privati, potendo, al limite, richiedere un bilanciamento più "forte" di quello

previsto al paragrafo 1 in caso di conflitto con altri diritti fondamentali. In questo senso, la pronuncia pone le premesse per un dibattito circa lo statuto del principio di non discriminazione nel diritto privato, e segnatamente nel diritto dei contratti; dibattito che da anni interessa la dottrina italiana».

[226] In generale, l'espressione "diretta applicabilità" è da intendere come «affermazione della idoneità delle norme costituzionali – anche quando esse siano (espresse in forma di) principi – a fornire immediatamente e direttamente la disciplina di un rapporto tra privati» (G. D'Amico, *Problemi (e limiti) dell'applicazione diretta dei principi costituzionali nei rapporti di diritto privato (in particolare nei rapporti contrattuali)*, in *Giustizia civile*, 3, 2016, 443 ss.)

[227] T. Martines, *Diritto costituzionale*, Milano, 2010, 581. Con particolare riferimento ai profili costituzionali dell'autonomia privata, v. F. Macario, voce *Autonomia privata (profili costituzionali)*, in *Enc. dir.*, *Annali*, VIII, Milano, 2015, 61 ss.; M. Esposito, *Profili costituzionali dell'autonomia privata*, Padova, 2003; S. Rodotà, *Per un costituzionalismo di diritto privato*, in *Rivista critica del diritto privato*, 2004, 1, 11 ss.

[228] *Ex multis*, R. Bin, *L'applicazione diretta della Costituzione, le sentenze interpretative, l'interpretazione conforme a Costituzione della legge*, in AA.VV., *La circolazione dei modelli e delle tecniche del giudizio di costituzionalità in Europa*, Napoli, 2010; F. Mannella, *Giudici comuni e applicazione della Costituzione*, Napoli, 2011; G. D'Amico, *Problemi (e limiti) dell'applicazione diretta dei principi costituzionali*, cit., cui si rinvia per l'approfondimento delle posizioni emerse nella dottrina civilistica e per gli ampi riferimenti bibliografici. L'A. evidenzia come i maggiori dubbi interpretativi emergano in ordine alla diretta applicabilità delle disposizioni costituzionali «che contengono (o che consistono in) principi. È evidente, infatti, che l'applicabilità diretta delle norme costituzionali espresse in forma di "regole" (ad es. art. 36 Cost. nella parte in cui stabilisce che il lavoratore ha diritto al riposo settimanale e a ferie annuali retribuite) non è in discussione». Inoltre – rileva ulteriormente l'A.- «l'applicazione diretta dei principi costituzionali (pertinenti) deve ritenersi possibile – e su questo concorda la dottrina prevalente – quando manchi una regolamentazione legislativa (e non soccorrano gli ordinari procedimenti di integrazione analogica dell'ordinamento), e soprattutto quando il giudice si trovi a fare applicazione di "clausole generali" (contenendo queste ultime – per riprendere una formula abbastanza ricorrente – una sorta di "delega", che legittima il giudice a individuare, senza la mediazione di una norma legislativa, la regola da applicare alla fattispecie da giudicare). Tuttavia […] una parte (ancora minoritaria, ma) sempre più ampia della dottrina privatistica e della giurisprudenza civile tende ad ammettere (più o meno consapevolmente) modelli di argomentazione e di soluzione delle controversie […] basate su un uso dei principi costituzionali viepiù pervasivo, che "pratica" (o, comunque, presuppone la possibilità di) una applicazione diretta di tali principi […]» (465 ss.).

la necessità di riconoscere perdurante attualità[229] alla tesi, autorevolmente sostenuta[230], secondo cui, in considerazione del carattere unitario dell'ordinamento, le garanzie di libertà affermate nei confronti del potere pubblico, non possono che valere anche nei rapporti tra consociati.

D'altra parte, come è stato di recente affermato, «[…]Per valutare la portata della "rivoluzione" costituzionale del secondo dopoguerra è necessario abbandonare una visione trascendente della Costituzione, in favore di una sua concezione immanente, pane quotidiano […] pure per privati cittadini nella loro vita di relazione, improntata appunto al principio di solidarietà», sicché «[…]di fronte alle inevitabili lacune che si manifestano nei rapporti concreti, inibire al giudice l'applicazione diretta dei princìpi [costituzionali] significherebbe astenersi dalla tutela effettiva dei diritti […]»[231].

L'applicazione delle suddette coordinate interpretative alla questione che ci occupa potrebbe determinare un duplice ordine di effetti, in vista di un esito di più ampia portata: consentirebbe, come anzidetto, di colmare i vuoti di tutela potenzialmente derivanti dalle scelte regolatorie contenute nel DSA; in secondo luogo – ma si tratta di un aspetto non disgiunto dal primo – garantirebbe la piena ed effettiva tutela giurisdizionale dei diritti degli utenti, scongiurando il rischio che la "legittimità procedimentale", ovvero il rispetto delle regole procedurali codificate dal Regolamento per l'adozione delle misure limitative finisca, per paradosso, con l'offuscare la necessità e l'intensità del controllo (anche) di ordine sostanziale - da operare alla stregua dei parametri costituzionali (artt. 2, 3, 21, 41 co. 2 Cost.) - sulla ragionevolezza e proporzionalità delle limitazioni decise dai prestatori di servizi digitali.

L'esito di maggior pregnanza perseguito con l'applicazione diretta delle norme costituzionali nei termini sopra esposti, sarebbe quello della piena attuazione del principio personalista, che, quale principio cardine sotteso all'intero impianto costituzionale, non può non permeare anche i rapporti che si esplicano in Rete; al contempo, come già auspicato dalla dottrina, troverebbe piena realizzazione il ruolo del diritto costituzionale, nel nuovo come nel precedente secolo: «farsi carico della libertà per il futuro delle nostre società»[232].

---

[229] In tal senso, v. F. Paruzzo, *I sovrani della rete* cit., 136 ss.; A. Lamberti, *Libertà di informazione, poteri privati e tutela dei dati personali nell'era digitale*, in *Dirittifondamentali.it*, 3, 2023, 22 ss.

[230] G. Lombardi, *Potere privato e diritti fondamentali*, Torino, 1970, 53.

[231] G. Silvestri, *Drittwirung*. Relazione al convegno annuale dell'Associazione civilisti italiani *Costituzione e diritto privato. Dialoghi*-Firenze, 13-14 dicembre 2024, reperibile in https://www.civilistiitaliani.eu/images/convegni/Firenze_13_e_14_dicembre_2024/Gaetano_Silvestri_Relazione_provvisoria_Drittwirkung_Firenze_13_14_dicembre_2024.pdf. Evidenzia ulteriormente l'A.: «Il vero problema che torna sempre a galla, benché spesso sommerso da complesse argomentazioni, è quello della natura delle norme costituzionali di principio, la cui piena giuridicità fu da subito negata dalla giurisprudenza dominante dell'epoca del "gelo" costituzionale e difesa invece dalla magistrale dottrina di Vezio Crisafulli. Se nessuno dubita della nullità di un contratto o di una singola clausola per contrasto con "norme imperative", giacché lo dispone l'art. 1418 c.c., non ritenere valido questo assunto per le norme costituzionali di principio vuol dire che non si considerano queste ultime vere e proprie norme giuridiche o quanto meno le si ritiene norme, per così dire, a giuridicità depotenziata».

[232] A. Simoncini, *Sovranità e potere nell'era digitale*, cit., 36.

Libertà di espressione e verità artificiali. Quale *marketplace of ideas* nella società dell'algoritmo?*

Luca Catanzano

Abstract

Il nuovo ecosistema dell'informazione - rivoluzionato dalle nuove tecnologie e dominato dalle piattaforme digitali - ha avuto un impatto significativo sulla libertà di espressione e sul diritto ad una informazione veritiera. La stagione del costituzionalismo digitale si caratterizza per il tentativo europeo di limitare il potere delle big tech, che da attori economici si sono trasformati in poteri privati digitali, incidendo significativamente su una pluralità di diritti degli utenti ed esercitando de facto funzioni di natura para-costituzionale. Partendo da una riflessione sulla inattuabilità di un *free marketplace of ideas* ci si sofferma sui rischi derivanti dal perfezionamento del *deepfake* e dal tentativo di una sua regolazione nell'AI act, alla luce delle problematiche che le strategie di contrasto alla disinformazione pongono - nella prospettiva del diritto costituzionale - alla limitazione della libertà di espressione.

The new information ecosystem - revolutionized by new technologies and dominated by digital platforms - has had a significant impact on freedom of expression and the right to truthful information. The season of digital constitutionalism is characterized by the European attempt to limit the power of big tech companies, which have transformed from economic actors into private digital powers, significantly impacting on a plurality of user rights and exercising "de facto" para-constitutional functions. Starting from a reflection on the impracticability of a free market place of ideas we focus on the risks deriving from the improvement of the deepfake and the attempt to regulate it in the AI Act, in light of the problems that strategies to combat disinformation pose - from the perspective of constitutional law - to the limitation of freedom of expression.

Sommario

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

Keywords
disinformazione – piattaforme – Intelligenza Artificiale – regolazione - costituzionalismo

1. Introduzione

La rivoluzione digitale ha stravolto tutti i settori dell'esistenza collettiva e in particolar modo quella giuridica[1], tanto da potersi considerare – con Marcel Mauss – un fatto sociale totale[2] che mette in moto, in alcuni casi, la totalità della società e delle sue istituzioni. La filosofia dell'informazione ha dimostrato come gli sviluppi della digitalizzazione abbiano creato le condizioni per la diffusione dei sistemi di Intelligenza Artificiale[3], il cui utilizzo interagisce con un ampio spettro di diritti fondamentali[4], con particolare riferimento alla libertà di manifestazione del pensiero e al diritto all'informazione[5]. Quello che appare come uno dei principali temi del momento per chi si occupa di diritto dell'informazione nell'attuale "stagione algoritmica", emerge in maniera dirompente nella recente *European Declaration on Digital Rights and Principles for the Digital Decade* che nel capitolo IV, dedicato alla partecipazione allo spazio pubblico digitale, dopo aver enunciato nell'art. 13[6] il diritto di ogni persona alla libertà di espressione e informazione nell'ambiente digitale, nell'art. 15 lett. d)[7] enuncia l'impegno (di Parlamento europeo, Consiglio e Commissione) a creare un ambiente digitale in cui le persone siano protette dalla disinformazione e dalla manipolazione delle informazioni.

Nella consapevolezza dello stretto legame tra libertà di espressione e innovazione tecnologica[8] l'obiettivo di questo saggio concerne i profili costituzionalmente problematici dell'impatto dei sistemi di intelligenza artificiale sulla libertà di manifestazione del pensiero e sul diritto ad una informazione veritiera. La multiformità dei poli che costituiscono l'oggetto di questo tipo di analisi rende particolarmente complesso il tentativo di riassumere le possibili intersezioni tra libertà di espressione e IA[9]. In particolare, ci si concentrerà sui rischi derivanti dall'utilizzo dell'IA nell'ambito di strategie

[1] A. Garapon – J. Lassegue, *Justice Digitale: Révolution Graphique et Rupture Anthropologique*, Paris, 2018; trad. it. *Giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021, 79.

[2] M. Mauss, *Saggio sul dono. Forma e motivo dello scambio nelle società arcaiche*, Torino, 2002, 134.

[3] L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, 21.

[4] La relazione dell'Agenzia dell'Unione europea per i diritti fondamentali *Preparare un futuro giusto-L'intelligenza artificiale e i diritti fondamentali.*

[5] O. Pollicino – P. Dunn, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2024.

[6] *European Declaration on Digital Rights and Principles for the Digital Decade*, art. 13.

[7] Ivi, art. 15 lett. *d*).

[8] G.E. Vigevani, *La libertà di manifestazione del pensiero*, in M. Bassini - M. Cuniberti – C. Melzi d'Eril – O. Pollicino - G. E. Vigevani, *Diritto dell'informazione e dei media*, Milano, 2022, 4.

[9] C. M. Reale – M. Tommasi, *Libertà di espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nel nuovo ecosistema costituzionale*, in *DPCE Online*, 1, 2022, 326 ss.

di contrasto alla disinformazione da un lato[10] e all'utilizzo di tali sistemi per diffondere contenuti disinformativi come *fake news* e *deepfake* dall'altro. Per rispondere a tale interrogativo, dopo una analisi della libertà di manifestazione del pensiero, della metafora del libero mercato delle idee e del formante tecnologico della disinformazione, si proverà a riflettere su alcune conseguenze derivanti dal mutamento della sfera pubblica[11], dominata da attori economici che sono divenuti poteri privati digitali incidendo significativamente sulle libertà fondamentali in questione e sul connesso ripensamento delle strategie regolatorie da parte dell'Unione. La dottrina ha da tempo messo in luce il come dopo una prima fase di liberismo digitale e una seconda di attivismo giudiziale[12], siamo entrati in una nuova stagione rappresentata dal costituzionalismo digitale, definita come il plesso di interventi legislativi dell'Unione volto a limitare i nuovi poteri privati con il fine di salvaguardare e promuovere i valori intrinseci del costituzionalismo europeo. Il fenomeno in corso non va considerato come una semplice disintermediazione, ma più propriamente come una "reintermediazione", in cui i mediatori sono sostituiti dalle piattaforme, che orientato la nostra vita quotidiana in misura maggiore rispetto ai mediatori tradizionali[13]. Una delle caratteristiche principali della società dell'algoritmo[14]- oltre al sempre maggiore rilievo dello strumento che la definisce – è proprio l'emersione di questi nuovi attori privati. Le "compagnie del digitale" hanno un potere politico che nessuno ha mai avuto[15], creano opinioni, hanno una funzione regolatrice della vita dei privati e degli Stati condizionando l'attività privata e pubblica. In questo contesto, problemi classici del costituzionalismo come la difesa dagli abusi dei pubblici poteri si arricchisce di un capitolo nuovo: «Condurre le oligarchie del digitale all'interno dei valori propri delle democrazie occidentali. È il costituzionalismo digitale»[16].

2. Libertà di espressione e disinformazione nell'era dell'intelligenza artificiale. Quale libero mercato delle idee?

Norberto Bobbio, quasi sessant'anni fa, sostenne che lo sviluppo della tecnica, l'ampiamento delle conoscenze e l'intensificazione dei mezzi di comunicazione avrebbero

[10] L'intelligenza artificiale nella sua funzione "selettiva" rappresenta uno dei più promettenti argini alle *fake news*. Ivi, 328

[11] J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Milano, 2023.

[12] G. De Gregorio, *The Rise of Digital Constitutionalism in the European Union*, in *International Journal of Constitutional Law*, 19-1, 2020, 41.

[13] L. Violante, Prefazione a *Intelligenza artificiale e democrazia*, cit., 2.

[14] J. M. Balkin, *Free Speech in the Algoritmich Society: Big Data, Private Governance, and New School Speech Regulation*, in *U.C Davis Law Review*, 51, 2018, 1149 ss.

[15] Tra i contributi apripista che hanno denunciato aspetti e ricadute problematiche di tale fenomeno vi è S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, 2019; trad. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, 2019.

[16] Ivi, 4.

profondamente mutato l'ordine della vita e dei rapporti sociali, creando occasioni favorevoli alla nascita di nuovi bisogni e quindi, a nuove richieste di libertà e poteri[17]. Nello sviluppare questa sua riflessione fece riferimento all'ecosistema dell'informazione: «La crescente quantità e intensità di informazioni cui l'uomo oggi è sottoposto fa sorgere sempre più forte il bisogno di non essere ingannati, eccitati, turbati da una propaganda assillante e deformante; si profila, di contro al diritto di esprimere le proprie opinioni, il diritto alla verità delle informazioni»[18]. Alcuni decenni dopo rispetto al suo discorso – precisamente nel 2016 – *post-truth* si affermerà come parola dell'anno nel *Oxford Dictionary* e le strategie di contrasto alla disinformazione si situeranno al centro di un approfondito dibattito dottrinale e istituzionale soprattutto nelle società libere e tecnologicamente avanzate[19]. Il *Digital Service Act*, l'*Artificial Intelligence Act*, nonché i recenti Regolamenti sulla trasparenza e il *targeting* della pubblicità politica e il Regolamento sulla libertà dei media - oltre al nuovo *strengthened Code of Practice on Disinformation* - dimostrano che sul versante europeo, pur avendo tale plesso di interventi un campo applicativo più amplio, i temi del contrasto disinformazione e della tutela della libertà di espressione sono a tutti gli effetti al centro dell'agenda europea.
Secondo il noto paradosso di Ernst-Wolfgang Böckenförde: «Lo Stato liberale secolarizzato vive di presupposti che non è in grado di garantire»[20]. La dottrina costituzionalistica italiana ha utilizzato questa formula per inquadrare la difficoltà che le misure di contrasto alle *fake news* pongono – nella prospettiva della tutela costituzionale – alla libertà di manifestazione del pensiero, che dello Stato liberale rappresenta la "cartina tornasole"[21]. La Corte costituzionale italiana ha più volte ribadito che il riconoscimento del diritto alla libertà di espressione costituisce la pietra angolare e il cardine dell'ordinamento democratico[22]. Tale libertà sembra attualmente attraversare un periodo sotto sforzo e potremmo dire che la sua stagione di fioritura sia ormai alle spalle[23]. Una nuova tensione censoria ha individuato nel *free speech* il principale bersaglio da combattere e i confini del dicibile appaiono progressivamente sempre più incerti. Il dibattito si alimenta di posizioni apparentemente inconciliabili tra le istanze più libertarie di chi rivendica un ampio spazio alla libertà di espressione e chi – facendosi interprete di nuove sensibilità – invece sostiene la necessità di un'ulteriore limitazione di

[17] Il saggio *Presente e avvenire dei diritti dell'uomo* è costituito dal testo della conferenza tenuta a Torino nel dicembre 1967 in occasione del Convegno nazionale sui diritti dell'uomo promosso dalla Società italiana per l'organizzazione internazionale, in occasione del ventesimo anniversario della Dichiarazione universale. N. Bobbio, *L'età dei diritti*, Torino, 1997, 28.

[18] *Ibid.*

[19] H.E. Kissinger – E. D. Schmidt - D. Huttenlocher, *L'era dell'intelligenza artificiale. Il futuro dell'identità umana,* Milano, 2023, 19 ss.

[20] E. W. Böckenförde, *Diritto e secolarizzazione. Dallo Stato moderno all'Europa unita*, Bologna, 2010.

[21] G. Pitruzzella - O. Pollicino - S. Quintarelli, *Parole e potere. Libertà d'espressione, hate speech e fake news*, Milano, 2017.

[22] Corte cost., 19 febbraio 1965, n. 9; Corte cost., 14 aprile 1965, n. 25; Corte cost., 23 marzo 1968 n. 11; Corte cost., 10 luglio 1968, n. 98; Corte cost., 17 aprile 1969, n.84; Corte cost., 8 luglio 1971, n. 168; Corte cost., 2 maggio 1985, n.126.

[23] M. Manetti, *Una stagione di fioritura della libertà di pensiero è ormai alle spalle,* in *Rivista AIC*, 3, 2016, 1.

tale libertà[24]. La dicotomia sembra essere quella da un lato, della democrazia militante[25] impegnata nella difesa di valori che l'esercizio degli stessi diritti fondamentali rischia di minacciare e dell'altro quella di una democrazia tollerante e quindi critica verso l'eccessivo controllo sull'esercizio delle libertà[26]. Il tema del possibile contrasto alla diffusione di *fake news* e *deepfakes* intercettando l'essenza del dilemma formulato da Böckenförde tocca dunque le radici del costituzionalismo, evocando tra le varie domande di ricerca possibili quella – complessa quanto attuale – del rapporto tra verità e Stato costituzionale[27], che si staglia dietro ad ogni riflessione giuridica sulla libertà di espressione e il diritto ad una informazione veritiera. Nel modello di tutela della libertà di espressione si gioca la partita determinante in questo campo e la scelta di intervenire per depurare la rete dalle informazioni false sembra evocare misure che limitano la libertà di manifestazione del pensiero dentro limiti più rigorosi di quelli stabiliti dalle costituzioni degli ordinamenti democratici[28]. Può essere utile ai fini del presente saggio fare riferimento a uno dei principali dibattiti che storicamente ha caratterizzato la dottrina costituzionale italiana in tema di libertà di espressione è quello sulla qualificazione di tale libertà come individuale o funzionale. Nel suo celebre studio su la libertà di manifestazione del pensiero nell'ordinamento italiano Esposito critica quella dottrina che dichiarava superata la tesi del fondamento individualistico e liberale della libertà di espressione[29]. A favorirne il capovolgimento, aprendo a una lettura di questa libertà come "esercizio di una funzione" – da qui il nome della dottrina funzionalista – è stata l'idea che la contrapposizione tra individuo e Stato sia un'astrazione e che il rapporto tra l'uno e l'altro sia quello della parte con il tutto; alla luce di questa lettura non solo i doveri ma anche le libertà dovrebbero in tal modo essere inquadrate come modo di partecipazione del singolo alla vita della comunità. Osservava come in progresso di tempo la tendenza fosse quella di accentuare di fatto il significato sociale di diritti che tradizionalmente venivano intesi come individualistici, unita alla struttura intimamente sociale dei diritti di comunicazione, manifestazione o dichiarazione del pensiero. L'Autore, critico verso questo tipo di lettura, analizzava nella prospettiva del diritto comparato alcuni casi estremi di inquadramento funzionale di tale libertà con la finalità di dimostrare che l'ordinamento italiano si distingue profondamente da questi[30]. L'articolo 21 Cost. al

[24] Il riferimento è alle nuove ondate culturali rappresentate dal *politically correct*, dalla *cancel culture* e dall'ideologia *woke*. G. Pino, *La strada dell'inferno è lastricata di buone intenzioni. Luci e ombre della cancel culture, in Rivista italiana di filsoofia del linguaggio*, 1, 2024, 12. La *Critical race theory*, da alcuni autori considerata come una sorta di *postmodern censorship*, ha evidenziato come la protezione assoluta della libertà di espressione giochi spesso contro gli appartenenti alle minoranze quando le opinioni, le parole e i comportamenti espressivi sono funzionali a veicolare e rimarcare la differenza tra la maggioranza e la minoranza. G. Pino, *Discorso razziale e libertà di manifestazione del pensiero*, in *Politica del diritto*, 39 - 2, 2008, 287.

[25] K. Loewenstein, *Militant Democracy and Fundamental Rights*, in *American Political Science Review*, 1937, 31, 417 ss.

[26] O. Pollicino, *La prospettiva costituzionale della libertà d'espressione nell'era di internet*, in questa *Rivista*, 1, 2018, 2.

[27] G. E. Vigevani, *La libertà,* cit., 6.

[28] O. Pollicino, *La prospettiva costituzionale,* cit., 49 ss.

[29] C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958, 4 ss.

[30] Ivi, 7 Dove analizza l'art. 125 della Costituzione dell'URSS in cui la libertà di espressione è garantita dalla legge in armonia con gli interessi dei lavoratori ed allo scopo di rafforzare l'organizzazione socialista.

contrario, riconoscendo a tutti il diritto di manifestare liberamente il proprio pensiero, non attribuisce un diritto funzionale ma, secondo un'altra interpretazione, un diritto individuale. Attraverso una lettura sistematica della Carta costituzionale che tiene conto dell'inserimento dell'art. 21 nell'ambito dei "rapporti civili", dell'attribuzione a tutti e non solo ai cittadini di tale libertà e della mancanza – contrariamente ad altre enunciazioni come quello sull'iniziativa economica – di accenni alla funzione sociale di questo diritto, Esposito sostiene l'interpretazione individualistica[31] di tale libertà, divenendo un riferimento assoluto per la dottrina fino ai nostri giorni. Quando si afferma che nella Costituzione è garantito il diritto di manifestazione del pensiero in senso individualistico si intende dire che esso è garantito al singolo in quanto tale, indipendentemente dagli svantaggi che possono derivarne allo Stato; quindi la divulgazione di un pensiero critico non verrà riconosciuta in misura differente a seconda di quelli che possono essere i vantaggi o gli svantaggi per la comunità statale ed ogni limitazione, lungi dal potersi dedurre dalla natura del diritto riconosciuto dovrà fondarsi in particolari disposizioni che ne giustifichino l'affermazione[32]. Esposito, tra quelle che sono le ragioni tradizionalmente addotte contro la censura e gli impedimenti al libero uso dei mezzi di diffusione del pensiero, cita esplicitamente, oltre alla sempre possibile fallibilità dei censori, l'incertezza tra una netta distinzione tra vero e falso[33], questione oggi centrale nel dibattito sulle *fake news*.

Le strategie di contrasto alla disinformazione si inseriscono nell'ambito della questione dei limiti alla libertà di espressione che, come dimostrato dalle migliori riflessioni in chiave comparata, variano fortemente a seconda delle tradizioni costituzionali di riferimento. Un approccio ormai classico che si utilizza nell'affrontare il tema della regolazione della libertà di espressione è quello di guardare, in ottica costituzionalmente comparata, al paradigma statunitense. La tradizione o famiglia giuridica raccoglie quei sistemi giuridici che condividono un complesso di atteggiamenti radicati e storicamente condizionati sulla natura del diritto e sul suo ruolo nella società e sul funzionamento di un sistema giuridico[34]. Per Ugo Mattei, la *Rule of Professional Law* è quella famiglia che si caratterizza per l'egemonia del diritto come modello di organizzazione sociale: la tradizione giuridica occidentale dove diritto e politica, così come diritto e religione, sono separati, all'interno di questa famiglia, quella tra *common law* e *civil law* è una sotto distinzione. La tradizione costituzionale europea e quella statunitense, pur rientrando nella medesima famiglia giuridica della *Rule of Professional Law*[35] presentano significative differenze in relazione alla presente indagine. Il costituzionalismo europeo, infatti, pur riconoscendo una centralità alla libertà di manifestazione del pensiero, ereggendola a pietra angolare di ogni società democratica, considera il campo di applicazione di questo diritto fondamentale come potenzialmente «oggetto di limitazioni o restrizioni dovute all'esigenza di prevenire abusi o bilanciarne l'esercizio con altri diritti fonda-

[31] Ivi, 8.

[32] Ivi, 10.

[33] *Ibid.*

[34] J.H Merryman, *The Civil Law Tradition*, Stanford 1969; trad. it. *La tradizione di civil law nell'analisi di un giurista di common law*, Milano, 1973, 9.

[35] U. Mattei – G.P Monateri, *Introduzione breve al diritto comparato*, Padova, 1997, 51 ss.

mentali meritevoli di tutela»[36] con cui gioca alla pari[37] e – diversamente dall'ordinamento giuridico statunitense - la sua stella polare è la dignità umana, non a caso primo dei valori fondanti dell'Unione citati nell'art. 2 del TUE. Se osserviamo, inoltre, i parametri fondamentali espressi dalla CEDU in materia di libertà di espressione, possiamo capirne il margine di protezione di quelle espressioni che ne concretano un esercizio e provare a delineare il perimetro entro il quale circoscrivere il tema della rilevanza costituzionale delle *fake news* e dei *deepfakes* e l'impatto sulla libertà in questione che possono causare eventuali misure di contrasto alla loro diffusione.

L'art. 10 della CEDU[38] dopo la solenne affermazione di una libertà di manifestazione del pensiero – che in questa formulazione include il diritto a ricevere informazioni – la norma convenzionale esplicita le possibili limitazioni a tale libertà e le loro caratteristiche essenziali: devono essere previste da una legge, essere proporzionali e orientate al raggiungimento di uno degli obiettivi enunciati esplicitamente da tale articolo, che inquadra quelle che nella topografia di un conflitto tra libertà di espressione ed altre situazioni giuridicamente protette dall'ordinamento convenzionale possono comportarne una limitazione. I limiti sopra enunciati, con particolare riferimento alla sicurezza pubblica e alla protezione della salute, rappresentano ipotesi tipiche di situazioni che possono essere lese da una informazione falsa o da un *deepfake*; basti pensare solamente alle *fake news* che sono state essere divulgate nel contesto pandemico dall'area no-vax.

In tema di copertura convenzionale della libertà d'espressione l'armamentario della Corte di Strasburgo è arricchito dall'art.17 della Convenzione, che – con un accento di democrazia militante – disciplina l'abuso di diritto, ai sensi di tale disposizione «Nessuna disposizione della presente Convenzione può essere interpretata nel senso di comportare il diritto di uno Stato, un gruppo o un individuo di esercitare un'attività o compiere un atto che miri alla distruzione dei diritti o delle libertà riconosciuti nella presente Convenzione o di imporre a tali diritti e libertà limitazioni più ampie di quelle previste dalla stessa Convenzione»[39]. Mentre l'applicazione dell'art.10 della Convenzione postula una logica di bilanciamento con altri interessi meritevoli di tutela con cui l'esercizio della libertà di espressione dovrà conciliarsi, il ricorso all'art.17 risponde a logiche giuridiche differenti andando a punire l'abuso di un diritto sancito dalla CEDU in funzione strumentale alla distruzione di altri diritti o libertà o alla limi-

[36] O. Pollicino, *La prospettiva costituzionale della libertà d'espressione nell'era di internet*, in questa *Rivista*, 1, 2018, 51.

[37] O. Pollicino – P. Dunn, *Intelligenza artificiale*, cit., 11.

[38] Art. 10 CEDU: «Ogni persona ha diritto alla libertà d'espressione. Tale diritto include la libertà d'opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera. Il presente articolo non impedisce agli Stati di sottoporre a un regime di autorizzazione le imprese di radiodiffusione, cinematografiche o televisive. 2. L'esercizio di queste libertà, poiché comporta doveri e responsabilità, può essere sottoposto alle formalità, condizioni, restrizioni o sanzioni che sono previste dalla legge e che costituiscono misure necessarie, in una società democratica, alla sicurezza nazionale, all'integrità territoriale o alla pubblica sicurezza, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, alla protezione della reputazione o dei diritti altrui, per impedire la divulgazione di informazioni riservate o per garantire l'autorità e l'imparzialità del potere giudiziario».

[39] Art. 17 CEDU

tazione degli stessi in modo più severo rispetto a quanto stabilito dalla convenzione; tale "strumento" viene utilizzato nei casi in cui si esclude *ab origine* la possibilità di un bilanciamento, come si è verificato per esempio nel caso Garaudy contro Francia[40] in tema di negazionismo. Questo strumento è stato considerato dalla dottrina che si è occupata di comparazione del *free speech* tra Europa e USA come qualcosa di irricevibile nell'ordinamento statunitense. Occorre dire sin da subito che in quel sistema la libertà di espressione non è illimitata, ma meno limitata – per ragioni storiche e giuridiche - rispetto al continente europeo e che anche lì sono state elaborate delle critiche sulla eccessiva espansività della sua tutela. All'impostazione giuridica statunitense sul *freedom of speech* corrisponde l'intento di legittimare la repressione del pensiero solamente a partire dal danno, oggettivamente valutabile, che esso possa arrecare agli interessi della collettività, conducendo alla distinzione tra puro pensiero e principio di azione applicata dapprima con la dottrina del *clear and present danger* e poi con la categorizzazione di messaggi non protetti in quanto dannosi in via di principio[41].

Per quanto riguarda la disciplina della libertà di espressione nel contesto dell'Unione Europea occorre considerare il fatto che l'attenzione per i diritti fondamentali di prima generazione si è sviluppata in ambito sovranazionale di matrice euro-unitaria solo a partire dalla Carta di Nizza, che attraverso l'Art.11 offre il parametro più rilevante per quanto riguarda la libertà d'espressione, ai sensi del quale: 1. Ogni persona ha diritto alla libertà di espressione. Tale diritto include la libertà di opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera. 2. La libertà dei media e il loro pluralismo sono rispettati.

Sin dalla sua fondazione, il *dna* della libertà di espressione nel contesto giuridico europeo, si caratterizza per la sua sostanziale cedevolezza, forgiandosi dentro quella che potremmo definire come una "filosofia del limite" fondata sulle seguenti coordinate: possibile abuso del diritto, non assolutezza, bilanciamento e pari ordinazione rispetto agli altri diritti fondamentali. il modello statunitense, invece, si caratterizza per una eccezionalità del *free speech*, di cui emblematica è la formulazione del Primo Emendamento, stella polare del diritto costituzionale statunitense e libertà che segna la caratterizzazione di quell'ordinamento[42].

Una delle espressioni che ha avuto maggiore successo nella letteratura sulla libertà di manifestazione del pensiero è quella rappresentata dalla metafora del *market place of ideas*, coniata dall'ampia riflessione statunitense in tema di primo emendamento della Costituzione[43]. Secondo Giuliani la storia del diritto potrebbe essere studiata, dal punto di vista linguistico, come un susseguirsi di metafore. In questo campo di studi è divenuta celebre la leggendaria *dissenting opinion* del *Justice* Holmes nel 1919[44], con cui viene utilizzata la suddetta metafora – appartenente all'immaginario economico – che

[40] CEDU, *Garaudy c. Francia*, ric. 65831/01 (2003)

[41] A. Pace – M. Manetti, *Commentario della Costituzione. Art. 21 Rapporti civili*, Bologna, 2006, 228.

[42] O. Pollicino, *Potere digitale*, in *Enciclopedia del diritto – I tematici*, 5, 2023, 411 ss.

[43] V. Zeno-Zencovich, *La libertà di espressione. Media, mercato, potere nella società dell'informazione*, Bologna, 2004, 95.

[44] Corte suprema degli Stati Uniti, *Abrams c. Stati Uniti*, 250 US 616, 1919.

successivamente verrà riproposta dalla Corte suprema americana nella sfida della regolazione di internet[45]. Per provare ad inquadrare questa metafora dal punto di vista giuridico può essere utile una breve ricognizione della sua elaborazione sul piano più strettamente filosofico. Nella storia del pensiero l'elaborazione di questa teoria si suole far risalire a John Milton (1608-1674) e a John Stuart Mill (1806-1873), pur dovendosi considerare sempre che queste costituiscono difese della libertà di espressione proprie del loro tempo. John Milton, tra i primi difensori della libertà di espressione in Età moderna, nel suo celebre pamphlet di Areopagitica[46], aveva lanciato uno storico appello contro la censura[47], rivendicando il libero scambio delle idee e delle opinioni come requisito ineludibile del progresso della conoscenza volto al raggiungimento della verità. Secondo questa teoria la libertà di espressione è uno strumento necessario per la ricerca della verità, di conseguenza, lo Stato non deve intervenire in questo processo perché la verità ha sufficiente forza per imporsi di fronte all'errore[48]. È grazie alla successiva riflessione di John Stuart Mill che si deve una delle principali argomentazioni classiche in lingua inglese secondo cui la libertà di espressione farebbe emergere spontaneamente la verità attraverso il libero mercato delle opinioni che consente di scansare ciò che è falso. In *On Liberty,* ad esempio egli sostenne che impedire l'espressione delle opinioni «è un crimine, […] perché significa derubare la razza umana, i posteri altrettanto che i vivi, coloro che dall'opinione dissentono ancor più di chi la condivide»[49] privando l'uomo di una doppia opportunità: «se l'opinione è giusta, sono privati dell'opportunità di passare dall'errore alla verità; se è sbagliata, perdono un beneficio quasi altrettanto grande, la percezione più chiara e viva della verità, fatta risaltare dal contrasto con l'errore»[50]. Come dimostrato dal dibattito filosofico politico successivo, che ha messo in luce alcune criticità di queste teorie, la libertà di manifestazione del pensiero è una condizione sicuramente necessaria ma non sufficiente per salvaguardare "standard aletici" decenti: i pesi e contrappesi del *free marketplace of ideas*, che la tradizione liberale ha storicamente considerato come il più efficace argine contro la falsità delle informazioni, «hanno in realtà finito per conferire a chiunque licenza di mentire»[51]. Nella società dell'algoritmo, le democrazie costituzionali hanno mostrato di poter lasciar spazio alla proliferazione massiva di informazioni false, considerabili come una sindrome autoimmune della formula democratica. È come se l'anticorpo della libertà di espressione attaccasse sé stesso, invece di generare la verità attraverso la collisione con l'errore e facendo sì che le informazioni che sopravvivono ai processi di selezione possano essere quelle non veritiere[52], potendo ledere beni giuridici individuali, come l'onore e la reputazione o incidere direttamente o indirettamente sulla libertà di cittadini di esercitare il diritto di voto e quindi sul corretto funzionamento

[45] Corte suprema degli Stati Uniti, *Reno c. ACLU*, 521 US 844, 1997.

[46] J. Milton, *Areopagitica. Discorso per la libertà di stampa*, Milano, 2022.

[47] G. E. Vigevani, *La libertà*, cit., 6.

[48] F. J. Ansuategui Roig, *Libertà di espressione: ragione e storia*, Torino, 2018, 65 ss.

[49] J. S. Mill, *Saggio sulla libertà*, Milano, 2014, 35.

[50] *Ibid.*

[51] F. D'Agostini – M. Ferrera, *La verità al potere. Sei diritti aletici*, Torino, 2019, 82.

[52] Ivi, 83 ss.

delle istituzioni democratiche. Sono infatti numerosi gli interessi pubblici e privati che circondano la libertà di espressione: dalla tradizionale tutela dell'onorabilità delle persone, all'interesse a proteggere i mercati passando per il più generale interesse di tutti ad un'informazione attendibile e quindi "vera"[53].

La metafora del libero mercato delle idee non appartiene alla tradizione costituzionale europea e la migrazione e importazione di metafore costituzionali appartenenti a tradizioni diverse (il campo di origine) può causare delle crisi di rigetto nel campo di destinazione[54]. Zeno-Zencovich, nel riflettere sulla validità di questa metafora, si domanda se il mercato delle idee possa davvero considerarsi un mercato oppure no e cosa sia ciò che lo distingue dagli altri mercati. Mentre in un mercato tradizionale i fattori dominanti sono i rapporti quantità/prezzo, correlati alla qualità e al tempo, nel *marketplace of ideas* paiono preminenti alcuni valori diversi: in primis tutte le idee sono qualitativamente uguali (con esclusione di alcune estreme); le idee sono beni non consumabili e condivisibili e quindi non vi è un rapporto diretto quantità/prezzo; l'obiettivo che si persegue non è l'efficienza economica ma la massima accessibilità a idee diverse. Non ci si trova quindi di fronte ad un mercato tradizionale e la formula magica secondo questo orientamento dottrinale nasconderebbe in realtà le incoerenze di una teoria politico-giuridica volta a espandere al massimo il principio della libertà di manifestazione del pensiero[55].

Oltre alle ragioni giuridiche summenzionate bisogna considerare un ulteriore fattore rappresentato dalla tecnologia: il nuovo ecosistema dell'informazione dominato dalle piattaforme digitali, che da operatori economici sono diventati veri e propri poteri privati che incidono significativamente sulle libertà e sul pluralismo del dibattito pubblico, rende impossibile l'applicazione di questa metafora. Nell'attuale stagione algoritmica la realtà digitale è ormai transitata verso una dimensione oligopolistica, nonostante la sua iniziale tensione libertaria, sottoponendo la sfera pubblica[56] a un cambio radicale[57]definito da alcuni autori come "piattaformizzazione". Questa metamorfosi ha fatto emergere problematiche considerevoli nella prospettiva del diritto costituzionale considerando che le big tech, in quando operatori privati sono orientati al profitto economico e non alla tutela del pluralismo informativo o dal diritto ad una informazione veritiera, ma allo stesso tempo hanno progressivamente ottenuto il potere di poter limitare delle libertà fondamentali tra cui quella di espressione esercitando *de facto* dei poteri di natura para-costituzionale[58]. Rodotà individuava come mito fondativo della comunicazione via Internet l'agorà democratica di Atene e tale discorso vale a maggior ragione oggi, considerando che le piattaforme, sono diventati dei veri e propri forum pubblici in cui si sviluppa il dibattito politico e il confronto delle idee, costituendo

[53] R. Bin, *Critica della teoria dei diritti*, Milano 2018, 50.

[54] Come di fatto avvenuto con il parziale fallimento del primo Codice di contrasto alla disinformazione che secondo una certa dottrina era orientato da tale strategia.

[55] V. Zeno-Zencovich, *La libertà*, cit., 100.

[56] J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Milano, 2023.

[57] O. Pollicino, *Potere*, Cit., 422 ss.

[58] M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati". Spunti di comparazione*, in *Rivista italiana di informatica e diritto*, 2, 2021, 43.

quindi delle – seppur anomale - "agorà digitali". È evidente come non si possa lasciare a tali piattaforme, orientate dalla logica del *business*, la possibilità di operare la pratica del bilanciamento sui diritti fondamentali. I grandi operatori delle piattaforme digitali, veri proprietari della sfera pubblica, governano con strumenti algoritmici che se inizialmente sfuggivano a un controllo pubblico diretto oggi, nel più ampio cambio di paradigma regolatorio, che dopo una fase essenzialmente liberista passa ad una marcatamente costituzionale, che è stata definita come la seconda fase del costituzionalismo digitale affronta i pericoli in questione. Con l'entrata in vigore del DSA in particolare, l'UE ha affrontato il tema della limitazione del potere di tali soggetti transnazionali che operano orientati da scopo di lucro tentando di salvaguardare i diritti fondamentali degli utenti tra i quali libertà di espressione e il diritto all'informazione nella cornice del pluralismo informativo.

3. Il formante tecnologico: dalle *fake news* al *deepfake*

Con il progressivo prendere forma della rete, ai suoi albori, si affermò come maggioritaria, in quella che comunque era una piccola *elite* di pionieri del *web* una filosofia cyberlibertaria, che vedeva internet come uno strumento rivoluzionario che avrebbe comportato un avanzamento de processi di democratizzazione, cogliendone le opportunità, ma non intravedendo quelle che sarebbero potute diventare problematiche infiorescenze del sistema. Emblematica di questa *weltanshauung* è la *Dichiarazione di indipendenza del cyberspazio* in cui l'autore John Perry Barlow descriveva la rete come spazio franco dove gli stati non hanno sovranità, riflettendo il sentire di un mondo che si identificava con una visione libertaria quasi fino all'anarchismo. Tra le pagine di *Code and other laws of cyberspace* di Lessig si respira questo clima palpitante di entusiasmo presente negli ambienti universitari, nei centri di ricerca e in generale nella società, nel momento in cui il cyberspazio si converte nell'obiettivo delle utopie libertarie[59]. Internet è il più grande spazio comune che l'umanità abbia mai conosciuto e la sfida epocale che ha posto al mondo giuridico – oggi accelerata nell'era dell'Intelligenza Artificiale – è stata paragonata ad altre avvenute nella storia del diritto quando le regole hanno dovuto abbandonare il tradizionale e rassicurante riferimento alla terra e fare i conti con realtà mobili come il mare[60], passando dall'assodato nomos della realtà tellurica all'incerta e nuova questione della regolamentazione talassica[61]. Come all'epoca ci si trovò di fronte a un diritto nuovo modellato sulla natura delle cose che, liberato da vecchi schemi, si apriva a una stagione inedita, così è avvenuto anche per lo spazio giuridico digitale. Per Rodotà non è un caso che la metafora dello stare in rete sia quella – appartenente all'immaginario talassico – del navigare e che proprio nel diritto del mare abbiano cercato risposte coloro che per primi hanno dovuto affrontare le sfide istituzionali volte a garantire la libertà e la sicurezza di questo mare[62]. Internet venne esaltato dai cy-

[59] L. Lessig, *El codigo y otras leyes del ciberspacio*, Madrid, 2001, 21.

[60] A. Celotto, *L'età dei (non) diritti,* 2017, 115.

[61] S. Rodotà, *Il mondo della rete. Quali i diritti, quali i vincoli*, Roma-Bari, 2014, 4.

[62] Ivi, 5 ss.

berlibertari, soprattutto oltreoceano, come amplificatore delle libertà preesistenti, con particolare riferimento a quella di espressione e come potenziale *new free market place of ideas*. Ma l'illusione cyberlibertaria che vedeva nel *web* uno spazio franco non regolabile e ne metteva in risalto i soli aspetti positivi è stata smentita dalla storia, infatti il cyberspazio è stato oggetto di iper-regolazione ponendo delle sfide ad alta complessità al costituzionalismo, tra le quali – con Peter Häberle – bisogna annoverare proprio il diffuso aumento delle *fake news*[63].

Le problematiche poste da questo "dispositivo", onnipresenti nel dibattito culturale tanto da aver quasi monopolizzato il dibattito tra gli studiosi di diritto dell'informazione negli anni passati[64] sono oggi tornate alla ribalta durante la nuova primavera dell'IA ripresentandosi con nuovi profili di complessità sia sul piano qualitativo che quantitativo. Se da un lato tale espressione rischia di essere strumentalizzata, causando una eccessiva limitazione della libertà di espressione e rischiando di costituire uno dei numerosi strumenti che nella storia del pensiero politico hanno contribuito alla demonizzazione del nemico con la finalità di eliminarlo dall'arena delle libere opinioni, dall'altro, se totalmente lasciato al libero mercato delle idee, rischia di portare a un punto di non ritorno, con un impatto significativo alla luce della sua incidenza su una pluralità di diritti fondamentali. Nonostante l'utilizzo di tale espressione sia andato progressivamente incontro a varie critiche sia in ambito accademico che istituzionale[65], soprattutto in relazione ai rischi di strumentalizzazione oltre che per il rischio di escludere una serie di manifestazioni del fenomeno della disinformazione, reputiamo necessaria una sua ricognizione alla luce della loro centralità nell'economia del tema oggetto di indagine. Pur essendo la prospettiva privilegiata quella del diritto costituzionale (soprattutto sotto il profilo della limitazione della libertà di espressione e di quello al diritto ad una informazione veritiera) il tema in questione necessita di una pluralità di lenti analitiche che vanno dalla filosofia alle scienze della comunicazione passando per la sociologia, aprendosi ad una interdisciplinarità oggi sempre più necessaria nell'economia di uno studio sulla disinformazione, fenomeno altamente complesso che coinvolge molteplici realtà sociali.

Nella comunità dei linguisti non vi è una posizione unanime sulla qualificazione delle *fake news*. L'enciclopedia Treccani così definisce il neologismo: «Locuzione inglese […] entrata in uso nel primo decennio del XXI secolo per designare un'informazione in parte o del tutto non corrispondente al vero, divulgata intenzionalmente o inintenzionalmente attraverso il Web, i media o le tecnologie digitali di comunicazione, e caratterizzata da un'apparente plausibilità, quest'ultima alimentata da un sistema distorto di aspettative dell'opinione pubblica e da un'amplificazione dei pregiudizi che ne sono alla base, ciò che ne agevola la condivisione e la diffusione pur in assenza di una verifica delle fonti»[66]; specificandone l'uso prevalentemente politico e l'entrata nel lessico gior-

[63] P. Häberle, *Il costituzionalismo come progetto della scienza*, in *Nomos quadrimestrale di teoria generale, diritto pubblico comparato e storia costituzionale*, 1, 2018.

[64] M. Bassini- G. E. Vigevani, *Primi appunti su fake news e dintorni*, su questa *Rivista*, 1, 2017, 13.

[65] O. Pollicino – P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *Federalismi.it Rivista di diritto pubblico italiano, comparato, europeo*, 12, 2024, 5.

[66] Treccani, *Fake news*.

nalistico grazie all'impiego fattone da Trump nell'anno della sua prima elezione, oltreché la sua stretta connessione con la post-verità[67], che se per alcuni autori rappresenta una nozione confusa e poco utile[68] per descrivere il presente, altri per altri è importante a tal punto da definire le caratteristiche essenziali dell'opinione pubblica contemporanea, rintracciandovi, nel quadro di un forte appello all'emotività, il risultato di «un filone conservatore che ha trovato nel postmoderno la propria legittimazione filosofica e nel populismo la propria diffusione politica»[69]. La domanda centrale che gli addetti ai lavori si sono posti è se questa locuzione rappresenti una nuova modalità volta a definire processi di disinformazione che da sempre sono presenti nella lotta politica o il risultato di un lavoro di "ingegneria comunicativa" nuovo rispetto al passato[70]. Tra le loro caratteristiche fondamentali è stata infatti rilevata una capacità di influenzare soggetti con una velocità e un coinvolgimento inediti. Il concetto di informazione falsa è antico quanto la storia dell'umanità, mentre il termine *fake news* – di origine anglosassone – pur nascendo verso la fine del XIX secolo per descrivere una notizia inventata in ambito politico, esce dall'ambito specialistico degli addetti ai lavori e diventa popolare nelle elezioni americane del 2016, culminate con l'elezione di Donald Trump, rivelando la sua natura essenzialmente politica e una sua stretta relazione con il fenomeno populista. L'utilizzo di tale tecnologia modifica la percezione della realtà del soggetto influenzandolo a tal punto da spingerlo a condividere spontaneamente ed in tempo reale tali notizie; l'elemento principale che caratterizza le *fake news* moderne secondo gli psicologi della comunicazione sarebbe la capacità di impattare sui soggetti e sui gruppi sociali con una velocità e un coinvolgimento unici nella storia dell'informazione[71].
Ma se una storia falsa[72] – per dirla con Canfora – è sempre esistita, che cosa distingue le attuali *fake news* dalle tradizionali informazioni false? Oltre alla capacità di toccare la dimensione emotiva dello sciame digitale, popolo anonimo che vive sulla rete, vi è un elemento fondamentale costituito dal mezzo attraverso il quale le *fake news* vengono veicolate: l'utilizzo delle tecnologie digitali. Bisogna infatti considerare che il rapporto tra informazione e mezzo è tutt'altro che secondario, anzi potremmo dire con McLuhan che il "medium è il messaggio", poiché il modo in cui un *medium* organizza e struttura i propri contenuti non è neutrale, ma ne influenza in modo decisivo la ricezione e la comprensione[73]. I media digitali hanno sottoposto la sfera pubblica a un radicale cambio di struttura causandone la frammentazione; la sfera del discorso pubblico è stata minacciata dall'infodemia[74] suscitando nuove riflessioni alla luce

[67] M. Ferraris, *Post verità ed altri enigmi*, Bologna, 2017, 9.

[68] F. Paglieri, *La disinformazione felice. Cosa ci insegnano le bufale*, Bologna, 2020, 21.

[69] *Ibid.*

[70] G. Riva, *Fake news. Vivere e sopravvivere in un mondo post-verità*, Bologna, 2018, 15.

[71] *Ibid.*

[72] L. Canfora, *La storia falsa*, Milano, 2008.

[73] M. McLuhan, *Gli strumenti del comunicare*, Milano, 2008, 6.

[74] Secondo l'Accademia della Crusca per *Infodemia* si intende un abnorme flusso di informazioni di qualità variabile su un argomento, prodotte e messe in circolazione con estrema rapidità e capillarità attraverso i media tradizionali e digitali, tale da generare disinformazione, con conseguente distorsione della realtà ed effetti potenzialmente pericolosi sul piano delle reazioni e dei comportamenti sociali.

dell'inedita dimensione pervasiva dei fornitori di servizi digitali e ponendo domande rilevanti sull'adeguatezza degli strumenti disponibili per il diritto costituzionale nella società dell'algoritmo. Le questioni precedentemente sollevate hanno comportato un cambio di paradigma nelle scelte di politica del diritto dell'Unione Europea che, lasciatasi alle spalle l'impostazione marcatamente liberista, ha risposto alle sfide poste dallo spazio giuridico digitale attraverso un ripensamento delle proprie strategie regolatorie-[75]avviando una nuova stagione di costituzionalismo digitale[76]. Con tale espressione si intende: «il plesso di interventi legislativi dell'Unione volti a direttamente a regolare il fenomeno tecnologico e digitale, nelle sue varie forme. Al fine precipuo di salvaguardare e promuovere i valori intrinseci del costituzionalismo europeo»[77].

Il formante tecnologico – che potremmo definire anche formante algoritmico – è centrale nell'economia del presente saggio, poiché senza internet e i *social media* le *fake news* non avrebbero assunto la rilevanza che oggi hanno essendo la loro caratteristica fondamentale l'essere progettate e diffuse tramite il web[78].

Sul piano comunicativo, il fenomeno che si è venuto a creare è stato definito come disintermediazione, processo mediante il quale vengono eliminate le strutture di mediazione – come i corpi intermedi o i filtri – tra due o più utenti in un processo di comunicazione. La rivoluzione digitale ha infatti comportato un cambiamento radicale del rapporto lettore-*media* e la disintermediazione – costituente il terreno ideale per lo sviluppo delle *fake news* – ha fatto acquisire al cittadino un potere maggiore permettendogli di usufruire di un accesso diretto alle informazioni immergendosi in un flusso di dati che, oltre a poter controllare, può modificare creare e diffondere attraverso strumenti progressivamente sempre più alla portata di tutti. All'interno di tale ecosistema mediale completamente rivoluzionato si inseriscono le problematiche poste dall'amplificazione delle notizie false tramite sistemi di intelligenza artificiale. La disinformazione si fonda sul *confirmation bias* processo di accettazione delle informazioni che è legato alla tendenza di ogni individuo a conservare intatte le proprie credenze perturbandole il meno possibile. Tale processo si enfatizza notevolmente all'interno delle piattaforme digitali di grandi dimensioni, in cui gli spazi discorsivi vengono sostituiti da *echo-chambers,* termine che indica «quegli spazi che, sui media, determinano la creazione di uno stato di isolamento degli individui, in cui le informazioni, idee e credenze vengono amplificate e rafforzate all'interno di un sistema isolato»[79]e in cui le informazioni vengono amplificate da una ripetitiva trasmissione all'interno di un ambito omogeneo e chiuso, in cui visioni e interpretazioni divergenti finiscono per non trovare più considerazione, portando quindi l'utente a convalidare e rinforzare le proprie convinzioni e alimentan-

[75] O. Pollicino – P. Dunn, *Intelligenza Artificiale*, cit., 27.

[76] G. De Gregorio, *Digital Constitutionalism in Europe*, Oxford University Press, 2022.

[77] O. Pollicino – P. Dunn, *Intelligenza Artificiale*, cit., 30.

[78] T. Guerini, *Fake news e diritto penale. La manipolazione digitale del consenso nelle democrazie liberali*, Torino, 2020, 5. Il concetto di formante, notoriamente elaborato da R. Sacco, *Legal Formants: A Dynamic Approach to Comparative Law (Installment I of II)*, in *American Journal of Comparative Law*, 39(1), 1991, viene utilizzato nell'analisi delle *fake news* - in una prospettiva penalistica - da Tommaso Guerini; l'espressione "formante algoritmico" è presente in F. Sgubbi, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa. Venti tesi,* Bologna, 2019, 39.

[79] B.-C. Han, *Infocrazia. Le nostre vite manipolate dalla rete*, Torino, 2023, 39.

do la polarizzazione politica[80]. Gli algoritmi delle piattaforme, la cui rete epistemica di informazioni a cui siamo sottoposti è influenzata dalle reti sociali in cui siamo inseriti *online*, accelera questo fenomeno attraverso gli agenti di raccomandazione, che generando un fenomeno noto come *filter bubble* tendono a presentare contenuti sempre più simili a quelli che l'utente ha già consumato[81], richiamando l'idea di un pluralismo ovattato, filtrato attraverso le preferenze espresse da chi fruisce le informazioni[82].

I sistemi di raccomandazione permettono ai fornitori delle piattaforme di garantire agli utenti di essere raggiunti da contenuti maggiormente rispondenti ai loro gusti ed interessi ma anche al loro orientamento politico, giocando quindi un ruolo fondamentale nella diffusione di informazioni e di conseguenza anche nella formazione della coscienza pubblica, potendo influenzare le preferenze degli utenti e potenzialmente guidarne le scelte sia a livello individuale che collettivo[83]. Per descrivere tale nuovo assetto in cui il pubblico si dissolve in una miriade di bolle autoreferenziali in cui vengono rafforzati i meccanismi di polarizzazione la teoria politica ha utilizzato il concetto di *bubble democracy*[84]. Inoltre bisogna considerare lo sviluppo di tali sistemi sia orientato all'interesse economico e quindi al profitto; considerando che i contenuti polarizzanti (come le notizie false) tendono a catalizzare l'attenzione del pubblico di conseguenza un ulteriore profilo problematico messo in luce dalla dottrina è quello rappresentato dal fatto che l'algoritmo, pur di suscitare l'interesse degli utenti, non solo non è disincentivato a ridurre i contenuti disinformativi ma è incentivato a promuoverli[85].

Un ulteriore profilo riguarda i rischi derivanti dalla possibilità di organizzare flussi informativi a fini strategici. La struttura algoritmica delle piattaforme e la presenza di "*cyber-truppe*" di *troll* professionisti e *social bot* favoriscono la suddivisione degli utenti in reti costruite ad arte che possono condizionare gli esiti di battaglie elettorali, creando, secondo Byung-Chul Han, le condizioni strutturali per una degenerazione della democrazia in "infocrazia"[86]. Il diritto all'informazione, la cui elaborazione[87] dottrinale

---

[80] W. Quattrociocchi – A. Vicini, *Polarizzazioni. Informazioni, opinioni e altri demoni nell'infosfera*, Milano, 2023.

[81] N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, 2023, 144 ss.

[82] C. M. Reale – M. Tommasi, *Libertà*, cit., 331 ss.

[83] O. Pollicino – P. Dunn, *Disinformazione*, cit., 13.

[84] D. Palano, *Bubble democracy: la fine del pubblico e la nuova polarizzazione*, Brescia, 2020.

[85] O. Pollicino – P. Dunn, *Disinformazione*, cit., 13.

[86] La comunicazione nelle piattaforme digitali basate sugli algoritmi secondo Han, *Infocrazia*, cit., non è né libera de democratica; è una "comunicazione senza comunità" che rende impossibile la politica dell'ascolto e segna la fine dell'agire comunicativo. La presenza dell'altro è infatti fondamentale e costitutiva dell'agire comunicativo in habermassiano che ci obbliga a pensare l'altro come parlante e ascoltante, che riflette la propria visione del mondo nel mondo oggettivo, sociale o soggettivo attraverso affermazioni che possono essere accettate e discusse. La fine dell'agire comunicativo è determinata dalla scomparsa dell'altro – inteso come parlante e ascoltante - che segna la fine del discorso rafforzate dalla polarizzazione e dalla propaganda delle camere dell'eco che generano una personalizzazione algoritmica della rete fomentando una mentalità tribale in una guerra delle identità particolaristiche.
La comunicazione digitale – nella prospettiva piuttosto pessimistica dell'A. - disintegra la società e determina la crisi della verità divenendo indifferente qualsiasi concetto che aiuti a designare di una forma vincolante le cose facendogli perdere la sua funzione di regolatore sociale.

[87] Secondo Martines il fondamento costituzionale di tale diritto è da rinvenirsi principalmente nella

e giurisprudenziale risale a un momento storico in cui non si erano ancora sviluppate le nuove tecnologie, necessità in questo momento storico, secondo una certa dottrina, di una nuova configurazione che va oltre il tradizionale diritto ad essere informati per configurarsi come diritto a non essere disinformati .

L'ultima frontiera della disinformazione è costituita dal *deepfake*, neologismo derivante da *deep learning* e *fake*, insieme di tecniche che – attraverso *software* di Intelligenza Artificiale – partendo da contenuti reali riescono a modificare o ricreare in modo realistico le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una voce. Nella definizione fornita dal Garante: I deepfake sono foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce.

Le foto e i video *deepfake* sono prodotti da reti neurali profonde, interconnesse con strati di unità computazionali simili a neuroni[88], addestrabili tramite il *deep learning*, che regola le connessioni tra le unità in risposta al *feedback*. Il *deep learning* può addestrare le reti a svolgere molte attività, tra cui la produzione di immagini e video apparentemente veri, potendo mostrare qualcuno mentre fa qualcosa che non ha mai fatto e dice cose che non ha mai detto; la pericolosità di questo dispositivo dal punto di vista della disinformazione di massa o nel settore pornografico, dove si è principalmente sviluppato, è particolarmente alta potendo potenzialmente incidere su una pluralità di diritti fondamentali ed essendo particolarmente pericoloso soprattutto durante le campagne elettorali potendo costituire uno strumento di distorsione del dibattito pubblico e di manipolazione del consenso[89]. Il rischio è che queste tecnologie migliorino tanto che secondo David Chalmers[90] in futuro non potremmo distinguere video *deepfake* da quelli reali, non potendoci fidare delle immagini in maniera così diretta come abbiamo sempre fatto e probabilmente la loro rilevabilità sarà affidata a degli algoritmi avanzati. Bisogna inoltre considerare, come giustamente messo in luce dalla dottrina più recente, le minacce derivanti da una progressiva "democratizzazione"[91] degli strumenti per la realizzazione di contenuti *deepfakes*, sempre più accessibili a chiunque e non solo ad informatici esperti[92]. Lo scenario che si profila si caratterizza per una particolare pericolosità, potendo questo tipo di contenuti entrare a far parte anche della real-

democraticità dell'ordinamento. L'art. 21 Cost. oltre alla libertà di manifestazione del pensiero (libertà di dare e divulgare notizie, commenti) tutela secondo la sua lettura, dal punto di vista dei destinatari della manifestazione - sia pure indirettamente - l'interesse generale all'informazione che, in un regime democratico, implica pluralità di fonti, libero accesso alle medesime, assenza di ingiustificati ostacoli legali alla circolazione delle notizie e delle idee. T. Martines, *Diritto costituzionale*, Milano, 1992, 679.

88 D. J. Chalmers, *Più realtà. I mondi virtuali e i problemi della filosofia*, Milano, 2023, 303.

89 Per una ricognizione dei principali casi di *deepfakes* disinformativi si veda M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes,* in questa *Rivista,* 171.

90 D. J. Chalmers, *Più realtà*, cit., 304 ss.

91 E. Meskys - J.Kalpokiene - P. Jurcys- A.Liaudanskas, *Regulating Deep Fakes: Legal and Ethical Considerations*, in *Journal of Intellectual Property Law & Practice*, 15 (1), 2020, 24.

92 M. Cazzaniga, *Una nuova tecnica*, cit., 175.

tà aumentata o virtuale[93]. La direzione sembra essere quella di una verità artificiale, dentro un ecosistema in cui vengono diffusi massivamente contenuti creati tramite sistemi di intelligenza artificiale e a sua volta, questo ecosistema di tecnologie diviene sempre più necessario per verificare la veridicità dei contenuti, costituendo un fattore sempre più essenziale nella lotta alla disinformazione su internet venendo già adesso utilizzata dalle piattaforme online per individuare profili falsi (*bot* o *troll*) oltre ad essere utilizzata per individuare contenuti falsi sintetici o manipolati attraverso l'IA stessa, è infatti particolarmente diffusa la pratica di allenare modelli *deep learning*, fondati su reti neurali convoluzionali[94]. Un profilo problematico che emerge con riferimento a questa tendenza – e che va oltre, aggiungendosi, alla tradizionale problematica della rimozione dei contenuti o del *fact-checking* umano - è che l'intelligenza artificiale applicata a sistemi di moderazione e cura dei contenuti sia esposta a errori e *bias* a danno, soprattutto di minoranze, con significativi impatti sull'effettiva promozione di un ecosistema informazionale pluralistico[95].

4. La regolazione europea dell'Intelligenza Artificiale

Quando si parla di intelligenza artificiale si utilizza spesso la metafora del passare delle stagioni. È infatti un campo che ha visto vari cicli di espansione e contrazione, con periodi di grandi aspettative e altri di investimenti sensibilmente ridotti[96]. Questa volatilità, tipica nelle nuove tecnologie, caratterizza particolarmente la storia dell'intelligenza artificiale, i cui periodi caratterizzati da meno interesse e investimenti, sono definiti inverni. Basti considerare il fatto che questa espressione è stata coniata da John McCarthy nella proposta di finanziamento per la conferenza di Dartmouth del 1956, raduno dei ricercatori di quello che all'epoca era un nuovissimo campo di ricerca. Da allora ha transitato attraverso varie stagioni, passando attraverso diversi paradigmi filosofico-tecnologici e la stagione che stiamo attraversando possiamo considerarla a tutti gli effeti come una nuova primavera dell'intelligenza artificiale.
La regolazione del ricorso a sistemi decisionali automatizzati e all'algoritmo ha progressivamente assunto un ruolo sempre più centrale nel contesto delle politiche dell'UE sin dalla metà degli anni 2010 e, in tal senso, il GDPR rappresenta il capostipite della strategia euro-unitaria di *governance* di tali sistemi[97]. Ma tale regolamento pur avendo introdotto una importante previsione che costituisce una "riserva di umanità" con riferimento alla soggezione a decisioni prese secondo modalità automatizzate tratta tale fattispecie come residuale rispetto al trattamento umano e manca di tenere conto delle applicazioni di IA fondate sui *big data* non prendendo ancora in considerazione il fenomeno dei trattamenti di dati di massa[98]. Bisogna considerare che il GDPR è stato

93 D. J. Chalmers, *Più realtà*, cit., 304 ss.

94 O. Pollicino – P. Dunn, *Disinformazione*, cit. 14.

95 Ivi, 150.

96 N. Cristianini, *La scorciatoia*, cit., 125 ss.

97 O. Pollicino – P. Dunn, *Intelligenza artificiale*, cit., 59.

98 G. Finocchiaro, *Intelligenza artificiale. Quali regole?,* Bologna, 2024, 86.

approvato nel 2016 e che negli ultimi anni abbiamo assistito ad una forte accelerazione nello sviluppo dei sistemi di IA ed a livello europeo la risposta a tali mutamenti è arrivata con l'adozione dell'AI act a giugno 2024.

Con riferimento al tema regolazione dell'algoritmo e dei processi decisionali automatizzati e dell'AI occupa uno spazio centrale il DSA con cui il legislatore europeo si dimostra consapevole dell'ormai centrale ruolo ricoperto da tali strumenti nella governance dei contenuti in rete[99], provando ad "imbrigliare il potenziale del potere computazionale dei provider e dei loro sistemi algoritmici allo scopo di promuovere gli interessi ed i valori democratici dell'Unione"[100]. L'entrata in vigore del Digital Service Act e le misure finalizzate alla sua applicazione a quelle imprese designate dalla Commissione come *Very Large Online Platforms* e *Very Large Online Search Engines* , con riferimento ai loro rischi sistemici derivanti dai loro servizi e la conseguente istituzione del *European Center for Algoritmich Transparency*, a Siviglia, presso il *Joint Research Centre* della Commissione europea, vanno nella direzione della seconda stagione del costituzionalismo digitale, in cui il legislatore europeo nell'ambito di un più ampio processo di riaccentramento delle fonti si riappropria del ruolo di *law maker* attraverso "iniezioni" di garanzie procedurali costituzionalmente orientate con il fine di limitare l'influenza dei poteri privati e prevenirne gli abusi. Con il DSA[101] le piattaforme e i motori di ricerca con un "raggio d'azione" di oltre 45 milioni di utenti attivi mensili – corrispondenti al 10% dei consumatori europei – in conformità agli obblighi previsti dal Regolamento sui servizi digitali dovranno analizzare e valutare i rischi sistemici derivanti dai loro servizi, che vanno dalla diffusione e amplificazione di contenuti illeciti e disinformativi fino alle eventuali ripercussioni sulle libertà di espressione e di informazione, oltre ai rischi relazionati con la violenza di genere e la sicurezza dei minori online. Un ulteriore aspetto peculiare del Regolamento è il transito da un paradigma autoregolativo a un sistema di co-regolamentazione nel contesto della regolazione dei contenuti e nella mitigazione dei rischi sistemici stabilito dall'art. 45. Il nuovo Codice rafforzato del 2022 è il primo modello di codice di condotta che interviene per realizzare tale strategia di co-regolamentazione costituendo un ulteriore tassello nella promozione del costituzionalismo digitale con riferimento all'intersezione tra libertà di espressione e intelligenza artificiale[102]. Se il Codice di buone pratiche sulla disinformazione del 2018 era costituito da una serie di impegni che si risolvevano in scarne petizioni di principio, il nuovo Codice rafforzato, che costituisce anch'esso un atto di *soft law*, prevede 44 impegni e 128 misure specifiche in settori come la demonetizzazione della diffusione della disinformazione, la trasparenza della pubblicità politica, la responsabilizzare gli utenti e la cooperazione con i verificatori dei fatti. Un aspetto importante di tale Codice è rappresentato dall'implicito riconoscimento della centralità dei sistemi automatizzati di moderazione dei contenuti[103] tanto nel contrasto alla disinformazione quanto nella

[99] O. Pollicino – P. Dunn, *Intelligenza artificiale*, cit., 48.

[100] Ivi*,* 55.

[101] M. A. Aranzazu Toquero, *Entre Scilla y Caribdis: los intermediarios digitales y la moderación de contenidos*, in *Rivista AIC Associazione italiana dei Costituzionalisti*, 4, 2023.

[102] *Ibid.*

[103] *Ibid.*

disseminazione di contenuti falsi. Con il fine di limitare la circolazione di *fake news*, contiene delle previsioni anche in relazione al *deepfake*, in particolare il *commitment* numero 14 che stabilendo che i sottoscrittori devono mettere a punto, o implementare se già ne avevano predisposte prima della sottoscrizione del nuovo codice, una serie di *policies* di contrasto alla disinformazione veicolata attraverso tali tecnologie[104]. Mentre l'impegno 15 richiede ai firmatari i quali sviluppino o operino sistemi di IA e diffondano attraverso i loro servizi contenuti generati o manipolati attraverso l'uso di IA (come il *deepfake*) di applicare diligentemente le regole concernenti gli obblighi di trasparenza previste dall'*AI Act*. Le osservazioni della dottrina hanno messo in luce come con riferimento al tema dei *deepfakes* l'adozione del codice non pare essere stata particolarmente influente orientando in maniera decisiva le scelte delle piattaforme in tale settore[105]. Costituendo di fatto il primo tentativo al mondo di regolazione di questa tecnologia, l'Europa si afferma come leader a livello globale sotto il profilo della regolamentazione dell'Intelligenza Artificiale e di un suo sviluppo etico e affidabile nella prospettiva del rispetto dei diritti fondamentali[106]. È una sfida profondamente ambiziosa che si caratterizza per la sua particolare complessità: in primis quella di dover regolare, a livello continentale, una tecnologia che per sua natura è globale. Ulteriore difficoltà è quella di "inseguire" lo sviluppo tecnologico di un ecosistema in continua evoluzione, come ha dimostrato il caso dell'intelligenza artificiale generativa, inizialmente non prevista nella proposta di regolamento.

La complessità di questa sfida epocale è costituita inoltre dalla necessità di proteggere i diritti fondamentali e i valori europei, ma al tempo stesso non frenare – attraverso un eccesso di regolazione – sproporzionatamente lo sviluppo tecnologico, la libertà di iniziativa economica, la libertà di scienza e tecnica e il mercato, rischiando di aumentare eccessivamente il costo dell'immissione sul mercato di tecnologie IA. Questo è lo scopo principale che si prefigge il regolamento: promuovere lo sviluppo di una intelligenza artificiale antropocentrica garantendo un elevato livello di protezione della salute, dell'ambiente e della sicurezza. Lo sviluppo di un ecosistema di fiducia dentro un quadro giuridico per una intelligenza artificiale affidabile è in perfetta coerenza con gli obiettivi della Commissione 2019-2024 e con la pubblicazione nel 2020 del *Libro bianco sull'Intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia,* dove venivano delineati i principi fondamentali di un futuro quadro normativo dell'UE per l'IA, basato sui valori e sui diritti fondamentali europei ed avente l'obiettivo di garantire una tutela della persona e al tempo stesso incoraggiare le imprese a sviluppare tali sistemi, ma sempre nel rispetto di questi valori.

La scelta del regolamento come atto giuridico è giustificata dalla necessità di un'appli-

[104] M. Cazzaniga, *Nuove tecniche*, cit., 179.

[105] Per un'analisi del Report presentato il 9 febbraio 2023 dai firmatari. *Ivi*, 180

[106] Il primato europeo riguarda la capacità di elaborazione di requisiti legali finalizzati alla protezione di una intelligenza artificiale a dimensione antropocentrica, non un primato sul suo sviluppo tecnologico. L'esplosione di ricerca, sviluppo e commercializzazione dell'IA, con particolare riferimento all'apprendimento automatico, è come abbiamo detto, globale, ma si è concentrata in larga misura negli Stati Uniti e in Cina. Oggi è in atto una corsa tra queste due potenze per ottenere il vantaggio strategico nell'ambito dell'IA, come emerge dallo studio di H.E. Kissinger – E. D. Schmidt - D.Huttenlocher, *L'era dell'intelligenza artificiale,* cit., 70.

cazione uniforme delle nuove regole su questa tecnologia, poiché l'applicabilità diretta del regolamento dovrebbe ridurre i rischi di una frammentazione giuridica e facilitare lo sviluppo di un mercato unico per tutti quei sistemi che sono leciti e affidabili. Normative nazionali divergenti potrebbero infatti determinare una rischiosa frammentazione del mercato interno e allo stesso tempo diminuire la certezza del diritto per gli operatori che sviluppano o utilizzano sistemi di intelligenza artificiale.

Parte della dottrina[107] ha sostenuto che di fatto si tratta di una "direttiva mascherata" come in parte lo è stata anche il GPDR, perché vi sono delle clausole aperte che lasciano un margine discrezionale agli Stati membri e che quindi aprono a possibili scelte strategiche diverse.

Il regolamento europeo sull'Intelligenza Artificiale[108], finalmente pubblicato nella Gazzetta Ufficiale dell'UE il 12 luglio ed entrato in vigore il 1 agosto 2024, costituisce il primo tentativo al mondo di regolazione di questo ecosistema di tecnologie, pur avendo un campo applicativo più amplio rispetto a quello dello del contrasto alla disinformazione e della tutela della libertà di espressione e di informazione, ha introdotto una norma importante finalizzata a disciplinare l'utilizzo dei sistemi *deepfake*. Nella definizione fornita nell'art. 3, 60) del *Ai Act* per *deep fake* si intende: «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona». L'art. 50 che disciplina gli obblighi di trasparenza per i fornitori e i deployers di determinati sistemi di IA stabilisce che i fornitori che garantiscono che i sistemi di IA destinati a interagire direttamente con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche interessate siano informate del fatto di stare interagendo con sistema di IA, a meno che ciò non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo. Tale obbligo non si applica ai sistemi di IA autorizzati dalla legge per accertare, prevenire, indagare o perseguire reati, fatte salve le tutele adeguate per i diritti e le libertà dei terzi, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato.

I fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici, garantiscono che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente. I fornitori garantiscono che le loro soluzioni tecniche siano efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche. Per poi successivamente delineare delle eccezioni a questo dovere di etichettamento: tale obbligo infatti non si applica se i sistemi di IA svolgono una funzione di assistenza per l'editing standard o non modificano in modo sostanziale i dati di input forniti dal deployer o la rispettiva semantica, o se autorizzati dalla legge ad accertare, prevenire, indagare o perseguire reati. I deployer di un sistema di riconoscimento delle emozioni o di un sistema di ca-

[107] G. Rutelli, *L'AI Act sarà una direttiva mascherata. Dialogo con Pollicino (Bocconi),* in *Formiche.*

[108] Regolamento (UE) 2024/1689, art. 50.

tegorizzazione biometrica informano le persone fisiche che vi sono esposte in merito al funzionamento del sistema e trattano i dati personali in conformità dei regolamenti (UE) 2016/679 e (UE) 2018/1725 e della direttiva (UE) 2016/680, a seconda dei casi. Tale obbligo non si applica ai sistemi di IA utilizzati per la categorizzazione biometrica e il riconoscimento delle emozioni autorizzati dalla legge per accertare, prevenire o indagare reati, fatte salve le tutele adeguate per i diritti e le libertà dei terzi e conformemente al diritto dell'Unione.

I *deployer* di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un "*deep fake*" devono rendere noto che il contenuto è stato generato o manipolato artificialmente. Tale obbligo non si applica nei casi in cui l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati o qualora il contenuto faccia parte di un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi, gli obblighi di trasparenza di cui al presente paragrafo si limitano all'obbligo di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera. I *deeployer* di un sistema di IA che genera o manipola testo pubblicato allo scopo di informare su questioni di interesse pubblico devono rendere noto che il testo è stato generato o manipolato artificialmente. Anche in questo caso l'obbligo non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati o se il contenuto generato dall'IA è stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione di quel determinato contenuto.

Secondo il *consideranda* 134 oltre alle soluzioni tecniche utilizzate dai fornitori del sistema di IA, i *deployer* che utilizzano un sistema di IA per generare o manipolare immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi, entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri, dovrebbero anche rendere noto in modo chiaro che il contenuto è stato creato o manipolato artificialmente etichettando di conseguenza gli output dell'IA e rivelandone l'origine artificiale. L'adempimento di tale obbligo di trasparenza non deve però ostacolare il diritto alla libertà di espressione e il diritto alla libertà delle arti e delle scienze garantito dalla Carta, in particolare quando il contenuto fa parte di un'opera o di un programma manifestamente creativo, satirico, artistico, fittizio, o analogo fatte salve le tutele adeguate per i diritti e le libertà dei terzi. In tali casi, l'obbligo di trasparenza per i *deep fake* si limita alla rivelazione dell'esistenza di tali contenuti generati o manipolati in modo adeguato che non ostacoli l'esposizione o il godimento dell'opera, compresi il suo normale sfruttamento e utilizzo, mantenendo nel contempo l'utilità e la qualità dell'opera. Dobbiamo infatti considerare che i sistemi *deepfake* presentano anche delle opportunità rappresentate da alcune sue applicazioni positive nel settore cinematografico, in quello pubblicitario o nella realizzazione di opere d'arte. Possono rappresentare uno strumento attraverso cui manifestare liberamente il proprio pensiero, nelle sue varie declinazioni come la parodia o la satira, creando quindi dei contenuti ricompresi nell'alveo di protezione della libertà di espressione[109]. Può essere inoltre utile ricordare che la formulazione volutamente generica dell'art. 21 della Costituzione italiana, stabilendo

[109] M. Cazzaniga, *Una nuova tecnica*, cit., 177.

che tutti hanno diritto di manifestare liberamente il proprio pensiero con la parola, lo scritto e « ogni altro mezzo di diffusione[110] », ricomprende astrattamente, attraverso la scelta di una formulazione volutamente amplia, anche la libera manifestazione del pensiero tramite questo ecosistema di nuove tecnologie rappresentato dall'IA. La dottrina ha recentemente messo in luce come essendo molti i settori in cui il *deep fake* può essere utilizzato con finalità meritevoli in una prospettiva costituzionale prospettare un divieto assoluto di ricorrere a questa tecnologia sarebbe sproporzionato e una proibizione totale limiterebbe eccessivamente la libertà di manifestazione del pensiero[111] e limiterebbe la sperimentazione in molti ambiti.

Scopo della norma, alla quale si accompagna – ricalcando il modello co-regolativo del Digital service act – la possibilità di elaborare codici di buone pratiche a livello dell'unione per facilitare l'efficace attuazione degli obblighi relativi alla rilevazione e all'etichettatura dei contenuti generati o manipolati artificialmente è quello di ridurre l'impatto che l'intelligenza artificiale ha a livello informativo sui cittadini dell'UE mirando a ridurre i rischi di cattiva informazione e manipolazione su vasta scala.

L'ambito politico, con particolare riferimento al momento elettorale, è sia il terreno prediletto dove si inserisce il fenomeno della disinformazione, sia quello ontologicamente fondativo della libertà di espressione. I sistemi di IA, operanti sulla base di grandi raccolte di dati, influenzano progressivamente sempre più aspetti relativi a questo campo, come la struttura dei messaggi politici e la loro diffusione a diversi settori della popolazione. L'IA viene infatti utilizzata per elaborare campagne di disinformazione e a sua volta viene utilizzata «per individuare, identificare e controbattere la disinformazione»[112]. Con il crescere di tali funzioni nel definire e plasmare l'ecosistema dell'informazione sarà sempre più difficile prevedere l'uso di questi sistemi, a essere minacciate saranno sempre di più le prospettive di una libera società e secondo una certa lettura persino del libero arbitrio[113]. La natura globale del digitale unite alla capacità dell'IA di monitorare, bloccare, conformare, produrre e distribuire informazioni sulle piattaforme di rete in tutto il mondo riportano tali complessità allo spazio dell'informazione delle società più diverse[114] e con il suo progressivo perfezionamento contribuisce a plasmare gli ordinamenti sociali su scala nazionale e globale. È una sfida complessa che richiede cautela, considerando che qualsiasi tipo di approccio nella regolazione di tale fenomeno riflette giudizi di valore, tanto che potremmo definirlo un dilemma complesso con risposte imperfette che variano a seconda della visione del mondo e dalla tradizione costituzionale di riferimento. L'anno appena[115] appena attraversato è stato uno dei più elettorali di sempre con oltre 50 elezioni nel mondo, con 76 paesi alle urne e 2 miliardi di persone chiamate al voto e non si possono sottovalutare i rischi che possono emergere dall'amplificazione delle tecniche di disinformazione attraverso

[110] Art. 21 Cost.

[111] M. Cazzaniga, *Una nuova tecnica*, cit., 177.

[112] H.E. Kissinger – E. D. Schmidt - D. Huttenlocher, *L'era dell'intelligenza artificiale. Il futuro dell'identità umana*, Milano, 2023, 19.

[113] *Ibid.*

[114] Ivi, 98.

[115] O. Pollicino – P- Dunn, *Disinformazione*, cit., 4.

l'impiego di quell'ecosistema costituito dall'intelligenza artificiale. Nel contesto politico, che come abbiamo visto è quello che riguarda principalmente la disinformazione, secondo la dottrina i video *deepfake* porteranno inevitabili interferenze nella formazione del consenso e più in generale, nel sistema democratico[116]. Quando questi sistemi riescono a rappresentare personaggi politici possono infatti causare sensibili variazioni nelle opinioni dell'elettorato, compromettendone i diritti di autodeterminazione informativa e di libertà decisionale[117].

5. Conclusioni

Possiamo provare a tirare le file della nostra riflessione sul tema libertà di espressione e verità (artificiale) nella società dell'algoritmo. Nel ragionare sul fenomeno in questione abbiamo rilevato come al tema ormai classico rappresentato dalla regolazione delle *fake news* e quindi dei profili costituzionalmente problematici causati dal rischio di una eccessiva o meno eccessiva limitazione – a seconda della tradizione costituzionale di riferimento – si aggiungano dei profili innovativi rappresentati dal come le varie forme di IA interagiscano o interferiscano con la libertà di manifestazione del pensiero, problematizzando ulteriormente il tema oggetto d'indagine tanto sotto il profilo quantitativo quanto sotto il profilo qualitativo. La filosofia della tecnologia ha dimostrato da tempo la non neutralità dello strumento tecnologico. È emersa nel dibattito la consapevolezza che l'intelligenza artificiale applicata a sistemi di moderazione e cura dei contenuti sia esposta a errori e *bias* a danno, soprattutto di minoranze, con significativi impatti sull'effettiva promozione di un ecosistema informazionale pluralistico. Abbiamo condiviso l'impostazione di quella dottrina secondo cui la disinformazione si caratterizza per la sua natura essenzialmente politica (la lesione di beni individuali come l'onore e la reputazione sono una eventualità o effetto collaterale ma non il reale obiettivo delle strategie di disinformazione) e l'obiettivo principale delle strategie di disinformazione è l'alterazione del funzionamento delle istituzioni democratiche con particolare riferimento alla formazione del consenso degli elettori. Di qui la forte rilevanza sotto il profilo giuspubblicistico dello sviluppo di questi sistemi in chiave antropocentrica nel rispetto dei diritti fondamentali, nella consapevolezza dei rischi derivanti da un suo utilizzo a scopi disinformativi da un lato e nella lotta alla disinformazione e nella moderazione dei contenuti dall'altro. Abbiamo inoltre aderito alla più recente dottrina italiana secondo cui l'utilizzo di sistemi *deepfake* non può considerarsi illecito in quanto tale ma solo quando determini la lesione di beni giuridici di rilievo costituzionale. Una analisi dei formanti delle *fake news* e del *deepfake* ha rivelato un mosaico complesso caratterizzato da una centralità dell'elemento tecnologico, i cui rischi oggi sono principalmente rappresentati dall'ecosistema dell'intelligenza artificiale. Il formante algoritmico rivela inoltre l'inadeguatezza e l'inapplicabilità della metafora del *free market place of ideas* come opzione di politica del diritto e strategia di contrasto alla disinformazione. Considerando che l'utilizzo di video deepfake e più in generale dei

[116] V. Azzali – N. Ellecosta, *La questione deepfake in Italia: una panoramica, in questa Rivista,* 3, 2023, 79 ss.

[117] *Ibid.*

sistemi di intelligenza artificiale a scopo disinformativo è destinata ad aumentare oltre che a diventare sempre più sofisticata, per spostarsi inoltre nel metaverso aprendo una nuova parentesi del dibattito, in parte già cominciato, sul *free speech*. Nell'attuale ecosistema dell'informazione caratterizzato dalla piattaformizzazione della sfera pubblica a prevalere non è l'informazione veritiera né il pluralismo informativo, ma informazioni false e polarizzanti, camere dell'eco e filtri bolla in cui l'utente è "isolato". Ho inoltre aderito a quella dottrina che da tempo sostiene l'inadeguatezza dell'importazione di questa metafora in prima luogo poiché dal punto di vista strettamente giuridico vi è una differenza sostanziale tra la tradizione costituzionale di provenienza della metafora, dove vi è una maggiore valorizzazione del profilo attivo della libertà piuttosto che a quello passivo, che come abbiamo visto non significa che sia una libertà illimitata, perché nessuna libertà dal punto di vista giuridico può esserlo, neanche quella di espressione, forse la più importante tra le libertà, come venne rilevato in una storica sentenza della Corte costituzionale italiana. La tradizione costituzionale del continente europeo dove vi è una differente visione del *freedom of speech* con notevoli conseguenze dal punto di vista filosofico e di politica del diritto. Alle complessità delle sfide poste dall'attuale momento storico per l'ecosistema dell'informazione, tra cui un ruolo rilevante lo rivestono proprio i rischi derivanti dall'utilizzo dell'IA nel nuovo ecosistema mediale, le strategie con cui l'UE le sta affrontando vanno nella direzione di una maggiore tutela dei diritti degli utenti con particolare riferimento alla libertà di espressione e diritto all'informazione. Le scelte di politica legislativa dell'UE vanno nella direzione della seconda stagione del costituzionalismo digitale, in cui il legislatore europeo si riappropria del ruolo di *law maker* attraverso "iniezioni" di garanzie procedurali costituzionalmente orientate con il fine di limitare l'influenza dei poteri privati e prevenirne gli abusi[118]. Nell'ambito di questo processo europeo di riaccentramento[119]delle fonti rientra senza dubbio anche l' entrata in vigore dell'*AI Act* e i Regolamenti sulla trasparenza e il targeting della pubblicità politica e sulla libertà dei media. Questa operazione di recupero della centralità delle fonti può consentire all'UE di affrontare correttamente le attuali sfide poste dal potere digitale[120], tra cui rientra il fenomeno della disinformazione. Il dilemma è complesso e dalle risposte imperfette perché – a fronte di un ecosistema di tecnologie dalla natura globale – le opzioni di politica del diritto, soprattutto in materia di disinformazione(nel cocktail esplosivo disinformazione-IA), dipendono fortemente dalle tradizioni costituzionali di riferimento e dai giudizi di valore, il terreno della questione inoltre è fortemente politico. A ciò bisogna aggiungere la natura ontologicamente tanto dinamica quanto antagonistica della libertà di manifestazione del pensiero, la cui dimensione non è immutabile e dipende da una serie di fattori politici, prestandosi non solo ad ampiamenti ma anche a riduzioni[121].

Come sostenuto da Luciano Floridi, la nostra è l'ultima generazione a fare esperienza

[118] O. Pollicino, *Potere*, cit., 438.

[119] C. Colapietro, *La proposta di Artificial Intelligence Act: quali prospettive per l'Amministrazione digitale*?, in *CERIDAP Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche*, fascicolo speciale, 1, 2022, 3.

[120] A. Iannuzzi, *La "governance" europea dei dati nella contesa per la sovranità digitale. Un ponte verso la regolazione dell'intelligenza artificiale*, in *Studi parlamentari e di politica costituzionale*, 2021

[121] V. Zeno-Zencovich, *La libertà*, cit., 159 ss.

della chiara distinzione tra ambienti *online* e *offline*, dicotomia che tende progressivamente a svanire venendo sostituita da una forma di vita *Onlife*[122]. Mai come adesso il costituzionalismo è chiamato a difendere i tradizionali diritti e le libertà fondamentali nell'era dell'intelligenza artificiale dai rischi che essa comporta, senza rinunciare alle numerose opportunità offerte da questo ecosistema di nuove tecnologie. Il costituzionalismo digitale, concetto in divenire, può rappresentare la bussola attraverso cui orientarsi in questi tempi complessi, senza la disperazione degli apocalittici e l'eccessivo entusiasmo degli integrati[123], ma con la giusta dose di pessimismo dell'intelligenza e ottimismo della volontà.

[122] L. Floridi, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, 2017, 107.

[123] U. Eco, *Apocalittici e integrati. Comunicazione di massa e teorie della cultura di massa,* 1977, 361. Con la lungimiranza che lo caratterizzava l'A. scrisse: «Oggi noi viviamo in un universo dell'informazione; lo sviluppo tecnologico ha fatto sì che se dialogo e cultura potranno ancora sopravvivere (e c'è chi ne dubita) tutto questo non avverrà che sullo sfondo di una comunicazione intensiva di dati, di notizie, di aggiornamenti circa ciò che sta accadendo».

IA e moderazione dei contenuti sui social media: il principio del '*Human in the loop*' nel campo del diritto all'informazione e alla comunicazione*

Matteo Paolanti

Abstract

L'introduzione delle tecnologie digitali nella vita di tutti i giorni ha fatto sì che anche i diritti fondamentali della persona venissero toccati dal progresso tecnologico. In particolare, la libertà di manifestazione del pensiero si è scontrata con il sorgere delle piattaforme social, che ha reso ancor più frastagliata e complessa la gestione pratica di questo diritto. I proprietari dei media digitali, per controllare le loro stesse creazioni, hanno cercato di approntare soluzioni diverse che comprendessero l'utilizzo sia di forze umane che informatiche. Tuttavia, con il passare del tempo e con l'evoluzione tecnica, sembra che il vento spinga sempre di più verso l'abbandono del controllo umano a favore di una totale automazione. Nel prosieguo del paper si spiega come si è giunti a questo momento fatidico, ripercorrendo la breve storia che fa da cornice alla materia della moderazione dei contenuti su piattaforma e si analizzeranno le strategie che i legislatori hanno messo a punto affinché il principio umanistico, anche dinanzi al progresso tecnologico, non sia messo da parte e, anzi, sia rafforzato.

The introduction of digital technologies has meant that the fundamental rights have also been affected by progress. In particular, the freedom of speech clashed with the rise of social platforms, which made the horizon of this right even more jagged. The owners of digital media, in order to look after their own creations, have tried to come up with different solutions that include the use of both human and cyber forces. However, with the passage of time and technical evolution, the wind seems to be blowing more and more towards the abandonment of human control in favour of total automation. The paper will explain how this fateful moment has been reached, retracing the brief history that frames the subject of platform content moderation, and will analyse the strategies that legislators have developed so that the humanistic principle, even in the face of progress, is not sidelined but rather strengthened.

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

Sommario

Keywords

piattaforme digitali – libertà di manifestazione del pensiero – social network – algoritmi – diritto UE

1. La manifestazione del pensiero al centro della rivoluzione digitale

Nell'epoca attuale, la dimensione digitale[1] dell'essere umano si dimostra di grande importanza sia dal punto di vista sociale che giuridico. Da quando sono entrati nelle vite dei comuni cittadini – poco più che venti anni fa – i nuovi dispositivi elettronici (come smartphone, computer dotati di connessione veloce e i loro applicativi), è indubbio che qualcosa sia cambiato non solo nella vita di tutti i giorni ma anche nel modo di concepire ciò che fa parte della realtà.

Nel contesto di queste affermazioni, anche i diritti non sono esenti dagli effetti del cambiamento. Non a caso si può rimandare la memoria a tutte quelle categorie degli studi giuridici che sono state investite dall'onda del progresso tecnologico e che hanno dovuto ripensare ai propri principi per adeguarsi ai tempi che mutano[2].

Ugualmente, per quanto concerne il diritto costituzionale, ci si chiede se i risalenti schemi e le categorie ormai consolidate siano tuttora valide; infatti, com'è noto, il problema del diritto costituzionale non può essere costantemente mantenuto in quel c.d. "regno delle idee" astratto ma va necessariamente reso in un risultato tangibile[3].

[1] In questo passo si vuole sottolineare come, in contatto con la visione marcusiana, si renda necessario distinguere una dimensione digitale autonoma internalizzata dall'essere umano che vive la contemporaneità. Cfr. H. Marcuse, *L'uomo a una dimensione: l'ideologia della società industriale avanzata*, Torino, 1999.

[2] Senza avere la presunzione di voler essere esaustivi, ma guardando alle casistiche più affascinanti, basti andare con la memoria a come, attraverso la tecnologia, sia cambiato il diritto dei contratti (si pensi alla tecnologia *blockchain),* oppure come sia stato investito il diritto tributario dalla nuova realtà dei beni immateriali digitali (come NFT e criptovalute), ovvero ancora alle prime avvisaglie di quella che sarà la questione dell'eredità digitale, ossia la materia successoria di tutti gli asset personali (tra cui anche i profili su piattaforma) presenti in Rete.

[3] A questo riguardo, la parte di principio della Carta fondamentale non può essere considerata uno scudo dietro il quale nascondersi dinanzi alle vicende attuali della società ma, anzi, va utilizzata per

Nell'orizzonte fattuale fin qui descritto si rende emblematico il caso del diritto alla manifestazione del pensiero, dal quale discende come naturale conseguenza anche il diritto ad una corretta informazione.

La nuova acquisita rilevanza nel dibattito di questa libertà fondamentale, rispetto all'avvento della società digitale, va ricercata nell'affermazione del fenomeno dei *social media*. In breve, per tali s'intendono tutte quelle piattaforme digitali che predispongono un sistema informatico di servizi finalizzati all'incontro tra utenti diversi con il fine di riunire le persone intorno ad un certo ideale o interesse specifico[4].

Proprio all'interno di queste realtà la libertà di pensiero si è andata a scontrare con il problema delle notizie false e del c.d. *hate speech,* ossia di tutte quelle condotte discriminatorie che possono essere realizzate attraverso il libero uso della parola.

Dinanzi al sorgere di queste criticità gli stessi gestori delle piattaforme hanno cercato di predisporre delle soluzioni di tipo diverso, ciascuna contrassegnata dalle linee guida che rispecchiano le convinzioni e le idee degli stessi imprenditori digitali. A questo riguardo, non si può sottacere il fatto che tutto ciò abbia condotto a incidenti o veri e propri cortocircuiti in relazione ai diritti fondamentali tipici degli ordinamenti liberal-democratici. Tuttavia, anche gli stessi Stati nazionali non hanno potuto disconoscere la realtà che si era venuta a creare e, a loro modo, hanno provato a immaginare risposte legislative atte alla neutralizzazione del problema e al fine della tutela dei diritti dei cittadini-utenti.

Ad ogni buon conto, considerando quanto detto finora, in questo rapporto dicotomico è presente un "convitato di pietra", un'entità che aleggia e mantiene la prospettiva fluida, così da non permettere a nessuno degli attori in scena di trovare una soluzione apparentemente accettabile, ma soprattutto stabile e condivisa. Il riferimento non può che andare al progresso tecnologico e alla sua accelerazione senza precedenti.

In questa rincorsa ogni possibile schema di risoluzione della questione sembra venire a dissolversi, in quanto – citando il noto cantautore Lucio Dalla – per via della velocissima evoluzione degli ecosistemi digitali «quello che ieri era vero, non (sembra che) sarà vero domani». Questo fa sì che sia l'imprenditore privato che il legislatore finiscano per essere superati dagli effetti reali del progresso non appena provano a muoversi seppur solo progettualmente.

È in tale contesto in costante divenire che il presente contributo proverà a rendere una disamina di quanto sia stato sostenuto dai due schieramenti nella vicenda. Si analizzeranno i diversi approcci – libertari o meno – posti in essere da potere privato e potere pubblico, cercando di trarre una conclusione che sposti il *focus* dalla macchina (e dalla sua gestione) all'essere umano inteso come centro di imputazione di diritti, affinché l'Uomo e la sua intrinseca essenza di vita non siano sacrificati sull'altare di strumentali ambizioni prometeiche.

---

rafforzare il ruolo sociale del diritto, facendo sì che questo possa affrontare sia le sfide del presente sia quelle che il futuro gli porrà in seguito. Cfr. G. Zagrebelsky, *Il diritto mite*, Torino, 2024.

[4] Per una definizione tecnica delle piattaforme digitali v. A. Contaldo - F. Zambuco, *L'abuso di posizione dominante, piattaforme digitali e interventi statali: breve rassegna sugli interventi antitrust europei ed italiani nonché cenni sul Digital Service Act in USA,* in A. Contaldo (a cura di), *Le piattaforme digitali. Profili giuridici e tecnologici nel nuovo ecosistema*, Pisa, 2021, 1.

2. L'esperienza privata di moderazione del dibattito interno ai social network

Prima di addentrarsi nell'analisi del fenomeno va fatto un passo indietro, per capire a grandi linee quale sia la vicenda in cui ci si inoltra. Com'è noto, le piattaforme social si affermano come dei poli gravitazionali nell'universo digitale rappresentato dalla Rete. Da quando il Web è diventato mezzo di creazione di ricchezza e di potere, le istanze libertarie ed utopiche degli albori[5] hanno dovuto lasciare spazio a visioni più pragmatiche ed utilitaristiche. In questo senso, tuttavia, non si deve necessariamente intendere una mera economicità di fondo delle decisioni dei giganti social[6]; al contrario, deve considerarsi quanto detto in un'accezione di tendenziale equilibrio politico che faccia sì da non creare malcontento generale, in quanto quest'ultimo porterebbe a minori interazioni e, in seguito, a minori guadagni[7]. Non esiste, quindi, un metro comune di giudizio nelle azioni di queste entità, quanto semmai si può provare ad intravedere dei *patterns*.

La verità è che l'attività di moderazione non è esattamente un compito così semplice. Essa, infatti, rivela una natura intrinsecamente politica[8]: come riporta Gillespie[9], talvolta questo impegno, che in via iniziale spetta alle piattaforme ed ai loro controllori privati, può comportare delle scelte che non rispecchiano pienamente gli standard generali autoimposti o che a tutti gli effetti violano quella uguaglianza sancita dalle *policies* degli stessi social network[10]. Gli "incidenti", dunque, sono sempre dietro l'angolo. A questo riguardo, ancora Gillespie ricorda come queste stesse regole autoimposte non siano né di facile individuazione ma soprattutto di semplice attuazione in via unitaria in quello che si potrebbe definire "l'ordinamento giuridico social". A certificare tale situazione di fatto, si potrebbero segnalare in questa sede alcuni degli episodi che più hanno destato scalpore nei confronti della gestione dei contenuti all'interno delle piattaforme social: per ciò che concerne Facebook, si potrebbero citare i due casi di censura algoritmica relativi alle nudità presenti nella foto "*Napalm Girl*" di Nick Ut[11]

---

[5] Emblema di questo approccio è la *Dichiarazione dell'indipendenza del cyberspazio* scritta nel 1996 dal noto attivista John Perry Barlow.

[6] Per quanto a seguito della sospensione dell'account personale di Donald Trump e dei suoi profili collegati il titolo di Twitter in borsa abbia perso fino anche il 12% del valore di quotazione; P.R. La Monica, *Twitter's stock falls after Trump's account is suspended*, in *CNN Online*, 11 gennaio 2021.

[7] Quindi, come si può capire leggendo, il fattore economico rappresenta un tassello importante nell'equilibrio che si è venuto a creare con l'avvento dei social network. Tuttavia, non è allo stesso tempo ravvisabile una sua totale predominanza nel processo strategico di gestione di queste imprese digitali.

[8] A. Chander, *Who Runs the Internet?*, in *Research Handbook on the Politics of International Law*, Cheltenham, 2017, 418-42.

[9] T. Gillespie, *Custodians of the Internet Platforms, Content Moderation, and the Hidden Decisions That Shape social media*, Cambridge, 2018, 10 ss.

[10] A questo riguardo, Gillespie sottolinea come il criterio più semplice da individuare non sia tanto quello di un'uguaglianza per come la si riconosce nella maggior parte delle costituzioni contemporanee quanto semmai quello del c.d. "*right thing to do*" . T. Gillespie*, ivi,* 11.

[11] S. Levin, *Facebook backs down from 'Napalm girl' censorship and reinstates photo"*, in *The Guardian*, 9 settembre 2016.

o alla scultura della Sirenetta di Andersen posta all'entrata del porto di Copenaghen[12]. Tuttavia, maggiori problemi derivano dalla confusa gestione del dibattito su piattaforme quando si viene a contatto con la politica. Proprio su questo versante, si muove la problematica attinente all'altro social per eccellenza del Web e che fa da concorrente ai giganti di Meta, ossia X (ex Twitter). Negli anni, quest'ultimo si è attirato a sua volta un discreto numero di critiche, complice anche il fatto che esso mantenga da tempo una grande rilevanza come mezzo di comunicazione politica da parte dei principali leader nazionali ed internazionali. Una delle prime vicende ad aver aperto il vaso di Pandora è stata quella del parlamentare indiano Raja Singh[13], il quale affermava nel 2017 che i rifugiati di etnia Rohingya dovessero essere uccisi qualora si fossero rifiutati di tornare da dove fossero venuti. Solo a seguito di uno scandalo - il quale tuttavia era scoppiato con mesi di ritardo rispetto al tempo in cui erano stati caricati i contenuti discriminatori - si era giunti alla chiusura del profilo del politico sopra nominato.

Ma il caso di specie che ha portato la piattaforma alla ribalta è stato senza dubbio quello che ha riguardato l'ambiguo rapporto tra il social e il 45° presidente degli Stati Uniti Donald Trump. Durante i turbolenti anni del suo mandato, la dirompente comunicazione del presidente aveva da sempre sollevato critiche da parte di pubblico e di esperti. In ogni caso – nonostante le vicende alterne – Twitter aveva giustificato la permanenza dei vari contenuti caricati per via della discutibile politica del "*public interest framework*"[14], secondo la quale un determinato contenuto caricato sulla piattaforma sarebbe dovuto rimanere su di essa per via della possibile utilità per il pubblico (eludendo anche i basici controlli algoritmici a cui tutti gli altri contenuti erano sottoposti[15]). Ciò nonostante, dopo nemmeno due anni dalla formulazione di questa teoria, l'impresa di San Francisco è corsa ai ripari a seguito della drammatica vicenda dell'assalto a Capitol Hill del gennaio 2021, cancellando *ex abrupto* il profilo del presidente uscente[16].

Volendo essere completi nella trattazione, manca all'appello un lato dei *social* ancora poco esposto al pubblico ed alla vasta comunità di coloro che navigano su Internet: il riferimento va a quelle piattaforme di nicchia, come Reddit e 4chan e Truth, i quali a loro volta – forse più di tutti gli altri – hanno contribuito a creare un clima conflittuale grazie alla loro generica ambiguità in tema di moderazione dei contenuti su di essi

---

12 *BBC News*, *Denmark: Facebook blocks Little Mermaid over 'bare skin*, in BBC News, 4 gennaio 2016.

13 N.R.C. Assam, *BJP MLA Raja Singh says illegal immigrants refusing to go back should be shot*, in *Times of India*, 31 luglio 2018.

14 Twitter Inc., *Defining Public Interest on Twitter*, in *Twitter Blog*, 15 ottobre 2019; Twitter Inc., *World Leaders on Twitter: Principles and Approach*, in *Twitter Blog*, 15 ottobre 2019; Twitter Inc., *General Guidelines and Policies: About Public-interest Exceptions on Twitter*. Volendo puntualizzare, con quanto affermato non si intende concludere che la clausola in sé sia erronea a prescindere - si sa come sia stata molto spesso utilizzata in casi simili giurisprudenziali (uno per tutti il caso celebre *Google Spain)* - quanto semmai che attraverso l'uso di questa formula ci si sia potuto costruire sopra un abuso, complice soprattutto la rilevanza di chi esprimeva il proprio pensiero attraverso quell'account.

15 I criteri (cumulativi) da rispettare per poter godere di questo "regime speciale" erano i seguenti:
1) L'essere un ufficiale governativo o politico candidato/eletto alle cariche governative;
2) Avere una cifra di *followers* superiore o uguale a 100.000;
3) Avere un account verificato.

16 Twitter Inc., *Permanent Suspension of @realDonaldTrump*, in *Twitter Blog*, 8 gennaio 2021.

condivisi. Anche per loro si renderà necessario successivamente un focus specifico.
Al termine di questa analisi preliminare si giunge quindi ad una conclusione: per certi versi tutte le piattaforme social conservano allo stesso tempo similarità e differenze tra loro[17]. Come tutte le imprese private, queste hanno le proprie regole e ad esse sono gelosamente affezionate. Talvolta si traducono in politiche più o meno permissive e trasparenti nei confronti di coloro che vengono a contatto con esse; l'uso continuo di sistemi informatici automatici, complice la mole gargantuesca di dati da elaborare, negli anni ha prodotto vere e proprie distorsioni ed ha causato forti dissonanze tra ciò che dovrebbe essere permesso e ciò che, nella pratica, lo è davvero. In questo orizzonte desta ancor maggiore preoccupazione l'avvento dell'IA, che non fa che allontanare l'auspicabile ritorno ad un controllo prettamente di matrice umana.
Ciò premesso, dall'analisi dello schema comune, si rende utile un *excursus* singolare per capire le scelte dei padroni delle piattaforme e per trarre insegnamento da esse; per comprendere soprattutto se un'altra strada sia possibile e, qualora esista, se sia anche la più giusta da percorrere.

2.1 Il gruppo Meta e il sistema misto di controllo

Decidendo di partire dal gruppo più grande della "fauna social" – ossia Meta[18] –, bisogna notare come la questione della moderazione dei contenuti sia da sempre un tema molto caldo per la dirigenza della società.
Più nello specifico, si può evidenziare come una delle principali missioni della piattaforma sia la seguente: «dare agli utenti il potere di creare community e rendere il mondo più unito»[19]. Nel declinare questo assunto, il gigante di Menlo Park ha cercato negli anni di adottare soluzioni differenti per ovviare alle criticità nascenti dal compito di mantenimento dell'ordine pubblico-digitale. Tra di esse bisogna distinguere più livelli di controllo, in quanto essi si diversificano per mezzi, modalità e procedimento.
Partendo dal presupposto che ogni anno il volume di casi che pervengono all'attenzione del *team* di Meta preposto al controllo dei contenuti sia quasi incalcolabile (il *Transparency Report* 2023 conta decine di milioni di casi[20]), ben si può comprendere che non tutti gli incarichi di moderazione siano basati su una presenza umana. In questo contesto, Meta utilizza in prima istanza strumenti algoritmici per risolvere le questioni più elementari, come ad esempio per ovviare alla presenza di *posts* che siano manifestamente contrari ai principi sanciti dalla *policy* generale[21].

[17] Intendendo con ciò che ognuna ha le proprie linee guida che le differenzia, almeno in parte, dalle altre.

[18] Secondo il noto sito web Statista, nell'ultimo trimestre del 2023, le piattaforme facenti parte del gruppo (Facebook, Instagram, WhatsApp e Messenger) hanno totalizzato l'accesso mensile di ben 4 miliardi di utenti; per rendere la dimensione del tutto, una persona su due al mondo ogni mese usa strumenti Meta. Statista, *Meta Platforms - statistics & facts*.

[19] Sezione I - Condizioni d'uso: Servizi offerti da Facebook.

[20] Il *Meta community standards enforcement Report*.

[21] Si potrebbe discutere a lungo su come le linee guida predisposte dal gruppo Meta per i vari social satelliti possano risultare arbitrarie, ingiuste, etc. Sia la dottrina che la giurisprudenza (italiana ed estera),

Contro le decisioni automatizzate "di primo grado" solo una minima parte degli utenti chiede di promuovere una sorta di "appello" – rimanendo nel linguaggio giudiziario –, il quale, secondo a quanto tiene a specificare Meta, è gestito interamente da controllori umani. Citando pedissequamente quanto riportato dalla pagina web del Gruppo: «Se l'addetto al controllo accetta la decisione originaria, i contenuti non vengono ripristinati. Se invece l'addetto al controllo non è d'accordo con il controllo iniziale e decide che i contenuti non andavano rimossi, questi verranno sottoposti a un altro addetto al controllo, che deciderà se il contenuto deve essere ripristinato o meno[22]». In ultimissima istanza, frutto di anni di controversie e di elaborazione interna[23], è anche presente il c.d. "*Oversight Board*", il quale però è da considerarsi come un organo *sui generis* che può essere adito solo in circostanze speciali. Ne discende che questo organo di controllo non possa essere tenuto in considerazione se non per la sua mera esistenza[24].
In ogni caso, per riuscire a mettere in atto concretamente queste procedure, Meta afferma di poter contare su più di quaranta mila lavoratori[25], i quali sono distribuiti nei centri nevralgici del globo per diversi motivi, dalla migliore gestione del traffico informatico alla necessità di venire incontro alle sensibilità delle popolazioni e delle loro differenti culture[26]. Tuttavia, guardando agli orizzonti dello stesso Gruppo, l'IA

negli ultimi anni, si sono interrogate sul tema, giungendo a conclusioni che, in un certo senso, sono poi state trasfuse negli ultimi interventi europei per la regolamentazione delle piattaforme; tra questi, il più attinente è senz'altro il Digital Services Act. Cfr. Meta, *Normative*.

22 Meta, *Contenuti oggetto di ricorso*.

23 A seguito del caos mediatico creato dallo scandalo *Cambridge Analytica*, per primo Mark Zuckerberg, col supporto di studiosi del diritto, accademici a vario titolo e altri *stakeholders*, ha supportato la creazione di un nuovo organo per il controllo dei contenuti più sensibili e rilevanti all'interno dei propri *social*. È in questo orizzonte che si è venuto a creare l'Oversight Board; citando le parole di Mark Zuckerberg: «*You can imagine some sort of structure, almost like a Supreme Court, that is made up of independent folks who don't work for Facebook, who ultimately make the final judgement call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world*». A riguardo si veda K. Klonick - T. Kadri, *How to Make Facebook's 'Supreme Court' Work*, in *New York Times*, 17 novembre 2018.

24 La volontà sottesa al momento della creazione di tale Consiglio era quella di rendere quest'ultimo in qualche modo indipendente dalla dirigenza di Facebook (ancora non si chiamava Meta). Questa idea avrebbe rispecchiato l'intento di creare un meccanismo di separazione dei poteri simile a quello della teoria dello Stato di diritto (infatti ad esempio nella carta istitutiva si prevedeva anche l'obbligo di motivazione della decisione e particolari accorgimenti in tema di trasparenza). Tuttavia, guardando con occhi più attenti, si notano delle contraddizioni alla base del funzionamento di questo organo. Perciò, una parte dei commentatori ha parlato di questa mossa come di un programma di marketing definibile "*legal washing*"; v. M. Gaye-Palettes, *Between private and state justice: Facebook and the legal washing of its "supreme court"*, in *Pouvoirs*, 3, 2021, 119-129.

25 Merita segnalare le condizioni di questi lavoratori, i quali svolgono un lavoro che contempla la visione di contenuti deprecabili e detestabili, a cui non corrispondono giuste tutele. Durante il periodo della pandemia, per via dell'enorme ricorso alle risorse digitali, anche questa tematica è emersa nel dibattito europeo. Cfr. C. Criddle, *Facebook moderator: 'Every day was a nightmare'*, in *BBC Online*, 12 maggio 2021.

26 Per quanto affermi attualmente Meta, non è stato sempre così. Kate Klonick, nel 2017, attraverso le testimonianze di coloro che vi lavoravano, ha raccontato di come Facebook provvedesse in passato ad "allenare" i propri lavoratori ad eliminare i loro valori e *bias* culturali a favore delle *policies* dell'azienda. Il fine era quello di far sì che i *content moderators* rispecchiassero solo ed esclusivamente gli *standard* del social stesso. Chiaramente i problemi sono sorti quando i moderatori si sono trovati dinanzi a giudicare post che confliggevano con i valori più tipici della propria cultura (in particolare numerosi errori si sono potuti riscontrare in tema di contenuti afferenti a nudità/pornografia); K. Klonick, "*The New Governors: The People, Rules, and Processes Governing Online Speech*", in *Harvard Law Review*, 131, 2017, 1598-1670.

sembra rappresentare il prossimo *step* verso un progressivo miglioramento di questa attività interna di governo del dibattito. Non ne fa un mistero sul proprio sito neppure Meta stessa, la quale afferma di star lavorando su più fronti al fine di implementare gradualmente maggiori funzioni gestite da sistemi automatizzati[27]. Il dubbio è che questo progresso possa andare a colpire quel poco che rimane del controllo umano, relegando quest'ultimo all'estremo rimedio dell'*Oversight Board*, il quale però non è da tutti raggiungibile[28].
Attraverso il progresso tecnologico, dunque, la giustizia privata digitale mette da parte l'Uomo, rendendo il beneficio del rapporto con la valutazione umana un privilegio più che un diritto.

2.2 Il *"visibility filtering"* predisposto da X

Per quanto concerne l'altro protagonista nel *pantheon* digitale sono necessarie alcune precisazioni di carattere storico. Si badi bene, anche in questo caso l'esperienza nel campo imprenditoriale non va oltre i venti anni di vita della piattaforma, tuttavia è indubbio che, tra l'avvicendamento di Trump alla Casa Bianca e l'acquisto di Elon Musk, per X siano trascorsi - almeno in senso figurato - svariati secoli e non una manciata di anni. Questa affermazione sorge dalla statuizione di fatto per la quale siano evidenti le differenze che intercorrono tra il vecchio modello di gestione del social network e quello attuale. Di seguito, quindi, si delineano brevemente le divergenze di cui abbiamo fatto accenno prima.
Twitter (si usa il vecchio nome per caratterizzare la struttura precedente all'acquisto da parte di Musk) ha sempre manifestato un forte interesse verso la moderazione dei contenuti. La ragione di ciò si riscontrava nell'acquisita rilevanza politica del social. Come accennato, negli anni la piattaforma si era particolarmente prestata a rappresentare una bacheca informale dove i corpi politici potevano esternare i propri pensieri, pubblicizzare le proprie proposte e "catturare" il consenso popolare. Per non perdere il ruolo guadagnato, tuttavia, Twitter ha lasciato maglie piuttosto larghe riguardo la moderazione dei contenuti, soprattutto alle personalità di carattere pubblico istituzionale[29].
Con la presa di coscienza dettata dall'assalto a Capitol Hill e il vicino acquisto della società da parte dell'imprenditore Elon Musk, la neo-nominata X ha deciso di approntare una nuova strategia per la gestione del dibattito al suo interno. La vocazione libertaria manifestata dallo stesso CEO[30] si è venuta a scontrare con una realtà che si rivela molto più frastagliata, e per certi versi anche più subdola.
Il sistema creato dalla piattaforma, infatti, a differenza del procedimento tipico di Meta Group, non prevede in alcun modo che la libertà di pensiero sia negata, quanto

[27] Meta, *In che modo Meta investe nella tecnologia*, 19 gennaio 2022.

[28] Gaye-Palettes, *Between private and state justice: Facebook and the legal washing of its "supreme court", cit.*

[29] *Note supra* 15-16.

[30] D. Milmo, *Elon Musk defends stance on diversity and free speech during tense interview*, in *The Guardian*, 18 marzo 2024.

semmai silenziata[31]. A questo riguardo, secondo le linee guida di X[32], i contenuti ritenuti "inappropriati" dall'algoritmo verrebbero automaticamente de-indicizzati, facendo sì che questi appaiano a una parte ristretta del pubblico secondo criteri che non sono in alcun modo specificati. Il c.d. "*visibility filtering*", a cui ci si può contrapporre contattando il *team* di assistenza della piattaforma, è probabilmente lo strumento più pericoloso se si guarda alla materia di moderazione della libertà di manifestazione del pensiero digitale: esso, infatti, delega totalmente ad un sistema automatizzato la gestione dei diritti degli utenti, non evidenziando in alcun modo i principi alla base delle scelte attuate e – soprattutto – tenendo gli stessi *users* del social network all'oscuro del fatto che il loro diritto ad esprimersi e ad informarsi sia potenzialmente leso.
In questo quadro, la libertà di espressione non viene né superata né cancellata, quanto semmai essa è ridotta ad un'esistenza meramente formale, per la quale di fatto non sussiste e non rileva in alcun modo ad un'autentica applicazione del diritto fondamentale. La presunta garanzia della presenza umana in seconda istanza perde di qualsiasi significato, in quanto si rende difficile all'individuo di conoscere la necessità o no di appellarsi a rimedi per la tutela dei propri interessi.
Alla luce di questa breve descrizione si può intuire a cosa porti questo sistema: la conseguenza pratica più diretta si riscontra nella creazione di un meccanismo delatorio (basato sulle segnalazioni anonime tipiche del contesto digitale) a cui fa da collegamento un organo automatico dal funzionamento sconosciuto che può decidere se limitare la sfera dei diritti fondamentali della persona. Ben si può intuire come la strada per una semplificazione tecnologica che cerchi di curare le diverse opinioni e, allo stesso tempo, gli interessi aziendali non possa passare attraverso soluzioni di questo tenore senza ledere il rispetto dei diritti fondamentali della persona umana.

2.3 L'approccio autogestionale: Reddit, 4chan e Truth.

Nella disamina dei social network che si districano con le loro soluzioni nel complesso problema della moderazione dei contenuti vanno inserite anche tre piattaforme controverse[33]. Il riferimento ricade sul trio rappresentato da Reddit, 4chan e Truth. Soprattutto per questi *social media* si renderanno necessarie delle contestualizzazioni, utili a capire come la loro natura "alternativa" si sia venuta a creare.
Le prime due, nate proprio all'inizio del secolo, si caratterizzano per una forte carica libertaria rappresentata dal generale principio di autogestione. I loro creatori, soprattutto con riferimento a Reddit, hanno sempre rigettato l'idea di controlli superiori - in particolare con riferimento a quelli di tipo statale[34] - dando la possibilità di mantenere

[31] Che questa affermazione possa risultare contraddittoria è palese, ma la scelta di utilizzare questa espressione nasce dalla volontà di sottolineare l'incoerenza tra quanto viene detto in pubblico dalle figure di spicco dell'impresa e la realtà dei fatti.

[32] X Safety Team, *Freedom of Speech, Not Reach: An update on our enforcement philosophy*, in *X Blog,* 17 aprile 2023.

[33] S. J. Brison - K. Gelber, *Free Speech in the Digital Age*, Oxford, 2019, 162.

[34] P. Guest, *I'm Reddit's CEO and Think Regulating social media Is Tyranny,* in *Wired*, 17 aprile 2023.

l'ordine in primis agli utenti e ai frequentatori della "piazza digitale"[35].
Leggermente diverso è il caso di Truth, che comunque mantiene un approccio di tendenza libertaria[36]. Non avrebbe potuto essere altrimenti, viste le condizioni in cui si è deciso di fondarlo: dopo l'eliminazione coatta degli account di Donald Trump dagli altri social *mainstream* (Facebook, Instagram, Tik Tok e Twitter), lo stesso entourage dell'ex presidente si era determinato a creare una nuova piattaforma dove poter dare spazio al magnate. La regola generale, quindi, è sempre stata quella del *laissez faire,* con una generica clausola di esenzione della responsabilità della piattaforma nei confronti di quanto su di essa fosse stato riportato[37].
Fatte le dovute contestualizzazioni, in tutte e tre le casistiche si può notare come alla base del loro funzionamento pratico vi sia una chiara responsabilizzazione dell'utente, il quale si pone come principale - e unico - centro di controllo all'interno del confronto digitale.
Sebbene a prima vista questo spirito di libera collettività possa apparire come la realizzazione dell'utopia dei pionieri del Web, le criticità rivelano subito la loro inquietante presenza. L'esistenza di piattaforme gestite in senso anarchico-libertario dai soli utenti, senza che vi siano neppure dei correttivi regolamentari interni, dimostra come non sia opportuno lasciare l'attività di moderazione dei contenuti ai soli esseri umani, i quali, necessariamente, rischierebbero di trasporre le proprie convinzioni – valide o meno che siano – in un giudizio arbitrario e unilaterale sul diritto fondamentale alla libera espressione.
A quanto riportato va poi aggiunta un'ulteriore annotazione: complici i volumi di traffico[38] informatico dei social, non è materialmente possibile predisporre apparati completamente umani finalizzati alla supervisione di quanto accade nei vari account; ciò comporterebbe inevitabilmente delle falle nel sistema, il quale non potrebbe difendersi da eventuali comportamenti non adatti.
In conclusione, queste esperienze sono necessarie per comprendere come, in questa discussione, non sia possibile supportare scelte di tipo assoluto: il ritorno al controllo umano di stampo luddista ormai è soluzione superata dal tempo e dall'evoluzione tecnologica; l'unico modo per trovare una soluzione è abbracciare un compromesso equilibrato e sostenibile per la collettività e i singoli individui.

[35] Citando l'espressione coniata dalla Corte suprema degli Stati Uniti in *Packingham v. North Carolina*; US Supreme Court, No. 15–1194, 19 giugno 2017.

[36] Dopo la sua creazione, in molti si sono domandati cosa avesse in mente Donald Trump nel momento in cui ha deciso di fondare una nuova piattaforma digitale in risposta al suo *ban* dai social network *mainstream.* M. McCluskey, *What's Allowed on Trump's New 'TRUTH' Social Media Platform—And What Isn't,* in *Time*, 22 ottobre 2021.

[37] Truth, *Termini legali di servizio*.

[38] Non è possibile controllare singolarmente senza alcun aiuto informatico il traffico digitale di quello che è stato calcolato in 5 miliardi di utenti annuali attivi sulle piattaforme; Statista, *Number of social media users worldwide from 2017 to 2028*.

3. Le soluzioni pubbliche al problema

Come si è potuto comprendere dalla lettura dei paragrafi precedenti, il fenomeno digitale si rivela troppo diffuso e rilevante dal punto di vista sociale per essere lasciato in totale concessione a coloro che posseggono le piattaforme.
Alla luce di ciò, negli ultimi anni, i legislatori hanno provato a porre rimedi per contenere i danni emergenti da un settore che aveva iniziato a dimostrare di essere pericoloso anche dal punto di vista di tenuta democratica dei vari Paesi[39]. Nel prosieguo della trattazione si sceglie di distinguere due tipologie di approcci diversi: prima quello del singolo Stato e poi quello unitario, cercando di evidenziare come, a dispetto della volontà di proteggere i cittadini e l'economia, l'avanzamento tecnologico tenda a porre l'asticella del progresso giuridico sempre più in alto.

3.1 Gli interventi degli Stati sul tema: un passaggio sul Network Enforcement Act tedesco[40]

Guardando a ritroso nel tempo – ma senza andare troppo lontano, considerando la velocità di sviluppo delle tecnologie in questione – gli Stati hanno inizialmente percepito due principali criticità: l'hate speech e le fake news. La ragione di tale attenzione risiedeva in uno stato patologico del dibattito politico nei Paesi di impronta costituzionale e democratica. Durante il periodo 2016-2020, che va dalla campagna per la Brexit fino alle presidenziali statunitensi del 2020, si sono susseguite diverse risposte pubbliche volte a esercitare pressione sui colossi del Tech. Tra gli interventi normativi più rilevanti si possono citare quelli di Turchia, Russia e Regno Unito[41]. Tuttavia, questi esempi non possono essere considerati modelli significativi, sia per la natura non democratica dei primi due Stati, sia per le differenze giuridiche rispetto all'esperienza europea.
Concentrandosi su un contesto più vicino alla tradizione euro-unitaria, il Network Enforcement Act tedesco (NetzDG)[42] emerge come un esempio chiave, pur con i suoi

[39] Tra i lavori che hanno individuato meglio la questione si segnala: S. C. Woolley - P. N. Howard, *Computational Propaganda : Political Parties, Politicians, and Political Manipulation on Social Media*, New York, 2019.

[40] Con questo paragrafo si renderà necessaria una digressione che parte da presupposti diversi rispetto al tema centrale della trattazione. Difatti, non sarà immediatamente oggetto della discussione la questione della presenza umana nell'atto di moderazione e gestione dei contenuti sui social network, quanto semmai quest'ultima in sé e per sé. Il motivo per cui si preferisce questo approccio sta nella motivazione per la quale, a parere di chi scrive, non è possibile comprendere l'insorgenza delle varie necessità da parte delle legislazioni se non se ne segue il progresso nel tempo. Solo dopo aver affrontato per la prima volta la questione della moderazione ci si è chiesti quale potesse essere il ruolo dell'essere umano nel quadro generale. Per questo motivo, si ritiene opportuno iniziare la disamina a partire dai primi interventi che hanno riguardato le limitazioni nazionali alle varie piattaforme digitali.

[41] Su queste tre esperienze legislative si segnala l'analisi comparata contenuta in T. Kasakowskij - J. Fürst - J. Fischer - K. J. Fietkiewicz, *Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media,* in *Telematics and Informatics*, 46, 2020.

[42] Per comprendere come fu accolto e analizzato al tempo questo intervento normativo si rimanda, anche per vicinanza, a V. Claussen, *Fighting Hate Speech and Fake News. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation*, in questa *Rivista*, 3, 2018, 110-136.

limiti, per il successivo sviluppo del Digital Services Act (DSA) e, di conseguenza, dell'AI Act.
Ai fini di questa trattazione, la parte della normativa tedesca che deve interessarci maggiormente risiede nell'art. 1, sezioni 2-3, ossia la previsione per la quale si subordina il rispetto dell'ordine pubblico digitale all'azione esecutiva dei gestori delle piattaforme. In questo primo frangente, non si è pensato a porre limiti alle modalità di gestione utilizzate dai privati, quanto semmai di rafforzare la responsabilità in capo ad essi di mantenere un ambiente digitale rispettoso dei principi legislativi tedeschi. Per mantenere alta la pressione in capo ai social si sono associati termini temporali e multe particolarmente incentivanti[43].
Già dal momento dell'implementazione della normativa nell'ordinamento tedesco è stato sottolineato da una parte della dottrina che il modello non potesse essere sostenibile. In particolare, le criticità avrebbero toccato due questioni di particolare rilevanza: il potere discrezionale delle piattaforme su come far rispettare la legge e l'insidiosa pratica del c.d. "*over-blocking"* di matrice algoritmica[44].
È stato notato come i privati gestori dei social potessero decidere i mezzi da utilizzare per garantire l'applicazione della legge con piena discrezionalità, venendo meno a qualsiasi principio di garanzia nei confronti della libertà di pensiero degli utenti. A ciò si aggiungeva la tendenza per la quale, di norma, era più conveniente bloccare i profili degli utenti piuttosto che controllare compiutamente che fosse avvenuto un vero e proprio illecito sulle piattaforme.
Dunque, ben si capisce come l'esperienza accennata si sia rivelata un passaggio obbligato da parte della legislazione tedesca, la quale, attraverso l'esperimento di questo tentativo, ha permesso all'Unione Europea di individuare un'esigenza e di trasformarla in una risposta normativa che non ricadesse negli errori fatti dal tentativo prototipale.

3.2 La prospettiva europea per una costruzione di un ecosistema digitale a misura dell'Essere umano: DSA e AI Act

Come accennato, alla base del DSA e dell'AI Act (a cui si potrebbe aggiungere anche il Digital Markets Act, che tuttavia si sostanzia in previsioni di natura prevalentemente

[43] Network Enforcement Act, art. 1, sezione 3 e 4.8 (2).

[44] Più nello specifico, per quanto riguarda la prima annotazione, si è sottolineato come la concessione alle piattaforme di totale discrezionalità nella scelta dei mezzi per tutelare l'ordine avrebbe fatto sì da favorire scelte poco ponderate e di natura automatizzata rispetto a tutte le casistiche concernenti la manifestazione del pensiero. In questa direzione, per motivi simili, va anche la discussione relativa all'*over-blocking*. Le piattaforme, incalzate da termini brevi e multe gravose - nascondendosi dietro il legittimo uso di algoritmi - non avrebbero avuto problemi a rimuovere a prescindere tutti i contenuti "rischiosi", ledendo la libertà di espressione anche di coloro che, in realtà, non avrebbero commesso alcun illecito. Cfr. A. Bormann, *Dealing with Digital Social Networks: The German "Network Durchsetzungsgesetz" (Network Enforcement Act)-A Challenging Balance between Combating Hate Crimes and Protecting the Freedom of Expression*, in *Bulletin of the Transilvania University of Braşov. Series VII, Social Science: Law,* 11(2)-Suppl, 2018, 25-30; S. Schmitz - C. Berndt, *The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword?*, in *ssrn.com,* 9 gennaio 2019; S.Theil, *The German NetzDG: A Risk Worth Taking?*, in *Verfassungsblog.*

economica, pur rientrando nei solchi della stessa politica regolatoria) vi è stata, innegabilmente, una presa di coscienza da parte del legislatore europeo, derivante anche dall'esperienza del tentativo tedesco. Tuttavia, è importante precisare che il DSA e l'AI Act non siano sorti dal nulla in seguito a tale esempio statale; piuttosto, rappresentano i prodotti più noti di un'elaborazione che ha coinvolto per anni gli Stati membri dell'UE.

Una propensione progettuale[45] che ha portato prima al *Code of Practice on Disinformation*[46] e poi, contestualmente ai già menzionati interventi, all'emanazione, del Regolamento 2024/900 (relativo alla trasparenza e al targeting della pubblicità politica**)** e del Media Freedom Act. A questo proposito, in relazione al nostro tema, risulta particolarmente interessante l'approccio adottato da quest'ultimo in materia di gestione algoritmica. Il regolamento, infatti, prevede misure volte a migliorare la trasparenza degli algoritmi utilizzati dalle piattaforme, specialmente quelli che influenzano la visibilità e la monetizzazione dei contenuti giornalistici, con l'obiettivo di garantire che i media possano operare in condizioni più eque rispetto alle piattaforme, prevenendo pratiche sleali automatizzate, come la manipolazione del ranking dei contenuti o politiche discriminatorie sui ricavi pubblicitari.

Ma tornando a noi, nella trattazione che segue si proverà a dare uno sguardo d'insieme ai due atti legislativi europei sopra citati, evidenziando le strategie messe in atto per proteggere la centralità dell'uomo (considerando con ciò sia i cittadini utenti che gli umani controllori) rappresentata dal principio del c.d. "*Human in the loop*"[47] .

3.2.1 Digital Services Act

Guardando alla forma che ha assunto il DSA, si può notare come - nella pratica[48] - esso sia stato diviso in quattro capi, i quali rispettivamente riguardano:

Disposizioni generali

Responsabilità dei prestatori di servizi intermediari

Obblighi in materia di dovere di diligenza per un ambiente online trasparente e sicuro

Attuazione, cooperazione, sanzioni ed esecuzione.

---

[45] Il riferimento va all'Agenda digitale 2030 stilato dalla stessa Unione europea (ultima consultazione 23 dicembre 2024).

[46] Progetto avviato nel 2018, il Codice sulle pratiche contro la disinformazione è stato aggiornato nel 2022 per rafforzare gli impegni e introdurre un meccanismo di monitoraggio, incluso un Centro di Trasparenza per garantire la rendicontazione pubblica. Sebbene volontario, è strettamente collegato al Digital Services Act (DSA), che prevede obblighi legali per le piattaforme più grandi e rende alcune misure del Codice più stringenti. Cfr. The 2022 Code of Practice on Disinformation.

[47] Per comprendere cosa s'intenda con questa espressione, si rende necessario far riferimento ad alcuni lavori di recente pubblicazione: *ex multis*, di taglio sia scientifico che divulgativo, si segnalano P. Benanti, *Human in the loop. Decisioni umane e intelligenze artificiali*, Milano, 2022; I. Drori, *Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risks*, in *Journal of the Association for Information Systems,* 25(1), 2024, 98–111; D. Sele - M. Chugunova, *Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision making*, in *PLoS ONE,* 19(2), 2024.

[48] Ad essere precisi i capi sarebbero cinque. Tuttavia, il quinto riguarda le disposizioni finali e altre modifiche attuative minori di normative già esistenti, rendendo i primi quattro le parti realmente innovative del Regolamento.

Nel primo breve capo si trovano quelle che si potrebbero definire in maniera informale come le "regole del gioco": in questi primi due articoli si è deciso di fare chiarezza sull'ambito di applicazione (art. 2) e sulle definizioni (art. 3) dei soggetti destinatari delle successive norme.

Tra di esse si può scorgere un parallelismo con il Netz Dg tedesco nella parte in cui si è scelto di procedere con la fissazione di ciò che è considerato "contenuto illegale[49]". Continuando a scorrere la normativa, viene in risalto la nuova distinzione tra servizi di *hosting*[50], *catching*[51] e *mere conduit*[52]; nella pratica, differenti riscontri fenomenici attribuibili alla figura dell'intermediario digitale a cui conseguono diversi regimi di responsabilità.

Proprio i successivi tre articoli (artt. 4-6 DSA), i quali aprono il secondo capo, si occupano di distinguere queste tre diverse tipologie di responsabilità, le quali si muovono su un binario uniforme, ossia quello della presunzione di irresponsabilità[53] salvo che concorrano condizioni di fatto[54] che cambiano a seconda della delicatezza del servizio posto in essere dall'intermediario digitale (ovviamente l'*hosting* prevede maggiori attenzioni, complice la memorizzazione dei dati e delle informazioni a tempo indeterminato). Per tutti e tre i provvedimenti viene mantenuta una clausola generale di riserva di giurisdizione, che potrebbe generare problematiche future a causa delle differenze nell'organizzazione giuridica dei vari Stati membri. In particolare, si prevede in modo esplicito che «(I) presenti articoli lasciano impregiudicata la possibilità, secondo gli ordinamenti giuridici degli Stati membri, che un organo giurisdizionale o un'autorità amministrativa esiga al prestatore del servizio di impedire o porre fine ad una violazione»[55]. Resta da vedere se questa disposizione, in futuro, potrebbe essere utilizzata

[49] A proposito, è tale «qualsiasi informazione che, di per sé o in relazione a un'attività, tra cui la vendita di prodotti o la prestazione di servizi, non è conforme al diritto dell'Unione o di qualunque Stato membro conforme con il diritto dell'Unione, indipendentemente dalla natura o dall'oggetto specifico di tale diritto». Digital Services Act, art. 3 lett. h).

[50] Per *hosting* si intende l'attività di memorizzazione di informazioni riferibili agli utenti di un servizio digitale. V. Digital Services Act, art. 6.

[51] Per *catching* si intende la temporanea memorizzazione dei dati presso i server utilizzati dalla piattaforma coinvolta. V. Digital Services Act, art. 5.

[52] Per *mere conduit* si intende l'attività di mero trasporto dei dati attraverso i server di una piattaforma. V. Digital Services Act, art. 4.

[53] Che riprende, anche qui, il principio del "Buon Samaritano" di legislazione statunitense che era già stato introdotto nell'ordinamento europeo con l'art. 14 della direttiva 2000/31/CE. A proposito della sezione 4 di quest'ultima, non si può non sottolineare come questo approccio derivi dall'esperienza statunitense rappresentata dalla Section 230 del *Communications Decency Act,* la quale sancisce a sua volta un'esclusione di responsabilità in capo alle piattaforme elettroniche. Volendo immergerci nella materia, bisogna ricordare come negli anni questo intervento legislativo abbia richiamato su di sé critiche di ogni genere, in particolare per la fragilità insita dello stesso principio del "Buon Samaritano" e della presunta buona fede che farebbe capo alle piattaforme stesse; v. A.M. Sevanian, *Section 230 of the Communications Decency Act: A "Good Samaritan" Law Without the Requirement of Acting as a "Good Samaritan".* in *UCLA Entertainment Law Review,* 21(1), 2014, 121-146, ma anche le considerazioni contenute in M.G. Leary, *The Indecency and Injustice of Section 230 of the Communications Decency Act,* in *Harvard Journal of Law and Public Policy,* 41(2), 2018, 621.

[54] In questo caso si intende il riconoscimento della condizione del prestatore di servizio digitale come *Host, Catcher o Mere conduit* dei dati dell'utente. Cfr. *note supra 51-52-53*.

[55] Questa clausola è ripetuta in maniera identica agli artt. 4, par. 3, 5, par. 2, e 6, par. 4, del Digital

dai Paesi membri come un margine di manovra per introdurre ulteriori limitazioni ai singoli social network.

Tuttavia, l'accelerazione rispetto alla legiferazione tedesca a cui si è fatto riferimento in precedenza è insita negli articoli 8 e 9 del DSA, dove rispettivamente si regolano l'"Assenza di obblighi generali di sorveglianza o di accertamento attivo dei fatti" e gli "Ordini di contrastare i contenuti illegali".

Per quanto riguarda il primo degli articoli nominati, la norma statuisce che «Ai prestatori di servizi intermediari non è imposto alcun obbligo generale di sorveglianza sulle informazioni che tali prestatori trasmettono o memorizzano, né di accertare attivamente fatti o circostanze che indichino la presenza di attività illegali»[56]. In poche parole, si tratta della trasposizione di una chiara strategia di *nudging* affinché le piattaforme siano dissuase dal creare sistemi privati di sorveglianza che consentano loro di farsi una propria "giustizia privata". Tuttavia, è con il primo comma dell'articolo successivo che avviene il cambio di passo rispetto al Network Enforcement Act tedesco: in esso si afferma «Appena ricevuto l'ordine di contrastare uno o più specifici contenuti illegali, emesso dalle autorità giudiziarie o amministrative nazionali competenti, sulla base del diritto dell'Unione o del diritto nazionale applicabili in conformità con il diritto dell'Unione, i prestatori di servizi intermediari informano senza indebito ritardo l'autorità che ha emesso l'ordine, o qualsiasi altra autorità specificata nell'ordine, del seguito dato all'ordine, specificando se e quando è stato dato seguito all'ordine»[57].

Con questa norma si viene a creare una vera e propria riserva di giurisdizione, da collegarsi con quanto previsto dall'art. 4 di cui sopra - ovviamente statale; quindi, si è ben lontani da realtà come il *Facebook Oversight Board* - per la quale i prestatori di servizi digitali sono necessariamente tenuti a recepire l'ordine senza alcuna possibilità di discrezione su quanto deciso in seno alle autorità pubbliche[58]. Quindi, le piattaforme restano in minima parte "braccio armato" del potere pubblico, con la decisiva cessione di quel ruolo decisorio che richiamava il *Network Enforcement Act*.

Quanto al capo terzo, tra gli articoli più attinenti allo studio finora portato avanti c'è il ventesimo, il quale regola il "Sistema interno di gestione dei reclami".

Viste le esperienze negative legate alle decisioni arbitrarie dei gestori dei social network, l'art. 20 statuisce l'obbligo in capo ad essi di predisporre «[…] l'accesso a un sistema

Services Act.

[56] Digital Services Act, art. 8, par. 1.

[57] Digital Services Act, art. 9, par. 1.

[58] A riguardo si veda anche il paragrafo successivo della norma precedentemente richiamata (art. 9 DSA), nella quale si stabilisce che: «Gli Stati membri provvedono affinché l'ordine di cui al paragrafo 1 trasmesso al prestatore soddisfi almeno le condizioni seguenti: a) l'ordine contiene gli elementi seguenti: i) un riferimento alla base giuridica dell'ordine a norma del diritto dell'Unione o nazionale; ii) la motivazione per cui le informazioni costituiscono contenuti illegali, mediante un riferimento a una o più disposizioni specifiche del diritto dell'Unione o del diritto nazionale conforme al diritto dell'Unione; iii) informazioni per identificare l'autorità emittente; iv) informazioni chiare che consentano al prestatore di servizi intermediari di individuare e localizzare i contenuti illegali in questione, quali uno o più URL esatti e, se necessario, informazioni supplementari; v) informazioni sui meccanismi di ricorso a disposizione del prestatore di servizi intermediari e del destinatario del servizio che ha fornito i contenuti; vi) se del caso, informazioni in merito a quale autorità debba ricevere le informazioni relative al seguito dato agli ordini(...)».

interno di gestione dei reclami efficace, che consenta (ai destinatari del servizio) di presentare per via elettronica e gratuitamente reclami contro la decisione presa dal fornitore della piattaforma online all'atto del ricevimento di una segnalazione o contro le seguenti decisioni adottate dal fornitore della piattaforma online a motivo del fatto che le informazioni fornite dai destinatari costituiscono contenuti illegali o sono incompatibili con le condizioni generali […]:
Le decisioni di rimuovere le informazioni o disabilitare l'accesso alle stesse;
Le decisioni di sospendere o cessare in tutto o in parte la prestazione del servizio ai destinatari;
Le decisioni di sospendere o cessare l'account dei destinatari […]»[59].
Con questa chiosa legislativa, si può dire che il legislatore europeo abbia voluto mettere in catene il sistema di moderazione *sui iuris* che ha caratterizzato i precedenti anni di vita dei social media. Inoltre - per non ricadere in episodi grotteschi, come quello della Sirenetta di Andersen nel porto di Copenaghen - il par. 6 ha stabilito a chiare lettere come «I fornitori di piattaforme online provvedono affinché le decisioni di cui al paragrafo 5[60] siano prese con la supervisione di personale adeguatamente qualificato e non avvalendosi esclusivamente di strumenti automatizzati»[61].
Questa disposizione si pone come naturale collegamento con quanto sarà tradotto nel corpus normativo dell'AI Act: non vi si trova solamente un rigetto al ricorso alla "giurisdizione esclusiva" privata ma anche - e soprattutto - alla giustizia algoritmica, assoluto simbolo di spersonalizzazione di un ambito, come quello dei diritti fondamentali ed in particolare della libertà *ex* art. 21 Cost. (almeno guardando al nostro ordinamento), dove maggiormente dovrebbe essere considerata la persona umana.
In questo articolo viene trasposta chiaramente la necessità di realizzare un progresso tecnico e giuridico improntato al valore dell'Essere umano. Questo concetto, se ci si sofferma a ragionarci sopra, rappresenta un'accelerazione vertiginosa rispetto a quanto statuito negli anni precedenti: per la prima volta, infatti, si mette da parte l'obiettivo di protezione a qualsiasi costo dell'ordine pubblico digitale, riportando la persona al centro del *focus* del legislatore. Tale orientamento, inizialmente solo accennato, è stato poi ulteriormente ripreso e rafforzato dal testo dell'AI Act.

3.2.2 AI Act

Come si è detto in precedenza, l'*AI Act* è visibilmente frutto di un processo di elaborazione che ha portato il legislatore europeo ad intervenire su più materie in maniera "intersezionale"[62], cercando di regolare a più riprese l'universo delle tecnologie digitali

[59] Digital Services Act, art. 20, par. 1.

[60] Digital Services Act, art. 20, par. 5: «I fornitori di piattaforme online comunicano senza indebito ritardo ai reclamanti la loro decisione motivata relativa alle informazioni cui si riferisce il reclamo e la possibilità di risoluzione extragiudiziale delle controversie di cui all'articolo 21 e le altre possibilità di ricorso a loro disposizione».

[61] Digital Services Act, art. 20, par. 6.

[62] In questo caso il termine "intersezionale" opera al di fuori del suo campo semantico originario (ossia quello legato alle lotte per la giustizia civile e sociale a tutto tondo). Si decide di utilizzare comunque

con una visione complessiva del fenomeno.

Ad ogni buon conto, è bene fare (per l'ennesima volta) un passo indietro al fine di una migliore comprensione della storia del Regolamento qui analizzato. L'*AI Act* fa parte della Strategia digitale dell'Unione Europea[63] ed è stato originariamente proposto dalla Commissione nell'aprile 2021. Per più di due anni il Regolamento è stato in gestazione all'interno delle Istituzioni europee, per giungere solo di recente a vera e propria vita.

Sin da quella che si potrebbe denominare "fase di *pre-drafting*", la priorità dell'Unione è stata quella di «garantire che i sistemi di IA utilizzati nell'UE (fossero) sicuri, trasparenti, tracciabili, (che impedissero) pregiudizi e discriminazioni, […] e (assicurassero) il rispetto dei diritti fondamentali»[64]. In particolare, il principio personalistico è stato da subito posto in evidenza come faro su cui creare sistemi di IA non pericolosi[65] per la società e l'individuo.

Su questa onda lunga, il Parlamento UE ha mostrato sempre più vivo interesse nel perseguimento del nuovo principio del *"Legal protection by design*"[66], ossia l'idea di una costruzione dell'infrastruttura informatica all'insegna del controllo giuridico. Per questo motivo, nel Regolamento sull'Intelligenza artificiale, si è cercato di stabilire definizioni uniformi e tecnologicamente neutre che possano, in un futuro, essere applicate a tutti i sistemi di IA. A tal fine, l'Intelligenza artificiale è stata definita come «un software […] in grado, per un determinato insieme di obiettivi definiti dall'uomo, di generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono»[67].

Ciò premesso, è bene immergerci nello studio dell'atto giunto all'approvazione del Consiglio il 21 maggio 2024[68]. Guardando alla versione finale dell'AI Act vengono in evidenza alcune somiglianze con il DSA: oltre ad una sistematica simile, anche in questo caso l'UE ha preferito introdurre la normativa attraverso la definizione della materia oggetto di legiferazione. Non a caso, dunque, sin dall'art. 3 ci si trova dinanzi ad una lunghissima lista - i punti sono più di 40 - di definizioni, a partire da cosa sia un "sistema di IA"[69] e chi sia il suo "utente"[70].

---

questo lemma poiché, come per le battaglie per i diritti civili, anche in questo caso la lotta contro la degenerazione del fenomeno digitale è stata affrontata in un'ottica d'insieme, in quanto i singoli ambiti, pur differenti, si collegano tra di loro nel fine ultimo, ossia la protezione dell'Essere umano, della sua dignità e dei suoi diritti fondamentali.

63 Commissione Europea, *Un'Europa pronta per l'era digitale. Più opportunità grazie a una nuova generazione di tecnologie*.

64 European Parliament News, *AI Rules: What the European Parliament Wants*, 20 giugno 2023.

65 *Ibid.*

66 Tra i primi contributi che hanno iniziato a vagliare questa nuova filosofia si segnala M.Hildebrandt, *Saved by Design? The Case of Legal Protection by Design*, in *Nanoethics,*11(3), 2017, 307–311.

67 S.Lynch, *Analysing the European Union AI Act: What Works, What Needs Improvement*, in *Human-Centred Artificial Intelligence (HAI) Stanford University*,, 21 luglio 2023.

68 Consiglio, *Artificial intelligence (AI) Act: Council gives final green light to the first worldwide rules on AI*.

69 AI Act, art. 3, par. 1, n. 1): «Un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono».

70 AI Act, art. 3, par. 1, n. 4: «Qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro

Ai fini della nostra trattazione è di estremo interesse la lettura degli articoli successivi all'art. 3. L'UE ha pensato di attivarsi sulla materia approcciandosi a questa diversificando le tipologie di Intelligenza artificiale; in particolare si sono venuti a distinguere quattro livelli di rischio:

1) Altissimo rischio
2) Alto rischio
3) Rischio limitato
4) Minimo rischio.

Nel titolo II della normativa in discussione si possono trovare le norme relative alle diverse categorie sopraddette, con tutte le limitazioni e i correttivi congegnati per porre un freno alla libera azione dei sistemi automatizzati.

Analizzando il contenuto dei precetti giuridici contenuti dall'art. 5 (Pratiche di Intelligenza artificiale vietate) all'art. 52 (Obblighi di trasparenza per determinati sistemi di IA), al fine della riflessione sul diritto alla manifestazione del pensiero e al conseguente diritto alla corretta informazione, si può notare come non sia così semplice individuare la fattispecie che funge da guida per la risoluzione delle criticità di cui si è trattato finora. Nello specifico, al già citato art. 5, par. 1, lett. a), si fa riferimento a «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico»[71]. Ugualmente, volgendo l'attenzione all'art. 7, par. 1, lett. b), si considerano "sistemi ad alto rischio" «i sistemi di IA che presentano un rischio di danno per la salute e la sicurezza, o un rischio di impatto negativo sui diritti fondamentali»[72].

Ben si capisce come la scelta di abbracciare una tesi ricostruttiva rispetto all'altra, in questo caso, apra scenari totalmente diversi tra di loro. Da una parte ci si verrebbe a trovare dinanzi ad una realtà nella quale l'uso dell'IA sarebbe totalmente vietato per via del pericolo sotteso di manipolazione dell'essere umano attraverso la concessione del potere di controllo al sistema informatico. Dall'altra, le maglie, seppur strette, lascerebbero spazio al progresso tecnologico di venirsi a sviluppare in una maniera sostenibile.

Tuttavia, la previsione che più raccoglie le speranze di salvaguardia del principio "*Human in the loop*" è l'art. 14 (Sorveglianza umana). Ai sensi dell'articolo citato, l'IA ad alto rischio deve poter essere sottoposta a controllo umano in tutti i passaggi legati al suo funzionamento, fino anche al momento precedente alla sua immissione nel mercato o alla sua messa in servizio. Questa sarebbe la soluzione per «prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali»[73].

Delineato questo quadro, va però sottolineato come ancora non ci sia alcuna certezza su quale delle due fattispecie riguardi l'ambito della moderazione dei contenuti condivisi dagli utenti. Dunque, dinanzi a questa iniziale confusione, non ci si può che rimettere

organismo che utilizza un sistema di IA sotto la sua autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

[71] AI Act, art. 5, par. 1, lett. a).

[72] AI Act, art. 7, par. 1, lett. b).

[73] AI Act, art. 14, par. 2.

all'interpretazione della normativa (in questo primo frangente sarà di sicura rilevanza la lettura della nota introduttiva della proposta di Regolamento e la sua spiegazione, i quali fanno fede per un'eventuale ricostruzione della volontà del legislatore)[74].

3.2.3 Riscontri positivi e negativi di DSA e AI Act

Al termine dell'analisi delle norme, emergono alcune considerazioni fondamentali, in linea con le diverse opinioni sollevate dall'introduzione di questi due storici interventi legislativi.

Il Digital Services Act rappresenta una pietra miliare nella regolamentazione dell'ecosistema digitale, offrendo soluzioni innovative alle sfide sempre più urgenti dell'era digitale, che negli ultimi anni avevano evidenziato la necessità di un intervento giuridico organico. Proprio questo carattere complessivo rappresenta per molti uno dei punti chiave del DSA. Gestire un fenomeno transnazionale richiede un lavoro coordinato ed equilibrato, capace di rispecchiare gli interessi di tutti gli Stati europei. Inoltre, gli obblighi di trasparenza e responsabilizzazione delle piattaforme assumono un forte valore simbolico, riaffermando il ruolo centrale del cittadino e, più in generale, della società nella transizione digitale[75].

D'altra parte, vengono sottolineati anche dei possibili punti deboli della normativa, soprattutto con riguardo al ruolo delle Big Tech. Nonostante gli auspici, si teme che i gestori dei social network possano reagire al nuovo regolamento adottando misure eccessivamente severe. Per evitare sanzioni, potrebbero continuare a fare affidamento esclusivo sugli algoritmi, rimuovendo anche contenuti legittimi nel dubbio che possano risultare illegali[76]. Questo fenomeno, al contrario delle aspettative, potrebbe portare a una riduzione dello spazio per il dibattito pubblico online. Inoltre, il DSA delegando ampie responsabilità alle piattaforme per identificare e rimuovere contenuti dannosi o illegali, potrebbe acuire lo stato attuale delle cose, anteponendo le loro necessità a quelle del confronto democratico e degli utenti.

Passando all'analisi dell'AI Act, emerge chiaramente la volontà dell'Unione Europea di guidare il cambiamento, contribuendo alla costruzione delle infrastrutture digitali del futuro del continente[77]. Nello specifico, riprendendo la disamina delle norme affrontate in precedenza, convince – come per il DSA – la politica definitoria attuata dal Legislatore; complice il carattere liquido della realtà digitale[78] e il continuo divenire tecnologico si cerca di stabilire dei confini chiari nei confronti degli utenti e degli sviluppatori, precisando i limiti della materia. In egual maniera è stata accolta la scelta relativa alle

[74] Premessa al Regolamento europeo sull'Intelligenza artificiale.

[75] A. Turillazzi - M. Taddeo - L. Floridi - F. Casolari, *The digital services act: an analysis of its ethical, legal, and social implications*, in *Law, Innovation and Technology*, 15(1), 2023, 83–106.

[76] M. Rojszczak**,** *The Digital Services Act and the Problem of Preventive Blocking of (Clearly) Illegal Content*, in *Institutiones Administrationis*, 3(2), 2023, 44-59.

[77] Si segnala un interessante disamina ad ampio spettro sul tema: M.Woersdoerfer, *The E.U.'s Artificial Intelligence Act: An Ordoliberal Assessment*, in *AI Ethics*, 2023.

[78] Facendo eco alla terminologia usata dal celebre filosofo Bauman in Z. Bauman, *Modernità liquida*, Roma, 2011.

fasce di rischio: infatti, uno degli elementi più innovativi dell'AI Act è proprio il suo *risk-based approach*, che suddivide i sistemi di IA nelle già menzionate quattro categorie. Questo modello consente una regolamentazione proporzionata, evitando eccessivi vincoli per applicazioni innocue e concentrando le restrizioni solo su quelle potenzialmente dannose. Si ritiene che tale approccio possa favorire un equilibrio tra innovazione e tutela dei cittadini. Da un lato, evita un quadro normativo eccessivamente rigido che potrebbe soffocare lo sviluppo di tecnologie emergenti; dall'altro, garantisce una maggiore attenzione verso le applicazioni ad alto rischio, come quelle utilizzate per il riconoscimento facciale, la valutazione del credito o il reclutamento del personale[79].

Quanto ai rilievi relativi alla sorveglianza umana, invece, vengono in evidenza due questioni: la trasparenza e la centralità della supervisione.

La trasparenza, inevitabilmente, aumenta la responsabilità (*accountability*) degli operatori, riducendo il rischio di decisioni arbitrarie o incomprensibili. Questo aspetto, come visto in precedenza per il DSA, è cruciale per garantire una maggiore accettazione sociale dell'IA e per mitigare il fenomeno del "*black box*"[80]. Quanto al secondo punto, un simile sistema di governance garantisce una gestione adeguata, riducendo i rischi di utilizzi impropri dell'Intelligenza Artificiale e offrendo un meccanismo di controllo flessibile ma rigoroso. Inoltre, con la creazione del Comitato europeo per l'IA si auspica che, nell'Unione, possa sorgere un forum per lo scambio di buone pratiche e per il coordinamento tra gli Stati membri, favorendo un'applicazione uniforme delle norme[81].

D'altra parte, anche in questo caso, non sono mancate le opinioni dissenzienti, le quali non guardano con particolare ottimismo alla normativa appena promulgata. Una delle principali critiche è che l'intervento legislativo rifletterebbe alcune carenze rispetto all'ambito scientifico. Si sostiene che la Commissione e il Parlamento non abbiano adeguatamente tenuto conto delle richieste degli sviluppatori, che, in virtù della loro conoscenza degli strumenti, si considerano i principali attori nella gestione del fenomeno. Le categorizzazioni, nello specifico, rifletterebbero una concezione legalistica ed eccessivamente solida, non adatta a governare il progresso[82]. Sull'attività di controllo, invece, un'eventuale carenza di risorse dedicate all'IA potrebbe portare a un'applicazione disomogenea della normativa tra gli Stati membri, compromettendo l'armonizzazione normativa che l'AI Act intende promuovere. Le autorità nazionali potrebbero non essere sufficientemente attrezzate per monitorare efficacemente i sistemi ad alto

[79] Per una disamina onnicomprensiva sul tema si veda G. Natale, *Intelligenza artificiale, neuroscienze, algoritmi aggiornato al nuovo regolamento europeo AI Act*, Pisa, 2024.

[80] Sulle meccaniche algoritmiche "*black box*" si è scritto molto, soprattutto negli ultimi anni in relazione ai sistemi utilizzati dalle piattaforme digitali. *Ex multis* si segnalano: J. Burrell - Z. Tufekci**,** *Seeing with Algorithms: How Data Science and Machine Learning Shape and Limit Human Understanding*, in *Communication Studies*, 70(3), 2019, 270-287; E. Finn, *The Black Box of the Present: Time in the Age of Algorithms*, in *Social Research*, 86(2), 2019, 557–580; V. Hassija et. Al., *Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence*, in *Cognitive Computation,* 16, 2024, 45–74.

[81] Ufficio europeo per l'IA.

[82] H. Woisetßchläger et Al., *Federated Learning and AI Regulation in the European Union: Who is Responsible? -- An Interdisciplinary Analysis*, in *24° Workshop at the 41 st International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024; *There Are Holes in Europe's AI Act — and Researchers Can Help to Fill Them*, in *Nature,* 625, 2024, 216.

rischio, compromettendo l'efficacia della normativa e rendendo puramente teorico il richiamo al controllo umano delle risorse digitali[83]. Il timore, quindi, è che anche questa normativa non sia davvero "umano-centrica"[84].

4. Cosa resta dell'Uomo? L'educazione ai diritti fondamentali come soluzione tecnica e sociale

Giunti al termine di questa disamina, rimane da chiederci cosa resti dell'Essere umano in questa transizione digitale.

Si è potuto constatare come l'Unione Europea abbia cercato di rivestire un ruolo di capofila nello sviluppo dei sistemi tecnologici, imponendo, dove si poteva, regole che guardano al futuro di questa materia, smussando le criticità derivanti dagli abusi dei nuovi mezzi automatizzati. Il c.d. "*Bruxelles effect*"[85] rappresenta proprio quanto descritto: ossia l'azione di indirizzo dell'opinione pubblica finalizzata all'emulazione da parte dei legislatori esteri. Si tratta di un modo per creare sia un'egemonia legislativa che - in senso lato - un'egemonia culturale[86] sulla tematica.

L'espressione forte che abbiamo voluto utilizzare poco prima non rappresenta un'esagerazione, poiché, ragionandoci sopra, il diritto non può trovare una forte presa nella società se prima non viene assimilato dagli elementi che formano lo stesso tessuto sociale. In questo discorso conclusivo, dunque, vanno distinte due dimensioni, due cerchi concentrici che costituiscono il quadro su cui deve operare l'attività di elaborazione sul tema: ci si riferisce al settore dell'ingegneria informatica e alla società umana nella sua interezza[87].

Per quanto riguarda la prima categoria menzionata, il ritorno all'Essere umano si pone come cambio di paradigma necessario per la costruzione materiale di macchine che non rappresentino un pericolo per la generalità delle persone. Va prima di tutto con-

[83] H. Fraser - J. M. Bello y Villarino, *Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough,* in *European Journal of Risk Regulation*, 15, 2024, 431–446.

[84] Come già si chiosava anni addietro in G. De Gregorio - F. Paolucci - O. Pollicino, *L'intelligenza artificiale made in Ue è davvero "umano-centrica"? I conflitti della proposta*, 22 luglio 2021.

[85] Per antonomasia si suole attribuire la prima teorizzazione di questa tesi ad Anu Bradford in A.Bradford, *The Brussel Effect. How the European Union Rules the World*, Oxford, 2020.

[86] In questo caso si usa la classica espressione che fa capo all'ideologia marxista svuotandola del suo costrutto rivoluzionario. Come anche nella teorizzazione di eredità gramsciana, in questo caso si vuol propugnare la necessità di un cambiamento di tipo ideologico nei confronti di una materia che è sempre stata segnata da un generico liberalismo senza freni. Il cambio di paradigma prevederebbe una conversione di questo *status quo* verso una politica incentrata sul controllo statale e sulla salvaguardia dei diritti fondamentali degli individui. Grazie all'esempio positivo rappresentato da questo nuovo assetto del sistema, successivamente, sarebbe possibile intravedere scelte dello stesso segno anche nelle legislazioni degli Stati extra-UE. Per il riferimento alla teoria dell'egemonia culturale si rimanda a A. Gramsci, *Quaderni dal carcere, vol. III*, Torino, 2014, 2010 ss.

[87] Al termine di questo periodo si spiega l'uso della figura dei cerchi concentrici: le teorie e i moniti che si riporteranno di seguito sono validi per gli ingegneri in senso duale in quanto essi, oltre a rappresentare una certa *species* nella società, sono anche parte della cittadinanza; d'altra parte, invece, i moniti per il cittadino comune lo riguardano in senso singolare, in quanto questo non è necessariamente provvisto del *know how* tipico della scienza ingegneristica.

siderata la constatazione per la quale la tecnologia e il progresso non rappresentino in ogni caso dei fattori neutrali nell'equazione della convivenza sociale.
Il reale progresso non può che essere declinato secondo il principio di un miglioramento sostanziale della condizione umana. Per questa ragione, esso non può essere asservito alla teoria per la quale un certo sviluppo debba essere portato avanti sulla semplice base della sua idonea capacità ad esistere. Una creazione di questo tipo, per quanto affascinante, si rivelerebbe priva di educazione e coscienza, dunque anche più prona a diventare dannosa se utilizzata in maniera sconsiderata.
Questa riflessione la si può leggere sia nel senso assoluto del progresso scientifico ma anche nel senso microscopico della manifestazione del pensiero. Se si ripensa a come sono stati creati gli algoritmi di controllo dei contenuti su piattaforma e a come essi si sono comportati negli anni, ben si può notare come sia mancato un *input* di base che corrispondesse ad un'adeguata educazione al valore del pluralismo. Queste macchine non sono state costruite secondo un principio giuridico chiaro, esse sono rimaste ancorate ai *bias* e alle credenze dei loro stessi creatori[88]. Queste le ha rese fallaci e ha portato all'attenzione del pubblico la questione; qualcuno, per questo, potrebbe pensare che, in fondo, sia stato un bene che quanto riportato sia avvenuto. Tuttavia, se in fase di costruzione del mezzo si fossero seguiti i principi giuridici tipici degli ordinamenti liberal-democratici - riprendendo e abbracciando il concetto del *Legal protection by design* - probabilmente non sarebbe accaduto nulla di tutto ciò.
La responsabilità dell'Essere umano che svolge un'attività di creazione, quindi, è cruciale in questo frangente, e la collaborazione tra le diverse conoscenze - come quella del giurista con l'ingegnere - si può rivelare salvifica per la società. E proprio in questo contesto, la spiegazione e l'innalzamento a pilastro del principio dell' **"***human in the loop***"** assume un ruolo centrale[89]. Esso, infatti, implica che, anche nelle tecnologie più avanzate, sia sempre garantita la supervisione e l'intervento umano nei processi decisionali critici. Tale approccio si traduce in un sistema in cui la macchina non agisce in maniera autonoma e incontrollata, ma opera sotto il controllo consapevole di un operatore umano. Questa supervisione non solo riduce i rischi legati a errori sistemici o a decisioni arbitrarie, ma permette anche di calibrare meglio gli strumenti tecnologici sulle esigenze della società e sui principi democratici, come il pluralismo e l'inclusività. Il concetto precedentemente detto, quindi, potrebbe rappresentare un punto d'incontro tra le istanze di innovazione e la necessità di preservare il valore umano al centro del progresso tecnologico. E, soprattutto, potrebbe rappresentare la sublimazione di quella volontà politica che è stata inserita negli interventi normativi affrontati nei paragrafi precedenti. Proprio attraverso un principio chiaro che faccia sì che l'essere umano pos-

[88] Sulla questione dei *bias* cognitivi riflessi sugli algoritmi si è scritto molto, soprattutto a seguito di casi giudiziari di grande clamore come Loomis v. Wisconsin. Tuttavia, si rimanda a delle letture non giuridiche per avere uno sguardo d'insieme sulla tematica; Cfr. C. Bartneck - C. Lütge - A.Wagner - S. Welsh, *An Introduction to Ethics in Robotics and AI*, Berlino, 2021; F. Pethig - J. Kroenung, *Biased Humans, (Un)Biased Algorithms?*, in *Journal of Business Ethics*, 183(3), 2023, 637–652.

[89] Alcuni riferimenti bibliografici sul tema: I.P. Di Ciommo, *La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello Human In The Loop*, in *Federalismi*, 9, 2022, 68-90; X.L. Meng, *Data Science and Engineering with Human in the Loop, Behind the Loop, and Above the Loop*, in *Harvard Data Science Review*, 5(2), 2023; D. Martire, *Human in the loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale*, in *Rivista italiana di informatica e diritto*, 2, 2024.

sa essere solo un fattore condizionante e non condizionato dall'intelligenza artificiale, scegliendo attivamente tra l'essere e il dover essere costituzionale[90].

Diverso è il discorso per quanto riguarda il tessuto sociale considerato in senso generale. In questo caso, la questione su cui si va a posare la nostra riflessione riguarda la capacità di generare all'interno della cultura di massa gli "anticorpi" contro le distorsioni derivanti dagli abusi della tecnologia. Il diritto alla manifestazione del pensiero e il diritto a poter usufruire di un'informazione pluralista rappresentano i pilastri su cui si basa la convivenza civile democratica, in quanto essi fanno sì che le varie anime di un Paese possano esprimersi e mostrare all'esterno le loro idee. Il loro continuo rafforzamento è necessario affinché queste "abitudini" non vengano dimenticate e lasciate indietro dinanzi ad una nuova realtà su cui ancora non abbiamo mezzi di comprensione adatti e su cui buona parte della società non è ancora educata all'uso. Solo con l'aiuto dei valori costituzionali si può pensare di affrontare il cambiamento con una consapevolezza che renda immuni dalle regressioni democratiche[91] e dagli abusi della tecnologia. Soltanto attraverso un compromesso culturale e una mentalità aperta potremo abbracciare il progresso e saperlo gestire affinché l'Uomo e i suoi diritti ne escano non solo intatti ma addirittura rafforzati e si realizzi quella collettività solidale di individui liberi e responsabili che così bene definisce la nostra Costituzione .

[90] D. Martire, *ibid.*

[91] È noto come la mancanza di libertà di pensiero sia l'anticamera per un declino del sistema democratico, in quanto comporta l'impoverimento del dibattito pubblico-politico, il ristagno delle idee e l'assenza di confronto costruttivo. A questo riguardo, in relazione agli stessi social media di cui si è trattato in questo testo, si segnala la lettura di L. C. Bollinger - G. R. Stone (a cura di), *Social Media, Freedom of Speech, and the Future of Our Democracy*, Oxford, 2022. Nonostante il testo ponga il suo *focus* principale sugli Stati Uniti d'America, esso si rivela illuminante e di sicura rilevanza per tutti quei sistemi ascrivibili alle "democrazie occidentali", i quali vanno sempre più incontro ad una condizione che potremmo denominare di "Estraneità", la quale, citando Francesco Piccolo, "rende impermeabile la conoscenza"; F. Piccolo, *Il desiderio di essere come tutti*, Torino, 2013, 251.

Governing Social Media's Opinion Power: The Interplay of EU Regulations*

Urbano Reviglio, Konrad Bleyer-Simon, Sofia Verza

Abstract

Mainstream social media platforms nowadays play an unprecedented gatekeeping role in public discourse. Through rules set by design, terms and conditions, content moderation practices and algorithmic systems, these platforms wield significant influence over individual users and public opinion. While the control over social media's gatekeeping function was initially entrusted to self-regulation under the presumption of neutrality, it is increasingly acknowledged that their role is pivotal for the whole media ecosystem. This article examines the evolving policy landscape in EU media governance, offering a retrospective analysis of how policymakers have addressed social media's "opinion power"— and a prospective analysis of the most recent regulatory developments, most importantly the Digital Services Act and the European Media Freedom Act. Finally, we conclude highlighting limitations, challenges, and opportunities of the EU regulatory framework.

Table of contents

Keywords

social media – platform governance – EU media policy – media regulation – content moderation – public opinion

1. Introduction

Since the early 2000s, the control over the gatekeeping role of social media platforms has been left to self-regulation presuming their content neutrality[1]. As a matter of

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

[1] Note that in this article the terms "social media" and "platforms" are used interchangeably to

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

fact, they were considered technology providers, not media companies[2]. Nowadays it is commonly recognized that social media's content moderation and content curation fulfill a fundamental role in the overall media ecosystem, so their role is not to be considered neutral at all. Indeed, through prioritization, removal, labeling, amplification or reduction of content visibility, social media make "editorial" choices which are key in defining which information people can see in their social media feeds[3]. Though these actions may not fit the classic definition of editorial choices and lack comparable regulation, they eventually influence not only individuals and public opinion but also how news is produced, distributed and consumed, as well as the practices of political communication and the formation of users' political preferences.[4]

The increased awareness of the risks that an unregulated social media environment can pose to democracy, consequently induced a shift in internet governance, including a focus on a range of platforms and especially social media.[5] The European Union (EU) led the way in the global regulatory landscape. In 2016, it regulated the protection of privacy and the handling of data with the General Data Protection Regulation (GDPR)[6]. In 2018, it has initiated a self-regulatory endeavor to curb disinformation with the first Code of Practice on Disinformation (CoP)[7] – later extended in the Strengthened Code of Practice on Disinformation (2022), soon to be transformed into a co-regulatory code of conduct. In 2022, the Digital Services Act (DSA)[8] and Digital Market Act (DMA)[9] introduced a set of obligations to safeguard competition, as well as set the ground for the safety and transparency of the online environment, while the European Media Freedom Act (EMFA)[10], enacted in 2024, includes among other things, protections for journalistic content published through social media. In the same year, the European Commission enacted a Regulation on the transparency

generally refer to large social media companies. While these often include platforms classified as "Very Large Online Platforms" under the EU Digital Services Act, the usage extends to major social media platforms that might not meet this specific regulatory classification.

2 R. Caplan-P. M. Napoli, *Why media companies insist they're not media companies, why they're wrong, and why it matters*, in *Medias Res*, 60, 2018.

3 T. Gillespie, *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*, New Haven and London, 2018.

4 M. Moore-D. Tambini, *Regulating Big Tech: Policy Responses to Digital Dominance*, New York, 2021.

5 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

6 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

7 EU 2022 *Strengthened Code of Practice on Disinformation*.

8 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

9 Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

10 Regulation (EU) 2024/1083 of the European Parliament and of the Council establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act), 20 March 2024.

and targeting of political advertising (TTPA)[11], which is particularly relevant in this context as a significant share of content on social media is sponsored, and even organic content can be amplified through payments. The measures just listed were complemented by a set of guidelines and official communications of the Commission, as well as efforts in neighboring fields, such as the Artificial Intelligence Act (AIA, 2024)[12] that introduces transparency obligations for artificial intelligence services.
This article discusses the policy shift occurring in this area of platform governance by analyzing the emerging European legal framework, and how this regulates the "opinion power" of social media, and the challenges this entails for media pluralism.[13] The research question guiding this analysis is the following: how are the new EU regulations and related policies expected to influence the governance of social media's internal pluralism, particularly users' diversity of exposure? To answer this question, in Chapter 2 we discuss what social media's opinion power is and how we conceive it. This allows us to disassemble the main components of this power, specifying how social media affect internal pluralism and diversity of exposure. Chapter 3 offers an overview of how such opinion power has been regulated in the last two decades in the EU to highlight the challenges faced, and which ones are still relevant. In Chapter 4, we analyze how the emerging European governance regime regulates such opinion power and, in Chapter 5, we discuss whether this is capable of properly redistributing opinion power between different stakeholders. Finally, preliminary conclusions are drawn.

2. Unpacking social media's opinion power

Historically, media exerted an extraordinary influence over the formation of individual and public opinion[14]. Such influence unfolds in various ways; not only by informing but also by entertaining, distracting, and persuading citizens (i.e., infotainment and propaganda), by facilitating or constraining the diffusion of information (e.g., gatekeeping theory), by choosing which political issues are most salient (i.e., agenda setting theory), or by indirectly dissuading (perceived) minoritarian opinions to be expressed (i.e., the spiral of silence theory). To regulate and minimize these forms of

[11] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising.

[12] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[13] It is important to note that, while this article examines the legal interconnections opportunities related to the governance of the opinion power exercised by social media platforms to outline the main challenges and opportunities, it does not aspire to provide a comprehensive analysis of this complex and evolving landscape.

[14] M. E. McCombs-D. L. Shaw, *The Agenda-Setting Function of Mass Media,* in *The Public Opinion Quarterly*, 36 (2), 1972, 176 ss.; E. Noelle-Neumann, *The Spiral of Silence. Public Opinion-Our Social Skin*, Chicago, 1972; N. Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*, 1985; N. Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media*, 1988.

influence and favor a diverse media environment, the European policy community has traditionally put an emphasis on media pluralism, namely ensuring the availability of a range of different points of view in the media environment[15]. This principle is enshrined in international documents, such as in art. 11(2) of the Charter of Fundamental Rights of the European Union.

In European media policy debates, media pluralism has usually been invoked around issues of media ownership concentration, the role of public service media, and media subsidies[16]. It is a multi-faceted notion that can be understood in various ways. Three dimensions are probably the most critical ones; "external pluralism" which refers to how plural the structure of the media market is; "internal pluralism" which generally refers to the plurality of content and viewpoints that are provided by a single media company; finally, adapting the latter to the digital environment where content abundance has led to the deployment of personalization algorithms, the focus shifted to the diversity of content that users are ultimately exposed to, what is referred to as "exposure diversity"[17]. Nowadays much of the public debate occurs online and much of the news is indeed accessed or found through social media.[18] Much discussion has centered around the negative effects of social media to political discourse, notably the reduction of information diversity to which individuals are exposed to and that they eventually consume (so-called filter bubbles and echo chambers)[19] and the amplification of disinformation, conspiracy theories, and sensational and divisive content.[20] Of course, we still lack conclusive evidence about the impact these phenomena ultimately have, and it is not even clear how exposure diversity could be achieved given the notion of 'diversity' is multidimensional and unsuited for algorithmic operationalizations.[21] This is why in this paper we opted to focus more broadly on social media's "opinion power", how this may affect exposure diversity, and how – from a regulatory perspective – such power could be constrained, or even leveraged to promote more diversity.

---

[15] European Commission: Directorate-General for Communications Networks, Content and Technology, P. Parcu-E. Brogi-S. Verza, et al., *Study on media plurality and diversity online* – Final report, Publications Office of the European Union, 2022.

[16] K. Karppinen, *Problem definitions in European policy debates on media pluralism and online platforms*, in T. Dwyer-D. Wilding, *Media Pluralism and Online News*, Bristol, 2023, 96 ss.

[17] N. Helberger-K. Karppinen-L. D'acunto, *Exposure diversity as a design principle for recommender systems*, in *Information, communication & society*, 21(2), 2018, 191 ss.

[18] N. Newman-R. Fletcher-C.T. Robertson-K. Eddy-R. Kleis-Nielsen, *Reuters Institute Digital News Report 2023*.

[19] The two phenomena are similar but substantially different. Filter bubbles are conceived as cultural and ideological bubbles in which individuals continue to see and consume content that reinforces its opinions and interests. Echo chambers refer to a group situation where established information, ideas, and beliefs are uncritically spread and amplified, while dissenting views and arguments are ignored. The crucial difference is that the former may not depend on the user's autonomy and awareness – therefore it is mainly caused by technological affordances – while the latter pre-exists the digital age and thus it is primarily driven by social relations.

[20] U. Reviglio, *The Algorithmic Public Opinion: a Policy Overview*, in *osf.io,* 5 October 2022.

[21] F. Loecherbach-J. Moeller-D. Trilling-W. van Atteveldt, *The unified framework of media diversity: A systematic literature review*, in *Digital Journalism*, 8(5), 2020, 605 ss.

But what is "opinion power"? It can be generally defined as the ability to influence the processes of individual and public opinion formation. This is not a new term but "a normatively and constitutionally rooted notion that captures the core of media power in democracy and substantiates why that power must be distributed"[22]. Social media platforms, however, have given rise to new forms of influence and dominance over public opinion. How social media's opinion power eventually is exercised, how it differs from traditional gatekeeping power, and what legal provisions and policy interventions are needed to redistribute such power in line with the new and updated normative frameworks is still unclear. These questions also require continuous reexamination in the rapidly changing media-landscape.[23] . It is out of the scope of this article to systematically examine such concept[24] Nonetheless, we briefly describe our current understanding, guiding our subsequent analysis. We outline, on the one hand, the key elements in which such opinion power is exerted and, on the other hand, how this could play out beyond traditional gatekeeping, salience, and exposure.

To effectively understand social media's opinion power and analyze the impact of EU regulations, it is useful to deconstruct the concept of opinion power it as being composed of four main intertwined components: platform design, terms of service, content moderation, and content curation[25]; (i) "Platform design" is the informational architecture — most importantly the user interface design[26] — which prescribes and favors certain behaviors instead of others (also referred to as "affordances"[27]) which may eventually affect news exposure and consumption; (ii) the "terms of service" of social media services represent a form of privatized governance of the rules between the platform and the user, and between users.[28] This also includes the "community guidelines" that platforms usually set and regularly update, and that also define rules that ultimately prohibit certain content and behaviors and shape content moderation practices; (iii) "content moderation" involves all the actions and strategies undertaken by platforms to moderate content according to the platforms' rules and to national laws.[29] This also includes diverse algorithmic systems detecting content to moderate; in

[22] *Ibid.*

[23] J. Schlosberg, *Digital Agenda Setting: Re-examining the Role of Platform Monopolies*, in M. Moore-D. Tambini (eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, New York, 2018, 202 ss.

[24] For such debate see T. Seipp et al., *Dealing with opinion power in the platform world*, cit.

[25] These components are all explicitly defined in the Digital Services Act and they are provided in footnotes.

[26] Art. 3(m) DSA defines "online interface" as «any software, including a website or a part thereof, and applications, including mobile applications». Admittedly, this definition is rather broad and does not highlight the design ability to prescribe values and afford actions.

[27] T. Bucher-A. Helmond, *The affordances of social media platforms*, in *The SAGE handbook of social media*, 2018, 233 ss.

[28] Art. 3(u) DSA defines "terms and conditions" as «all clauses, irrespective of their name or form, which govern the contractual relationship between the provider of intermediary services and the recipients of the service».

[29] T. Gillespie, *Custodians of the Internet,* cit., 4; To be clear, content moderation on social media extends beyond content that is posted by users but includes the moderation of user accounts, comments, direct messages (DMs), live streams, advertisements, hashtags, search results, user interactions, content

particular, (iv) the algorithmic systems that recommend content (i.e., "recommender systems") are especially influential as they embed specific values of content curation, ultimately determining news visibility and exposure.[30] Of course, there are additional mechanisms that can influence social media's opinion power, such as monetization and advertising models or partnerships and publisher deals. However, these may have more indirect effects and are more closely related to issues of competition and copyright. Exploring how the four highlighted components are regulated provides a more practical and straightforward approach to gain insights into the governance of social media's influence on exposure diversity.

It is important to acknowledge how social media's opinion power is not limited to "which" content users see and "how much" exposure they (do or don't) have but, importantly, "how" and "when" content is shown or discovered. Social media have been reported to design their interfaces and order content in specific ways to affect users' behaviors and optimize their engagement.[31] The resulting addictive power and its ability to influence news consumption habits is still an underexplored area. Social media companies regularly conduct experiments on users and test algorithmic changes (i.e., A/B testing).[32] The understanding derived and its potential to enable them to imperceptibly influence patterns of news consumption is evident, though far from being fully understood. Think of the potential manipulative power of social media's recommender systems. By recommending content that users find strongly disagreeable, they can increase political polarization[33]; by recommending a lot of conflicting news accounts, they can generate "reality apathy" (i.e., people do not care about what is

flagging, bot activity, amongst others, ensuring a comprehensive content oversight. Art.3(u) DSA defines "content moderation" as «the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or that affect the ability of the recipients of the service to provide that information, such as the termination or suspension of a recipient's account».

[30] Art. 3(s) DSA defines "recommender system" as «a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service or prioritize that information, including as a result of a search initiated by the recipient of the service or otherwise determining the relative order or prominence of information displayed».

[31] V. R. Bhargava-M. Velasquez, *Ethics of the attention economy: The problem of social media addiction*, in *Business Ethics Quarterly*, 31(3), 2021, 321 ss.

[32] See R. J. Deibert, *The road to digital unfreedom: Three painful truths about social media*, in *Journal of Democracy*, 30(1), 2019, 25 ss; the persuasive and manipulative capability of social media can also be exemplified with two famous experiments conducted by Facebook more than a decade ago; one is the experiment on 61-million-person in social influence and political mobilization that showed their potential to nudge citizens to vote (R. M. Bons-C. J. Fariss-J. J. Jones-A. D. Kramer-C. Marlow-J. E. Settle-J. H. Fowler, *A 61-million-person experiment in social influence and political mobilization*, in *Nature*, 489(7415), 2012, 295-298); the other is the infamous study on the "emotional contagion effect" that showed how small changes in the algorithms can manipulate emotions on a mass level (A. D. Kramer-J. E. Guillory-J. T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, in *Proceedings of the National Academy of Sciences*, 111(24), 2014, 8788-8790).

[33] C. A. Bail-L. P. Argyle-T. W. Brown-J. P. Bumpus-H. Chen-M. F. Hunzaker-A. Volfovsky, *Exposure to opposing views on social media can increase political polarization*, in *Proceedings of the National Academy of Sciences*, 115(37), 2018, 9216-9221.

true or not)[34]; conversely, by suggesting conspiracy theories, they nudge users towards making sense of a complex reality through emotionally appealing – albeit simplistic or even untrue – explanations (e.g., the "rabbit hole effect" on YouTube)[35]; by changing the order of political candidates information in search queries, they can manipulate voting intentions, or even favor a specific viewpoint on a topic to people who have not yet formulated a strong opinion (e.g., the "search engine manipulation effect")[36]. Online platforms can influence what information their users consume in other subtle ways as well, notably through the search suggestions or autocomplete functions.[37] For example, TikTok has been recently shown to provide search suggestions through the "Others Searched For" function that may lead to questionable information or send users down contentious political rabbit holes.[38] Furthermore, it should be considered that only a minority of users regularly and proactively look for political news online, and as such, most users encounter political news in social media only or mainly incidentally.[39] This grants these platforms greater control over whether and how much political news a user receives, potentially creating "social media news deserts" where certain user groups are minimally or not at all exposed to political and public affairs content.[40] In this context, it is also important to mention the activities of third-party actors (foreign interference, bot activities, etc.), whose activities are aimed at manipulating other users, as platform design and the choice (or lack) of responses can have a significant impact on users' exposure to content. To effectively regulate the opinion power of social media, it is crucial to address its impact on the exposure and the consumption of diverse content in the short- as well as in the long-term, ultimately influencing not only opinion formation but, more broadly, and more subtly, individual and collective worldviews.

3. A brief history of EU social media governance

Historically, the debate on social media governance pivoted on the question of whether platforms can be held accountable for the content shared through them, legally and ethically. The European liability regime was influenced by the U.S. model of the Com-

[34] L. Thorburn-J. Stray-P. Bengani, *Is Optimizing for Engagement Changing Us?*, 10 October 2022.

[35] M. Yesilada-S. Lewandowsky, *Systematic review: YouTube recommendations and problematic content*, in *Internet policy review*, 11(1), 2022.

[36] R. Epstein-J. Li, *Can biased search results change people's opinions about anything at all? A close replication of the Search Engine Manipulation Effect (SEME)*, in *Plos one*, 19 (3), 2024.

[37] ; Autocomplete functions have been already subject of a case law in 2013 before the German Federal Court of Justice. See S. Wünsch, *Google don't complete. Deutsche Well*, in *dw.com*, 15 May 2013.

[38] M. Schüler-M. Degeling,-S.Romano-K. Meßmer, *Other search for … the opposition party,* in *tiktok-audit.com,* 16 July 2024.

[39] R. Fletcher-R. K. Nielsen, *Are people incidentally exposed to news on social media? A comparative analysis. New Media & society,* 20(7), 2018, 2450 ss.

[40] M. Barnidge-M. A. Xenos, *Social media news deserts: Digital inequalities and incidental news exposure on social media platforms*, in *New Media & Society*, 26(1), 2024, 368 ss.

munications Act, Section 230 (1996).[41] Under the E-Commerce Directive (2000/31/EC)[42], the early 2000s were marked by a relatively hands-off approach to online content. The case law of European Courts has been important in tracing the borders of platforms' liability exemptions, especially with regards to intellectual property rights and thanks[43] . In this situation, much of online content moderation remains at the discretion of platforms also in the EU. The set of rules that users are required to follow on the services are shared with them through terms of service, community guidelines and other internal policies. These rules are usually prepared and updated by the legal teams of the platforms – generally based in the country where the platform's headquarters are established, and reflecting the values and legal environment of that country — which is in most cases the United States.[44] .

Many of today's large online platforms started out as startups without clear use purposes or monetization strategies, not to mention an understanding of the potential role they might play in society. As such, their rules were developed "in an *ad hoc* manner", often by reacting to some imminent threats, among other things, driven by a «desire to prevent fraud, to assuage advertisers, avoid lawsuits»[45]. This approach can be also referred to as the "first wave" of content moderation governance[46]. In the mid to late 2000s and more prominently around the 2010s, detection algorithms and more proactive efforts towards content moderation were undertaken. Moreover, platforms started adapting these rules to the national legal frameworks they operate in, adding to the list of non-acceptable behaviors those considered illegal or harmful in specific country contexts. While some non-tolerable activities that might lead to certain forms of punitive action against the content or its publisher were illegal in most jurisdictions (such as inciting hatred), platforms also had the opportunity to ban certain legally acceptable forms of conduct on their platforms, from pictures containing nudity[47] and

[41] Notably, this was supported by the assumption that the size and speed of online user activity made it impossible to monitor online content effectively. Platforms in the U.S. still nowadays enjoy the same status as Internet Service Providers (ISPs), considered as mere hosting providers of content created by someone else. In addition to Section 230, Section 512 of the Digital Millennium Copyright Act (DMCA) is also relevant, providing the "Safe Harbor Protection" to online platforms, shielding them from liability for copyright-infringing content uploaded by users, provided they comply with notice-and-takedown procedures.

[42] In particular, art. 14 under the E-Commerce Directive (2000/31/EC) requires action from platforms when they become aware of the existence of illegal content on their services. Platforms must act quickly to remove illegal content to maintain their liability protection.

[43] See for example EctHR, *Delfi AS v. Estonia,* app. no. 64669/09 (2015); CJEU, C-324/09, *L'Oréal and Others* (2011) in contrast to *Tiffany (NJ) Inc. v. eBay Inc.* 600 F.3d 93 (2nd Cir.2010); CJEU, C-131/12, *Google Spain* (2014); T-201/04, *Microsoft v. Commission*, mentioned in M. Cantero Gamito, *Regulation of Online Platforms* in J. Smits-J. Husa-C.Valcke-M. Narciso, *Elgar Encyclopedia of Comparative Law*, Cheltenham-Northampton, 2021.

[44] R. Gorwa, *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*, Oxford, 2024, 21.

[45] *Ibid.*, 13.

[46] This was basically characterized by privatized vertical procedures that apply legislative-style rules drafted by platforms to individual cases and hears appeals from those decisions, deciding on a case-by-case basis. This approach still dominates content moderation practices nowadays. See E. Douek, *Content moderation as systems thinking*, in *Harvard Law Review*, 136 (2), 2022, 526 ss.

[47] Inevitably, sometimes users oppose such content moderation decisions. Consider the #Freethenipples

satire, for example, to the publication of fabricated and misleading information, that got to be known later as disinformation.

As social media platforms grew massively, the self-regulation approach has been increasingly questioned. The international policy community, media activists and scholars began to be concerned of the rising gatekeeping power of mainstream social media platforms, both for the moderating practices themselves and for their opacity, indicating a growing consensus for a regulatory change [48]. In this period an historical shift took place where traditional forms of speech regulation, which typically involve legal restrictions and eventually government censorship (what has been described as "old-school speech regulation") contrast with newer forms that arise from the nature of social media platforms, algorithmic content moderation, and the actual privatization of public discourse (i.e., "new-school speech regulation") [49].

European policymakers, civil society and scholars soon had to realize how complicated such exercises were, as there is indeed content on social media that does not violate existing laws but still could cause significant harm to society. In particular, policy makers were concerned about the proliferation of untrue statements that might risk the integrity of elections or hamper public health authorities' efforts to handle pandemics. Indeed, content moderation has always been particularly challenging as it requires striking a balance between free speech and other conflicting interests, such as reputation, public safety, or crime prevention, taking into account [50] At the same time, there has been a lack of specific forms of regulation that could be used as a role model for content on social media, as traditional media laws and policies are often inadequate in light of and the emergence of new actors involved in it. The calls for increased action created an environment in which platforms proactively formulated their own rules on what is allowed and what is unacceptable on their services – [51].

In 2015, the European Commission published its Digital Single Market Strategy[52], preluding the enactments of the DSA, DMA, AIA and EMFA. A pivotal shift towards a more active governance of social media was marked by the voluntary EU Code of conduct on countering illegal hate speech online (2016)[53]. Major platforms like Facebook, Twitter, and YouTube agreed to review and remove illegal hate speech within 24 hours. This initiative reflected the EU's growing concern for the societal impacts of

movement which gained traction particularly around mid 2010s and advocated for the normalization of female breasts challenging societal restrictions on their visibility. It generally argued that the censorship and sexualization of female breasts are discriminatory and perpetuate harmful stereotypes.

48 E. Morozov, *The net delusion: The dark side of Internet freedom,* New York City, 2012; C. Fuchs, *Social Media: a Critical Introduction*, 2013; T. Gillespie, *Custodians of the Internet,* cit.; J. Meese- S. Bannerman, *The Algorithmic Distribution of News,* in *Policy Responses,* 2022.

49 J. M. Balkin, *Old-school/new-school speech regulation*, in *Harvard Law Review,* 127, 2013, 2296.

50 I. Nenadić-S. Verza, *European Policymaking on Disinformation and the Standards of the European Court of Human Rights*, in E. Psychogiopoulou-S. de la Sierra, *Digital Media Governance and Supranational Courts Selected Issues and Insights from the European Judiciary*, Cheltenham, 2022, 175 ss.

51 T. Flew, *Regulating Platforms*, Cambridge, 2021.

52 Communication COM(2015) 192 final from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 6 May 2015 on A Digital Single Market Strategy for Europe.

53 See *EU Code of Conduct on Tackling Illegal Hate Speech*, 2016.

online hate, setting the stage for more stringent measures.[54] At times, national measures also played a role in this process: in 2017 the German Network Enforcement Act (NetzDG) introduced strict requirements for social media platforms to remove "obviously illegal" content within 24 hours under threat of fines of up to €50 million. This law was a critical moment for EU content moderation policy, emphasizing the responsibility of platforms in policing content, and influenced discussions across Europe about balancing free speech with the need to control harmful online behavior.[55]
In 2018, the revelation that the company *Cambridge Analytica* had harvested the personal data of millions of Facebook users without consent and used it for political profiling and targeting – among others, on behalf of pro-Brexit and pro-Trump campaigners – catalyzed a broader reckoning regarding the role of digital platforms in societal harm, mis- and disinformation, and electoral interference. Following this, the EU's Code of Practice on Disinformation (CoP) was introduced in 2018, addressing the spread of manipulative content in the context of new techniques and tactics, reflecting concerns exacerbated by the scandal's revelations. This policy instrument, however, was still non-binding and voluntary. In 2018, the General Data Protection Regulation (GDPR) came into effect, representing a fundamental step in strengthening human rights in the digital realm as well as illustrating the growing EU digital regulatory worldwide influence (i.e., Brussels effects)[56].
In this same period, the debate on content moderation practices became particularly intense in relation to "high-intensity events" (elections, terrorist attacks, natural disasters, pandemics), as certain kinds of content communicated in these contexts, even if otherwise legal, may pose a risk to human life and interfere with human rights, including the right to receive and impart information and to form and develop an opinion[57]. Key examples are concerns related to COVID-19 disinformation[58] and geopolitical pressures for information sovereignty. The latter have been on the European policy agenda at least since 2015, when the European Council asked the High Representative of the EU for Foreign Affairs and Security Policy to address information manipulation attempts originating from Russia. This was followed by a number of measures

[54] Other relevant regulations affecting content moderation include: the Directive on combating terrorism 2017/541, which provides for similar obligations against public online incitement to acts of terrorism; The revised Audio-visual Media Service Directive (AVMSD) 2018/1808, which includes new obligations for video-sharing platforms to tackle illegal online content (such as terrorist content, child sexual abuse material, racism and xenophobia) and specific categories of hate speech; The Directive on Copyright in the Digital Single Market 2019/790 which establishes obligations for copyrighted-materials.

[55] R. Gorwa. *Elections, institutions, and the regulatory politics of platform governance: The case of the German NetzDG*, in *Telecommunications Policy*, 45(6), 2021, 102 ss.; V. Claussen, *Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation*, in *Rivista di diritto dei media*, 3, 2018, 110 ss.

[56] A. Bradford, *The Brussels effect: How the European Union rules the world,* New York, 2020.

[57] I. Nenadić-Verza, *European Policymaking on Disinformation,* cit.

[58] For example, the Communication on Tackling COVID-19 disinformation – Getting the facts right in 2020 as well as the European Democracy Action Plan, in the same year). See Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling Covid-19 Disinformation – Getting The Facts Right*, JOIN/2020/8 Final, of 10 June 2020.

related to so-called hybrid threats, such as the establishment of the East Strategic Communication Task Force in the same year.[59]

In the early 2020s, we entered a new era for content moderation governance, which reflects a growing understanding of the complexity of human communication online and the limits of previous moderation technologies. Facebook, for example, established in 2020 the "Oversight Board" as an independent body that can issue non-binding recommendations to Facebook and Instagram, or even make binding decisions on whether specific content should be allowed or not on the platforms[60]. The Digital Services Act (DSA), in particular, enshrines this policy paradigm shift. Enacted in 2022, it aspires to promote a European digital sovereignty by establishing systemic regulation of platforms like those operated by Meta or Alphabet to address their impact on public discourse regulating various governance areas previously almost untouched, including terms of service, content moderation, recommender systems, and interface design. In 2024, the European Union also enacted the European Media Freedom Act (EMFA), a legislation designed to safeguard media freedom and pluralism, partially on digital platforms too. The same year marked a significant milestone with the introduction of the Artificial Intelligence Act (AIA), the first major regulation globally to specifically govern the use of artificial intelligence, that aims to address risks specifically posed by AI applications. Across all these regulations, the governance of content moderation has evolved from a focus on removing explicitly illegal content to strengthening the users' autonomy, among other things through algorithmic transparency and data protection, and to mitigate the risk posed by harmful content, such as hate speech and disinformation, by considering them as symptoms of so-called systemic risks, that stem from the design or functioning of platforms.[61]

Despite this policy shift, the EU's response to the challenges posed by large social media platforms was rather slow. This can be traced back to the history of Internet devel-

[59] G. Abbamonte -P. Gori, *European Union*. In: O. Pollicino (ed.), *Freedom of Speech and the Regulation of Fake News,* Antwerp, 2023, 129 ss.

[60] Criticism stems from the fact that the OB depends on Meta's funding (it is funded by an independent trust established by Meta); also, it is not that diverse in terms of geographical and gender background; and despite being an innovative body for a company like Meta, its case-by-case ex post review of content moderation practices – setting precedents for future similar cases- exemplifies the standard picture approach of judicial review–style solutions, thus falling in the traditional approach of speech regulation. On the other hand, subjecting platforms' content moderation decisions to judicial review may be expensive and time-consuming. A relationship between the OB's rulings and the domestic legal frameworks consists in the obligation for Meta to implement the OB's ruling unless doing so "could violate the law" in the relevant jurisdiction; moreover, ideally the OB cannot review cases that clearly violate national laws and could make Meta and its employees legally vulnerable. See D. Wong- L. Floridi, *Meta's oversight board: A review and critical assessment*, in *Minds and Machines*, 33(2), 2023, 261 ss.

[61] The DSA, for example, outlines four main categories of "systemic risks". A first category concerns the risks associated with the dissemination of illegal content, such as the dissemination of child sexual abuse material or illegal hate speech. A second category concerns the actual or foreseeable impact of the service on the exercise of fundamental rights, including but not limited to freedom of expression and of information, including media freedom and pluralism. A third category of risks concerns the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, as well as public security. A fourth category of risks stems from similar concerns relating to the design, functioning or use, including through manipulation, of VLOPs and VLOSEs with an actual or foreseeable negative effect on the protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence.

opment itself which has been driven by the US and its values.[62] In the early days, social media were widely seen as a powerful force for good, promoting free expression, fostering connections among people, and driving a global democratic revolution. Once the US-based digital platforms came to dominate the market showing their negative consequences and potential risks, there was a lack of collective determination and financial resources, in addition to the complexities in the coordination of the EU's digital policies and the digital companies' lobbying pressure. Over time, a recurring cycle has been observed where public outrage over specific incidents temporarily disrupted the status quo, prompting platforms to make only superficial adjustments that appease public sentiment without leading to substantial regulatory changes[63]. Public pressure and sustained negative coverage, though, seem to have had a role in shaping the governance of online platforms[64]. Ultimately, the history of the regulation of social media's opinion power is not only the history of unprecedented socio-technical challenges, but also a history of U.S. technological and cultural hegemony over the EU, and how big tech companies' increasing influence over people's life have been slowly recognized, increasingly opposed and, finally, regulated.[65]

4. The interplay between EU regulations

At the time of writing, the DSA is the main EU regulation that can be deployed to govern social media's opinion power and its influence on media pluralism. It is complemented by other regulatory tools, primarily the European Media Freedom Act (EMFA, 2024), the Regulation on the Targeting and Transparency of Political Advertising (2024), the Artificial Intelligence Act (AIA) (2024), the Strengthened Code of Practice on Disinformation (CoP, 2022), the General Data Protection Regulation (GDPR, 2016), amongst others.

The DSA thus updates the European legal framework on platform liability, which was previously prescribed by Directive 2000/31 on E-commerce. It develops a more thorough regulation on the techniques and the decision-making processes employed for content moderation. It contains many reporting and transparency obligations[66] By establishing Very Large Online Platforms (so-called VLOPs) and Very Large Search

[62] H. Nieminen-C. Padovani-H. Sousa, *Why has the EU been late in regulating social media platforms?*, in *Javnost-The Public*, 30(2), 2023, 174 ss.

[63] M. Ananny-T. Gillespie, *Public platforms: Beyond the cycle of shocks and exceptions*, IPP2016 The Platform Society, 2016.

[64] N. Marchal-E. Hoes-K. J. Klüser-F. Hamborg-M. Alizadeh-M. Kubli-C. Katzenbach, *How Negative Media Coverage Impacts Platform Governance: Evidence from Facebook, Twitter, and YouTube,* in *Political Communication,* 2024, 1 ss.

[65] A similar dynamic has been empirically observed in the United Kingdom. See M. Kretschmer-U. Furgal-P. Schlesinger, *The emergence of platform regulation in the UK: an empirical-legal study,* in *Weizenbaum Journal of the Digital Society–special issue:"Democracy in Flux–Order, Dynamics and Voices in Digital Public Spheres"*, 2022.

[66] See, in particular, arts. 15 (Transparency reporting obligations for intermediary services), 24 (Transparency reporting obligations for online platforms) and 42 (Transparency reporting obligations) and the EU official website with the DSA transparency reports: *transparency.dsa.ec.europa.eu.*

Engines (VLOSEs)[67] as new legal subjects, the DSA acknowledged the special role and reach of social media, which therefore shall «pay due regard to freedom of expression and of information, including media freedom and pluralism » (Recital 47). Platforms that have at least 45 million monthly active users (10% of EU population) (art. 33) have additional duties regarding activities that directly and indirectly influence media pluralism. VLOPs and VLOSEs are required to conduct periodical assessments for "systemic risks" potentially caused by their services (art. 34) and to take appropriate measures to mitigate them (art. 35). Content that is harmful, but not necessarily illegal, as well as a number of systemic risks that represent a challenge to freedom of expression, media pluralism, informed citizenship and electoral integrity fall under the "duty of care" of VLOPs (Recitals 80-83). The DSA also lists factors to consider when assessing risks (such as content moderation systems, recommender systems, advertising systems, and "data related practices") and a list of possible mitigation measures (such as adapting interface design, terms and conditions, content moderation processes, etc. under art. 35 DSA).

Furthermore, the DSA evolves the traditional "Notice-and-Takedown" system establishing a more detailed "Notice-and-Action" framework (art. 16 DSA). As well as under the e- Commerce Directive (ECD, 2000)[68], any individual can notify companies about specific supposedly illegal content, and companies are in charge of deciding whether to remove the content or not. Furthermore, the DSA evolves the traditional "Notice-and-Takedown" system contained under the e- Commerce Directive (ECD, 2000), in which any individual could notify companies about specific supposedly illegal content, and companies were in charge of deciding whether to remove the content or not. This approach led to low quality notifications, increasingly sent by algorithms, and not humans, and to the over-removal of content to avoid liability and reduce costs. The DSA thus establishes a more detailed "Notice-and-Action" framework (art. 16 DSA)[69].

Positive developments can be observed in the DSA in this regard: firstly, a new subject is introduced: "trusted flaggers", which are entities recognized by online platforms for their expertise and reliability in identifying content that violates community standards or legal requirements (art. 22). Moreover, platforms shall also include in their general terms and conditions information on the restrictions they impose on the use of their services, applying them in a transparent and non-discriminatory manner and with due regard for fundamental rights such as the right to freedom of expression and the freedom and pluralism of the media (art. 14). Such information shall concern, inter alia, the policies, procedures, measures and tools used for the purposes of content moderation, including algorithmic decision-making and human verification, as well as the procedural rules of their internal complaint-handling system. Article 17 also

[67] Find here the *list of designated very large online platforms and search engines under DSA*.

[68] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce).

[69] A. de Streel-M. Husovec, *The e-commerce Directive as the cornerstone of the Internal Market, Policy* Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament, 2020.

requires social media to provide "statements of reasons", meaning a clear and specific justification for a series of restrictions imposed on the grounds that the information provided by the recipient of the service constitutes illegal content or is incompatible with their general conditions. These restrictions include removal of content, disabling access to content, or demoting content[70], suspension, termination or other limitations of service provision, monetization opportunities or even the whole account. Finally, a specific internal complaint-handling system and out-of-court dispute settlement bodies related with the notice-and-action mechanism have been introduced, updating art. 17 ECD. The dispute settlement bodies (art. 20 and 21 DSA), are established by the EU member states and the national Digital Service Coordinators, do not establish precedents, handle higher numbers of complaints and not just selected cases, and have no binding powers on platforms[71].

The DSA is complemented by the European Media Freedom Act (EMFA), especially for the protection of news media pluralism. The EMFA aims to protect and improve the media sector in the EU, given the crucial democratic implication of guaranteeing access to free, plural and independent news[72]. For the scope of this article, it is especially worth considering the inability of the media to fulfill their social role of providing quality and independent news if they are overwhelmed by the competition of digital actors that impact their revenues and news distribution[73]. Especially relevant in this context is art. 18 which regulates the relationship between VLOPs and news media services ("media service providers" (MSP), art. 2 EMFA), providing a so-called "special treatment for the media" or "media privilege"[74]. In particular, it regulates the suspension of the «provision of its online intermediation services» based on its terms and conditions, and outside of the systemic risk cases listed in the DSA. The goal is to

[70] Demotions are reductions of content visibility. Another common term is "shadowban" which specifically refers to demotions that have not been announced by a platform and, thus, they are only suspected. To better understand the topic, see T. Gillespie, *Do Not Recommend? Reduction as a Form of Content Moderation*, in *Social Media+ Society*, 8(3), 2022, and P. Leerssen, *An End to Shadow Banning? Transparency rights in the Digital Services Act between content moderation and curation*, in *Computer Law & Security Review*, 48, 2023, 105790.

[71] This process could also lead to issues related to workload, de facto limiting the efficiency and the effectiveness of ODS. The list of certified out-of-court dispute settlement bodies is available online.

[72] M. Monti, *The missing piece in the DSA puzzle? Article 18 of the EMFA and the media privilege*, in *Rivista Italiana di Informatica e Diritto*, 2, 2024.

[73] The proposal acts on several aspects: the concentration and transparency of media ownership, the governance of media policies by European and national regulatory authorities, journalists' safety from surveillance through spywares, editorial independence, and the relationship between media service providers and platforms in the context of content moderation. See E. Brogi-D. Borges-R. Carlini-I. Nenadic-K. Bleyer-Simon-J. Kermer-U. Reviglio-M. Trevisan-S. Verza, *The European Media Freedom Act: media freedom, freedom of expression and pluralism*, 2023.

[74] A similar clause had been discussed as an obligatory "media exemption" or "non-interference principle" in the phase of drafting the DSA, encompassing terms and conditions and notice-and-action procedures. But at that time, no political agreement was reached on this issue; however, a debate started which fed into the debate over art. 18 EMFA. See C. Papaevangelou, *'The non-interference principle': Debating online platforms' treatment of editorial content in the European Union's Digital Services Act*, in *European Journal of Communication*, 38(5), 2023, 466 ss. On the media privilege in EMFA, see D. Tambini, The EU is taking practical measures to protect media freedom. Now we need theory, in *cmpf.eui.eu*, 9 May 2023; M. Z. van Drunen-C. Papaevangelou-D. Buijs- R. Ó. Fathaigh, *What can a media privilege look like? Unpacking three versions in the EMFA,* in *Journal of Media Law*, 15 (2), 2023, 152 ss.

offer protection against the unjustified removal by VLOPs, in case the media content was produced in line with professional standards[75]. To minimize the impact of any restriction to that content on users' right to receive and impart information, and to preserve media outlets and journalists from unjustified content removals or suspensions, VLOPs should submit their statement of reasons to the MSPs prior to the suspension or restriction of visibility taking effect, and MSP's complaints to platforms should be handled with priority «to minimize the impact of any restriction to that content on users' right to receive and impart information» (Recital 50). Thus, art. 18 EMFA aims to contribute to protect media pluralism by establishing a privileged procedure for content moderation of news content produced by MSPs over other types of content, ultimately leading to a prominent and diverse provision of news content offered by independent sources, that are already subject to other regulations, either hard media law or self-regulation aimed at guaranteeing the quality of their content.

The DSA also introduces specific regulations for recommender systems. It provides the first detailed legal definition of recommender systems worldwide[76] (art. 3 (s)) and requires platforms to (1) notify its users when these systems are being used, (2) disclose the employed algorithmic parameters in plain and intelligible language, (3) allow users to manually alter the criteria used for content recommendations (art. 27), and (4) allow users to opt-out from recommender systems based on profiling (art. 38). Furthermore, under the DSA independent audits of algorithmic systems are mandated. According to art. 37 platforms must undergo, at their own expense and at least once a year, independent audits aimed at assessing compliance with the regulation. These will complement and interact with the assessment and mitigation of systemic risks, which includes risks to media pluralism, in accordance with the delegated regulation on the performance of audits (DRPA) (2024/436) (arts. 13 and 14).[77] To implement and monitor the DSA provisions, the European Centre for Algorithmic Transparency (ECAT) has also been established, which is committed to research algorithmic systems for policy purposes, through platform assessment and investigations, scientific projects, networking, and community building. This legal framework complements the GDPR which guarantees European consumers a set of individual rights in relation to the collection of data useful for user profiling, thus relevant in relation to recommender systems on social media platforms. Importantly, the GDPR regulates more in detail profiling and, even if it is not explicitly mentioned, the "right to an explanation"[78].

Social media's recommender systems are also regulated under the Artificial Intelligence Act (AIA). At present, however, they are classified only as "minimal risk" AI systems,

[75] Media service providers should self-declare to VLOPs ex art. 18 (1), stating among other things, to abide by standards of editorial independence and not to provide AI generated content without editorial review.

[76] Except, to our knowledge, a more general definition in China's regulation. U. Reviglio-G. Santoni, *Governing Platform Recommender Systems in Europe: Insights from China*, in *Global Jurist*, 23(2), 151 ss.

[77] Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines.

[78] L. Edwards-M.Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, in *Duke Law & Technology Review,* 16 (18), 2017, 18 ss.

which are systems free to use for which the only suggested regulatory action is the promotion of voluntary codes of conduct. AIA's framework, in fact, defines four levels of risk in AI: unacceptable, high, limited and minimal or no risk, carrying with them different obligations. Initially, recommender systems were deemed as high risk by the EU parliament[79]. This would have led to a whole series of additional obligations for platforms that would have increased the safety, transparency and accountability of these systems, most importantly risk management (art. 9) and technical requirements (art. 15). However, the EC is empowered to adopt delegated acts to amend the list of AI systems categorized under each risk typology (art. 73). The possibility therefore remains that such systems in the future may be subject to more stringent provisions[80]. There are, however, already applicable norms of the AIA. First, according to art. 5, AI models using "subliminal techniques" beyond a person's consciousness or that are intentionally manipulative or designed to exploit a person's vulnerability in a manner that causes or likely to cause physical or psychological harm are to be banned. This would prevent recommender systems from manipulating users. In parallel, the DSA addresses the issue of "dark patterns" (art. 25), which states that social media's interface should not be designed in a way that hinders users' ability to make informed decisions[81]. Additionally, the emerging framework established by the DSA for data access for research (art. 40), which also operates in tandem with the GDPR rules on research processing, is another fundamental tool to improve the understanding of social media functioning, and how this affects media pluralism, eventually providing evidence for policymaking.

Furthermore, a particularly sensitive area of content moderation is undoubtedly disinformation. In this governance area, the central instrument of the EU is the Code of Practice on Online Disinformation (CoP), a self-regulatory instrument set up in 2018 by the leading online platforms and the advertising industry established and evaluated through a process guided by the European Commission (EC). While acknowledged as a significant first step, the CoP in its initial execution contained some critical flaws and limitations.[82] Because of the noted shortcomings, the EC modified this self-regulatory instrument in a "strengthened" version (2022) and is expected to convert it into a code of conduct under the DSA. By signing the CoP platforms such as Facebook and TikTok have also committed to making changes to their algorithms based on so-called "trust indicators" that would reduce the risk of users being misled by ambiguous content. Signatory platforms should amplify "authoritative information", allowing users to introduce trust signals into recommendation systems, and providing metrics to evalu-

[79] See *Commission welcomes political agreement on Artificial Intelligence Act,* press release, 9 December 2023.

[80] N. Helberger, *FutureNewsCorp, or how the AI Act changed the future of news*, in *Computer Law & Security Review*, 52, 2024, 105915.

[81] Recital 67 DSA refers to dark patterns and lists a number of unfair practices that platforms often engage in.

[82] E. Culloty-K. Park-T. Feenane-C. Papaevangelou-A. Conroy-J. Suiter, *Covidcheck: assessing the implementation of EU code of practice on disinformation in relation to Covid-19*, Project Report. Broadcasting Authority of Ireland and FuJo, 2021; I. Nenadic- E. Brogi- K. Bleyer-Simon, *Structural indicators to assess effectiveness of the EU's Code of Practice on Disinformation*, European University Institute, Working Paper, 2023.

ate the effectiveness of fact-checking, and information on user engagement with the options provided to modify the output of algorithms. Trust indicators are expected to provide the basis for platforms to improve the discoverability of trustworthy content sources and decrease the visibility (demotion) of their untrustworthy counterparts[83].
Political advertising has already become a key component of platform regulation in the context of the initial Code of Practice on Disinformation, as well as its later iteration. As content aimed at one's friends and followers can easily be turned into advertising, by boosting or amplifying a post through payments, it becomes hard to determine the differences between organic content or professional communications and political advertising. As previously said, it was especially the scandal around Cambridge Analytica that made it clear how risky political advertising can be, and how broad the concept of political advertising can become, depending on its definition and on the context.[84]
While the CoP still depended on the voluntary cooperation of online platforms, the EU regulation on the transparency and targeting of political advertising, which will entry into force in 2025, has introduced mandatory rules. These cover the labeling and targeting of campaign messages, as well as give authorities the opportunity to impose sanctions on those who violate the rules, including the technology companies that provide advertising services. As such, users of online services are expected to be exposed to less deceptive and manipulative advertisement and will be given tools to understand why certain messages target them.

Regulatory areas	**Main regulatory instruments**	**Main provisions**
Liability	E-Commerce Directive (2000/31/EC) DSA (2022/2065) AVMSD (2018/1808) Directive on Combating Terrorism (2017/541) Directive on Copyright in the Digital Single Market (2019/790)	Art. 14 E-Commerce Directive Arts. 4-10 DSA (Chapter 2 – Liability of providers of intermediary services)

[83] In the CoP, "trustworthiness" refers to the source or publisher of information. An information publisher can be trusted when the chance that users will be exposed to false or misleading content from that source is relatively low. Furthermore, a reputable publisher is expected to have a process in place to make sufficient and timely corrections, in case it publishes false or misleading content.

[84] In its narrowest sense, political advertising refers to advertising placed by political parties and candidates running for office, with the aim of securing votes. However, Cambridge Analytica's advertisements were not directly placed by a party or candidate, and were not always advocating for a vote – sometimes, they just wanted to convince certain voters to stay at home, and not cast their votes. See C. Timberg- I. Stanley-Becker, *Cambridge Analytica database identified Black voters as ripe for 'deterrence,' British broadcaster says*, in *washingtonpost.com*, 29 September 2020.

Content moderation	DSA (2022/2065) EMFA (2024/1083) CoP (2022) Political Ads (2024/900)	Art. 16 DSA (Notice and action mechanisms) Art. 17 DSA (Statement of reasons) Art. 22 DSA (Trusted flaggers) Art. 18 EMFA (Content of media service providers on very large online platforms)
Content curation	DSA 2022/2065 DRPA (2024/436) EMFA (2024/1083) AVMSD (2018/1808)	Art. 27 DSA (Recommender system transparency) Art. 38 DSA (Recommender systems) Art. 3 EMFA (Right of recipients of media services) Commitments 18-22 CoP Art. 7a AVMSD (N.B. enforced by member states and only for video-sharing platforms, i.e., YouTube) Art. 13 AVMSD (N.B. enforced by member states)
Transparency & accountability	DSA (2022/2065) GDPR (2016/679) DRPA (2024/436) CoP (2022) EMFA (2024/1083)	Art. 15 DSA (Transparency reporting obligations for intermediary services) Art. 24 DSA (Transparency reporting obligations for online platforms) Art. 37 DSA (Independent Audits) Art. 40 DSA (Data access and scrutiny) Art. 42 DSA (Transparency reporting obligations) Art. 18 EMFA, para. 2 and 8 (Content of media service providers on very large online platforms)
Privacy and manipulation	GDPR (2016/679) AIA (2024/1689) Data Governance Act (2022/868) Data Act (2023/2854)	Art. 22 GDPR (Automated individual decision-making, including profiling) Art. 25 DSA (Online interface design and organisation) Art. 5 AIA (Prohibited AI Practices)
Risk management	DSA (2022/2065) DRPA (2024/436) AIA (2024/1689)	Art. 34 DSA (Risk assessment) Art. 35 DSA (Mitigation of risks) Art. 36 DSA (Crisis response mechanism) Art. 9 AIA (Risk management system)

Table 1. An overview of the mentioned provisions affecting social media's opinion power and exposure diversity.

The 2022 Strengthened CoP (just like its 2018 precedent) includes commitments for signatories that require them to propose appropriate definitions for both political and so called "issue advertising" – the latter being paid content not clearly advocating for the support of the candidate, but still capable of influencing electoral decisions –, as well as asking for increased transparency on many aspects.[85] The Regulation (EU) 2024/900 on the transparency and targeting of political advertising is even clearer, and provides a number of binding transparency and integrity obligations, highlighting the need for a comprehensive assessment of what constitutes political advertising.

Finally, addressing the concentration of opinion power in a few platforms may also require actions related to competition law. However, traditional media concentration law seems to be largely ineffective in this domain.[86] In this sense, the Digital Markets Act (DMA) should be considered as the EU regulatory tool targeting dominant platforms, imposing stricter obligations to counterbalance market concentration. The DMA designates platforms providing core services as "gatekeepers" if these have a "significant impact on the internal market" (art. 3)[87] One of the DMA's novel goals is to achieve a contestable platform market for effective market pluralism, leveling the playing field and lowering entry barriers in the platform market[88]. Provisions such as the prohibition of self-preferencing (i.e., favor one's content) and the interoperability of gatekeeper could also have spillover effects on the exposure to media diversity. Furthermore, other EU regulations such as the Data Governance Act 2022/868 and the Data Act 2023/2854 which emphasize fairer access to and sharing of data could be effective in countering platforms' dominant data power and, indirectly, opinion power. From the perspective of external media pluralism, it will be interesting to observe to what extent the DMA and other regulations aimed at restoring competition will foster

[85] This is fundamental when it comes to social media's opinion power: if there is no regulation of political advertising on social media in the EU member states (as regulation in EU member states still focuses on traditional legacy media) or the definition of political advertising allows for loopholes (making it possible for manipulative ads to go under the radar), certain members of the online audience will find themselves more exposed to manipulation on issues that are most important for opinion formation related to elections – not even knowing about it.

[86] T. Seipp et al., *Dealing with opinion power in the platform world*, cit.

[87] A platform service is presumed to be a gatekeeper if the company, in each of three consecutive financial years, achieved an annual turnover within the EU of at least EUR 7,5 billion, or if the company's average market capitalisation or its equivalent fair market value amounted to at least EUR 75 billion in the last financial year, or if it had on average at least 45 million monthly active end users established or located in the Union in the last financial year, and at least 10 000 yearly active business users established or located in the Union (art. 3 DMA). As of now, four social media platforms have been designated as core platform services of gatekeepers, namely TikTok, Facebook, Instagram, and LinkedIn. Here can be found the full list of designated gatekeepers.

[88] Two concrete benefits for pluralism could be expected from (i) requirements related to fairness and transparency in crawling, indexing, and ranking, namely requiring non-discrimination in content organization; (ii) privacy protection obligations, strengthening consent requirements from users for targeted content. Thus, the DMA directly addresses the core platform service of data trade, whereas the DSA tries to regulate the user's experience that is provided in exchange of such data. See J. Bayer, *Digital Media Regulation within the European Union. A Framework for a New Media Order*, Baden Baden, 2024, 265. According to Bostoen, «the DMA's contestability goal is reminiscent of the pluralism pursued by the Audiovisual Media Services Directive», see F. Bostoen, *Understanding the Digital Markets Act*, in *The Antitrust Bulletin*, 68(2), 2023, 263 ss.

diversification in the news industry, especially considering how newsrooms are dependent on platforms' logics, so these may have spillover effects.

5. Limits, challenges, and opportunities of the EU governance emerging model

While several mechanisms of transparency have been established by the DSA, one of the main concerns is that these will remain a form of "transparency theater", namely that these measures serve more to legitimize online platforms than to exert pressure on their power structure, considering that they are mostly framed as publicity, procedural fairness, and access to datasets, but they do not necessarily solve the problem of information asymmetry[89] nor the limited resources dedicated to content moderation.[90] While transparency disclosures could be appealing for various subjects such as researchers, lawyers and policymakers, they provide relatively little useful information, especially to users.[91] More broadly, it can be questioned if the emerging EU model enables a governance of content moderation which is iterative and dynamic, forcing social media to truly engage in a dialogue about the value judgements behind their choices, and explain each of them[92]. There are, for example, concerns regarding the implementation's effectiveness of provisions related to content visibility. If art. 17 DSA were applied rigorously and all demotions duly disclosed, it would still not be effective enough to avoid shadowbans. Indeed, what makes demotions particularly problematic is that they can be technically undetectable (for example, social media can minimize the risk of detection by demoting content gradually over time rather than instantaneously). Amplified content, on the other hand, must not even be disclosed in the DSA, albeit such strategy (sometimes referred as "shadow-promotion") has been repeatedly observed, even during Russia's war of aggression in Ukraine[93]. VLOPs' operations in extra-EU countries also remain outside the scope of the legislation, albeit these may still affect European citizens' public opinion. These represent critical limitations of EU regulation. It can even be questioned whether to make VLOPs truly accountable, the amplification of content should be disclosed, beyond the disclosure of general criteria of recommender systems (art. 27 and recital 70 DSA) and despite

[89] See for example: M. Maroni, *'Mediated transparency': The Digital Services Act and the legitimisation of platform power*, in *Legal Studies Research Paper Series*, University of Helsinki, 2023.

[90] To illustrate the shortage of human content moderators for languages other than English, the *DSA transparency reports* indicate that in Italy, Alphabet has 229 moderators for YouTube, Meta has 164 for Facebook and Instagram, TikTok employs 439 moderators, and Twitter has just one moderator for the Italian language. All things considered, this amounts to fewer than 850 individuals who understand the Italian language moderating content for approximately 140 million accounts. This indicates that in Italy there is roughly one moderator for every 168,000 accounts.

[91] A. Trujillo-T. Fagni-S. Cresci, *The DSA Transparency Database: Auditing self-reported moderation actions by social media*. arXiv preprint arXiv:2312.10269, 2023.

[92] E. Douek, *Content moderation as systems thinking*, cit., 533.

[93] S. Romano-N. Kerby-M. Schüler-D. Beraldo-I. Rama, *The Impact of TikTok Policies on Information Flows during Times of War: Evidence of 'Splinternet' and 'Shadow-Promotion' in Russia*. AoIR Selected Papers of Internet Research, 2023.

content amplification is hard to assess[94].

Limitations and challenges of the current regulatory framework have been already observed in the recent Israeli-Palestinian conflict. The risk of censorship that shadow-bans entail, for example, when content about the Israeli attacks in the Palestinian territories and posts in support of Palestine have been reportedly demoted by social media.[95] In this case, social media's opinion power remained substantially unchallenged. At the same time, the spread of illegal content following Hamas' terrorist attack in Israel contributed to the first enforcement of the DSA[96]. Criticism was moved by digital rights' civil society organizations towards both the VLOPs involved and the EC for the actions undertaken which may disproportionately affect oppressed groups and human rights defenders[97]. This example shows the difficulties in defining what constitutes violent and harmful content under the DSA. In the occasion of very conflictual and polarizing events/issues, the DSA enforcers' (i.e., public authorities) framing of the situation might boost a certain narrative over another, adding to the issue of opaque content moderation choices by private social media platforms.[98]

Illegitimate influence over public opinion could be unveiled through algorithmic auditing (art. 37 DSA).[99] Their outcome, however, can be limited due to the platform providers' influence in the market.[100] Moreover, due to information asymmetries and superior technological capabilities[101], VLOPs are distinctly better equipped than national governments to monitor disinformation, foreign interference, and other malicious activities. Their advantage lies not only in accessing extensive private and real-time social media data but also in their technical expertise. This critical role inevitably enhances their influence over the formation of public opinion, and it is still unclear whether the EU governance model will be able to make their power fully accountable — as well as

[94] L. Thorburn-J. Stray-P. Bengani, *What will "amplification" mean in court,* in *techpolicy.press,* 19 May 2022; L. Belli- M. Wisniak, *What's in an Algorithm? Empowering Users Through Nutrition Labels for Social Media Recommender Systems*, in *knightcolumbia.org,* 22 August 2023.

[95] K. Paul, *Instagram users accuse platform of censoring posts supporting Palestine, in theguardian.com*, 18 October 2023. Two years earlier, Meta was accused of censoring posts in support of Palestine on its platforms in occasion of the Sheikh Jarrah crisis in 2021. Facebook's Oversight Board entrusted a company to conduct a due diligence exercise that led to recommendations for Meta. See H. Elmimouni et al., *Shielding or Silencing?: An Investigation into Content Moderation during the Sheikh Jarrah Crisis,* Proceedings of the ACM on Human-Computer Interaction, Volume 8, Issue GROUP, Article No.: 6, 1 ss.

[96] The EC submitted formal requests of information to X, Youtube, TikTok and Meta on risk assessment and mitigation measures against illegal content and disinformation (based on art. 67 DSA); later, it adopted a Communication against hate speech which established a dedicated network of trusted flaggers specialized in antisemitic content online; finally, formal proceedings (based on art. 66 DSA) were opened also against X and TikTok. full list of the main enforcement activities.

[97] Accessnow, *Precise interpretation of the DSA matters especially when people's lives are at risk in Gaza and Israel*, in *accessnow.org*, 18 October 2023.

[98] K. Bleyer-Simon-U. Reviglio, *Defining Disinformation across EU and VLOPs*, European Digital Media Observatory, forthcoming.

[99] P. Terzis-M. Veale-N. Gaumann, *Law and the Emerging Political Economy of Algorithmic Audits*, in *The 2024 ACM Conference on Fairness, Accountability, and Transparency,* June 2024, 1255 ss.

[100] J. Laux-S. Wachter-B. Mittelstadt, *Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA,* in *Computer law & Security review*, 43, 2021, 105613.

[101] J. Black, *Decentring regulation: Understanding the role of regulation and self-regulation in a 'post-regulatory'world*, in *Current legal problems*, 54(1), 2021, 103 ss.

the emerging powers of the EU institutions.

The role of users in improving content moderation and identifying illegitimate influence over public opinion could also be leveraged. As said above, the out-of-court dispute settlement bodies can be a powerful instrument. However, it has been stressed that there might be issues related to workload, limiting in practice the efficiency and effectiveness of these bodies; however, there is hope that there will be a pull effect: the more users turn to these bodies, the greater the pressure on platforms to comply with the decisions.[102] Other forms of participatory governance should be considered in future policy developments. Most importantly, "social media councils" could be established, which are independent, multi-stakeholder bodies designed to oversee, advise, and sometimes enforce accountability in how social media platforms manage content moderation (similar to X's Community Notes).[103] Users could be further involved, among others, in countering and mitigating impacts of online misinformation[104], or even providing algorithmic feedback; indeed, retrospective, deliberative judgment on previous recommendations could indeed help users to align their preferences with the output of recommender systems.[105] Furthermore, a crowd-sourced approach to identify illegitimate influence such as shadowbans could be envisioned, where users signal their suspicions or evidence, especially during conflicts and high intensity events. In practice, they could already submit these in the DSA whistleblowing platform[106], the decision to disclose this information remains at the Commission's discretion. Developing more transparent and inclusive mechanisms could strengthen the EU's governance approach to social media's opinion power.

Fundamentally, to "redistribute" social media's opinion power, users should be able to consciously decide for themselves what content they want to see in social media. Although art. 27 DSA provides the criteria that recommender systems should be adjustable by users to their preferred preferences, we still don't know what is meant with "criteria" and "preferences" and, in fact, these same criteria are left to VLOPs to choose; moreover, the design of these same options remains at their discretion. Therefore, art. 27 DSA is still unclear in its implementation. Even if a delegated act or a code of conduct would clarify this, what options would be offered to users are clearly hard to

[102] H. Ruschemeier-J. P. Quintais-I. Nenadic-G. De Gregorio-N. Eder, *Brave New World. Out-Of-Court Dispute Settlement Bodies and the Struggle to Adjudicate Platforms in Europe*, in *verfassungsblog.de,* 10 September 2024. According to the authors of this blog post, out-of-court dispute settlements bodies have different options, ranging from a limited mandate that only covers the content moderated and not the justification provided by the platform, to a full review of, for example, all the requirements of art. 17 DSA (statements of reason). Also, ODS bodies are likely ill-suited to carry out assessments related to misinformation. Therefore, it would be advisable for them to cooperate with fact-checking organisations as well as with journalists and news media organizations.

[103] M.C. Kettemann-W. Schulz. *Platform://Democracy: Perspectives on Platform Power, Public Values and the Potential of Social Media Councils*, 2023.

[104] The Global Partnership on AI, *Responsible AI for Social Media Governance: A Proposed Collaborative Method for Studying the Effects of Social Media Recommender Systems on Users,* 2021 Report. The Global Partnership on AI.

[105] J. Stray-A. Halevy-P. Assar-D. Hadfield-Menell-C. Boutilier-A. Ashar-N. Vasan, *Building human values into recommender systems: An interdisciplinary synthesis*, in *ACM Transactions on Recommender Systems*, 2(3), 2024, 1 ss.

[106] See *digital-services-act-whistleblower.integrityline.app*.

establish. The risk of further promoting personal relevance at the expense of exposure to more diverse content needs to be carefully weighed by considering customization options. To further empower users, EMFA could complement the DSA's endeavor, since it has recognized the right to customize the media offer (art. 20). This, however, only applies to audiovisual media such as smart TV interfaces and applications. And yet, art. 3 EMFA asserts the right of recipients (i.e., users) « to have access to a plurality of editorially independent media content and ensure that framework conditions are in place (…) to safeguard that right, to the benefit of free and democratic discourse.». In conjunction with the spirit of art. 20 EMFA, as well as art. 10 of the European Convention of Human Rights (ECHR) recognising the right to freely receive and impart information and Article 11 of the EU Charter of Fundamental Rights (CFR), art. 3 can complement artt. 27 and 38 DSA on recommender systems in its future implementation.[107] Furthermore, the forthcoming list of media service providers that will result from self-declarations to VLOPs may also represent an additional option for users to customize their experience by receiving only, or mostly (i.e., prioritize), content from media following professional standards. While this may empower users, however, it may also further strengthen VLOP's opinion power, as they ultimately select these "professional" media service providers.[108]

With regards to the limitations of art. 18 EMFA, the idea of providing a "privilege" for media content is aimed at balancing the asymmetrical relation of power between media service providers, especially smaller ones, and VLOPs.[109] Critics, however, point out the ample discretion left to platforms when accepting the MSPs' self- declarations[110]. As a matter of fact, platforms can- but are not bound- to consult public authorities when deciding if to accept or to reject the status of MSPs declared by a user of their services. Moreover, art. 18 might divert attention from the content of communications towards the standing of the source – as even renowned news media outlets can publish untrue content – and possibly provides a loophole that can be exploited by content publishers that only formally comply with quality standards. The adherence to professional standards should indeed be the basic requirement for the identification of the media service providers that will enjoy the guarantees granted by the EMFA Regulation, also in light of the standards elaborated by Recommendation 2022/1634 "on internal safeguards for editorial independence and ownership transparency", accompanying EMFA[111]. To some extent, this is explicitly specified by art. 18 (1d) EMFA

[107] U. Reviglio-M. Fabbri, *Navigating the Digital Services Act: Scenarios of Transparency and User Control in VLOPSEs' Recommender Systems,* in *Proceedings of the 18th ACM Conference on Recommender Systems.*

[108] How EMFA's provisions will be implemented in the single EU Member States, however, remains to be seen, considering that EMFA is a "Regulation of principles"; namely, it is directly enforceable in the national legal frameworks, as EU Regulations are, but is very open in its possible interpretations, as if it was a Directive. See E. Brogi et al., The European Media Freedom Act, cit.

[109] O. Pollicino-F. Paolucci, *Unveiling the Digital side of Journalism: Exploring the European Media Freedom Act's opportunities and challenges*, in *La Revue des Juristes de Sciences Po*, 1, 2024.

[110] D. Tambini, *The EU is taking practical measures to protect media freedom. Now we need theory*, CMPF Discussion Series, 9 May 2023; M. Z. van Drunen-C. Papaevangelou-D. Buijs-R. Ó. Fathaigh, *What can a media privilege look like? Unpacking three versions in the EMFA*, in *Journal of Media Law*, 15(2), 2023.

[111] S. Verza, *What is journalism in the digital age? Key definitions in the European Media Freedom Act,* forthcoming in *Rivista Italiana di Informatica e Diritto.*

as one of the criteria for media service providers to be able to self-declare as media. A related debate in this context is the prominence of Public Service Media and public-interest content online. This is primarily governed by the Audiovisual Media Services Directive (AVMSD) but it does not apply in the online environment.[112] For example, Article 7a of the AVMSD requires EU Member States to ensure the appropriate prominence of audiovisual media services of general interest, but its enforcement is limited to video-sharing platforms (e.g., YouTube). The EMFA and the DSA lack effective mechanisms to guarantee that PSM and public value content are given due prominence online. While the DSA primarily focuses on mitigating harmful and does not place positive obligations on platforms to prioritize public interest content, the EMFA fails to address the need for prioritizing public value content in digital spaces. However, the implementation of the DSA and the revision of the EMFA offer an opportunity to close these accountability gaps within this governance framework.[113]

In the realm of social media's potential manipulative power, the main provisions highlighted above - the prohibition of dark patterns under art. 25 DSA and of subliminal techniques under art. 5 AIA - present significant challenges. On the one hand, dark patterns and nudging techniques constantly change, so there are concerns that it may be hard to identify and promptly ban new ones.[114] On the other hand, AI harms remain hard to detect, and the same harms they cause would be hard to prove.[115] Given the potential risks highlighted in the first paragraph of this article, the fact that the AIA does not consider recommender systems to be a high risk AI technology represents a substantial limitation of the EU approach to tame and redistribute social media's opinion power. To effectively prevent manipulation, more experimental research is needed. This is particularly true for the experimental protocols of VLOPs.[116] At present, however, the data access for research granted by the DSA shows several limitations and, above all, does not allow for experimental tests in social media. Article 40 DSA can represent a watershed in the understanding of how platforms can set the news agenda and, eventually, for collecting evidence for policymaking. However, data access is limited to "vetted researchers", which means researchers affiliated with universities and independent from commercial interests, while excluding traditional watchdogs such as journalists and NGOs. Moreover, the norm protects platform data, user privacy, and trade secrets and further restrictions delimit the grounds that justify data access (most importantly, the "systemic risks""), what data can be accessed, and more generally how these are managed afterwards.

Finally, in the realm of advertising, there are also possible challenges related to paid

---

[112] E. M. Mazzoli, *The politics of content prioritisation online governing prominence and discoverability on digital media platforms* (Doctoral dissertation, London School of Economics and Political Science), 2023.

[113] K. Rozgonyi, *Accountability and platforms' governance: the case of online prominence of public service media content,* in *Internet Policy Review*, 2023, 12(4).

[114] P. Cesarini, *Regulating Big Tech to Counter Online Disinformation: Avoiding Pitfalls while Moving Forward*, in *medialaws*.eu, 2021.

[115] D. Acemoglu, *Harms of AI*, National Bureau of Economic Research, Working Paper 29247, September 2021.

[116] D. Knott-J. Pedreschi- S. Stray-S. Russell, *The EU's Digital Services Act must provide researchers access to VLOPs' experimental protocols*, in *informationdemocracy.org*, June 2024.

and amplified content with the potential of manipulating opinions. The discussion on political and issue advertising has already shown that the categories are not clear-cut, and a lot of new ways of manipulating political opinion can become possible. In traditional journalism, the boundaries between editorial and commercial content are being blurred by native advertising and sponsoring of content, among other things. The targeting and microtargeting of advertising are by now subject to increased scrutiny, but new forms of personalizing advertising messages might emerge in the future. The evolution of sponsored formats with the potential of influencing political processes will therefore require policymakers and other stakeholders to consider political advertising a moving target,and constantly update its definitions to prevent misuses and undue manipulation of opinions.

6. Final remarks

Despite moderately positive expectations of the EU policies' potential in the areas of social media governance and media pluralism, more pessimistic viewpoints have also been raised. Notably, influential EU leaders have suggested the possibility of using the DSA during periods of civil unrest to shut down social media platforms completely.[117] How far this model might go remains to be seen. Given the nascent stage of the DSA's enforcement, and many of the other relevant regulations entering into full force, it would be premature to assess the overall effectiveness as well as the potential unintended consequences of this emerging governance model in taming and redistributing social media's opinion power.[118] Furthermore, despite the EU's regulatory efforts and its rather ambitious, comprehensive approach to digital governance, the influence of U.S. technological and political hegemony remains a subtle yet pervasive factor shaping these strategies. Indeed, almost all social media platforms are US-based and potentially

[117] C. Goujard-N. Camut, *Social media riot shutdowns possible under EU content law, top official says*, in politico.eu, 10 July 2023; L. Kayali-El. Bertholomey, *Macron floats social media cuts during riots*, in ***politico.eu,*** 5 July 2023.

[118] Key elements of the framework are still being rolled out, starting with the first algorithmic audit set for August 2024, which will serve as an initial litmus test for the DSA's ability to open social media's black boxes. Furthermore, the Digital Services Coordinators (DSCs), namely the national authorities critical to the enforcement and oversight of the DSA, are not expected to be fully operational until 2025, as well as the European Board for Digital Services, composed of the DSCs. Similarly, EMFA establishes the European Board for Media Services, gathering representatives of national media authorities (substituting ERGA – the European Regulators Group for Audiovisual Media Services). Due to the complex interplay between the DSA and EMFA regarding fundamental rights and the platforms' operationalisation of the systemic risks, competent authorities will play a crucial role in the process of enforcement. Their role, along with the appointment of trusted flaggers who will assist in identifying and addressing compliance issues, marks a significant step in operationalizing the Act. (See I. Nenadic- E. Brogi, *The Game of Boards: The role of authorities in concerting the Digital Services Act and the Media Freedom Act for protecting media freedom,* in ***medialaws.eu,*** 28 August 2024.) Additionally, the transition of the Code of Practice on Disinformation into a formal Code of Conduct in January 2025 will introduce enforceable obligations specifically tailored to combat disinformation. This shift, coupled with art. 40 of the DSA allowing access to data for researchers becoming fully operable by the end of the year, underscores the gradual implementation process.

aligned with and weaponized for U.S. interests[119], and it is unlikely to expect a change in this status quo with the creation of a "EU silicon valley", considering the hurdles posed by regulatory differences and bureaucratic complexities across EU member states or even the talent and infrastructural gaps.[120] This dominance, coupled with the rapidly evolving media landscape and the lack of robust evidence to guide effective policymaking, makes regulating social media's opinion power particularly challenging. Additionally, social media's considerable political and technical power may enable them to effectively lobby against and often circumvent regulatory measures. Even if users rely on their service for essential functions such as business, social interaction and, indeed, access to information, these continue to exercise a form of privatized governance through the terms and conditions they impose on users, creating a contractual relationship marked by structural information asymmetries where social media have significantly more information than users. These asymmetries, along with the platforms' business models that prioritize user engagement—a strategy that seem to contribute to most of the unintended and harmful consequences of social media[121]—remain largely unaddressed by current regulations. Most of these Regulations, in fact, find their legal basis in art. 114 of the Treaty on the Functioning of the European Union (TFUE), namely in the objective of harmonizing the internal market.[122] This market-based rationale is arguably not the perfect fit for comprehensively addressing

[119] In the realms of national security (e.g., terrorist propaganda, immigration security or foreign influence operations) and public health (e.g., COVID-19), there has been a natural collaboration between U.S. social media companies and the U.S. government. While direct evidence is lacking, the history of secret surveillance initiatives like the PRISM project provides a basis for legitimate speculation that such partnerships might also extend to manipulating other content, particularly in foreign countries and for supporting U.S. interests. This conjecture is further bolstered by the U.S. government's well-documented history of conducting "psychological operations" (PSYOPS), which encompass a set of techniques used by military and non-military organizations to manipulate public perceptions through the deliberate use of information, misinformation, and communication strategies, ultimately influencing decision-making processes and behaviors. Various investigations indeed highlighted shady U.S. army operations conducted by thousands of people secretly employed (see N. Fielding - I. Cobain, *Revealed: US spy operation that manipulates social media*, in *The Guardian*, March 17 2011; William M. Arkin, *Exclusive: Inside the Military's Secret Undercover Army*, in *Newsweek*, May 17 2021). Further suspicions of collusion between the U.S. government and social media also emerged from the 'Twitter files', which revealed a close relationship with the FBI, as well as from the fact that various former CIA agents are working, or have worked, at Meta for issues related to content moderation (see M. Koening, *Spooks infiltrate Silicon Valley,* in *Dailymail*, December 22 2022).

[120] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Report on the state of the Digital Decade 2023*.

[121] P. Bengani-J. Stray-L.Thorburn, *What's right and what's wrong with optimizing for engagement. Understanding Recommenders,* 27 April 2022.

[122] See P. Parcu et al., *Study on media plurality and diversity online,* cit., 36 ss.; S Vries- O. Kanevskaia-R.de Jager, *Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation*, in *European Papers,* 8(2), 2023, 590. It is fundamental to acknowledge that the EU is essentially an economic regulator. As van Drunen et al. argued «the more the EU enacts rules that affect public communication, the more urgent it becomes to integrate the relevant sets of expertise into EU decision making, strengthen the procedures that anticipate broader impact on the marketplace of ideas, as well as re-think more generally the legitimacy the European Union has for adopting speech-related measures under the legal bases to regulate the internal market and protection of personal data». See M. van Drunen-N. Helberger-W. Schulz-C. de Vreese*, The EU is going too far with political advertising!*, in *dsa-observatory.eu,* 16 March, 2023.

complex phenomena impacting on human rights and societal dynamics caused by the platformization of the public sphere. In fact, it legally recognizes the strong ramifications of power of private actors for citizens' fundamental rights, public interests and social values, breaking down the traditional public-private divide.

To counteract the roots of social media's harms and effectively tame social media's opinion power, more structural interventions could be deployed such as reforming the ad-driven business model, crowd-sourcing content moderation, social media full interoperability, mandating the disclosure of platforms' experiments and non-engagement signals in recommender systems[123], that can be designed to favor societal cohesion[124] or other democratic values[125], but even developing a public service social media or creating a market of recommender systems and let users choose which one to employ[126]. Without a more ambitious interpretation of these EU regulations, as well as the implementation of more structural and radical policies, the EU's governance of social media platforms is unlikely to mitigate the risks stemming from their opinion power—which, as highlighted throughout this paper, not only determines "personalized agenda-settings" but it is closely and concerningly tied to their ability to shape users' worldviews, to manipulate their information behavior and to manage the challenges of contemporary information warfare.

7. Conclusions

This paper has delineated and explored the key elements of the individual and public opinion-shaping power of social media and their interaction with the past and emergent regulatory landscape of the EU. Over time, the EU governance model has become more stringent and dynamic: the terms of service and community guidelines across social media have increasingly been scrutinized and standardized; content moderation practices have evolved from reactive to more proactive measures; recommender systems have been subject to regular transparency provisions and algorithmic audits; and lastly, platforms' interface design is being rethought to prioritize user autonomy, compliance, and the mitigation of systemic risks. The emerging EU regulatory model, however, has also been widely criticized. Above all, because it does not directly "fix" the business model of social media and its inherent objective to maximize for "user engagement" - with all the undesirable consequences this seems to lead to – but it tackles platforms' opinion power mainly by creating indirect incentives, such as mandating transparency measures and the assessment and mitigation of "systemic risks", in addi-

[123] T. Cunningham-S. Pandey- L.Sigerson-J. Stray – J. Allen - B Barrilleaux - B. Rezaei, *What We Know About Using Non-Engagement Signals in Content Ranking,* in *arXiv,* 2024.

[124] A. Ovadya-L. Thorburn, *Bridging systems: open problems for countering destructive divisiveness across ranking, recommenders, and governance,* in *arXiv*, 2023.

[125] J. Stray-A. Halevy-P. Assar-D. Hadfield-Menell-C. Boutilier-A. Ashar -N. Vasan, *Building human values into recommender systems: An interdisciplinary synthesis,* in *ACM Transactions on Recommender Systems*, 2(3), 2024, 1-57.

[126] J. M. Marella, *Middleware Technologies: Towards User-Determined News Curation in Social Media,* in *Cath,* in *UJL & Tech*, 31, 2022, 95.

tion to fines. While this model may be largely beneficial to the "public sphere", in the discussion chapter we briefly highlighted many of the legal and technical challenges and opportunities it faces.

All in all, the EU governance model's effectiveness in governing social media's "opinion power" remains to be seen and it would be premature to draw definitive conclusions. The rationale behind regulations such as the DSA, EMFA, and partly the AIA, nevertheless, must be recognised as an explicit will from the side of EU institutions to rebalance very large private platforms' powers over public opinion, regaining control over public interest objectives and overcoming the traditional presumed neutrality of online platforms regarding content moderation. And yet, many of the limitations we discussed possibly derive from the market-oriented logic of the recent EU media regulations, as articulated in their legal basis -Article 114 TFUE - which may have constrained opportunities for a more profound reconsideration of the dynamics of social media's opinion power.

Our retrospective analysis suggests that despite the gradual and reactive evolution of the EU regulatory model, public pressure has eventually led to more stringent regulation and public oversight. While the effectiveness of transparency measures and user empowerment provisions can still be contested at present, the significant volume of information that will be eventually disclosed can nonetheless be expected to sustain, and possibly enhance, the regulatory impact of this governance model. To comprehensively assess such impact on opinion power in general, and media pluralism in particular, it will be crucial to expand the analysis beyond a user-centered perspective on diversity exposure to understand the overall media ecosystem, including newsrooms, as well as the impacts on external pluralism of new EU laws that act also on competition issues.

Online hate speech, diritto penale e libertà di espressione. Utopia od opportunità?*

Matilde Bellingeri, Federica Delaini

Abstract

L'avvento delle nuove tecnologie nella governance del discorso pubblico impone di svolgere una riflessione circa l'opportunità di utilizzare il diritto penale quale strumento volto a contrastare (e prevenire) le condotte di online hate speech, le quali hanno assunto potenzialità lesive senza precedenti, portando con sé seri rischi per la sopravvivenza di una società democratica e plurale. A fronte della dimensione transnazionale del fenomeno, si procederà tracciando una mappatura dello stato dell'arte in Italia, mediante il ricorso ad una metodologia comparatistica che abbracci le soluzioni applicate in ambito europeo ed americano. Si rifletterà sull'efficacia delle scelte di politica criminale, valutando, quali rimedi alternativi, gli strumenti di "moderazione algoritmica" in capo agli Internet Service Providers. In conclusione, si indagheranno le interferenze tra i limiti apposti alla libertà di pensiero ed i principi che sorreggono un ecosistema penale costituzionalmente orientato, nel prisma della tutela dei diritti e delle libertà fondamentali degli utenti.

The advent of new technologies in the governance of public speech compels a reflection on the possibility of using criminal law as a means of counteracting (and preventing) online hate speech, which has taken on an unprecedented detrimental potential, leading to serious risks for the survival of a democratic and pluralistic society. In view of the transnational dimension of the phenomenon, it is intended to map the state of the art in Italy, through a comparative methodology that encompasses the solutions applied in the European and American frameworks. The effectiveness of criminal policy measures will be reflected upon, evaluating, by way of alternative remedies, the tools of algorithmic moderation in the hands of the Internet Service Providers. In conclusion, the interactions between the limits posed on freedom of thought and the principles that underpin a constitutional criminal ecosystem will be investigated through the prism of the protection of users' fundamental rights and freedoms.

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco". Il seguente articolo è frutto di riflessioni condivise. Tuttavia, i paragrafi 2, 4 e 7 sono attribuibili a Federica Delaini, mentre i paragrafi 3, 5 e 6 a Matilde Bellingeri. L'introduzione e le conclusioni (paragrafi 1 e 8) sono stati scritti congiuntamente dalle Autrici.

Contributo finanziato dall'Unione europea - Next Generation EU, Missione 4, Componente 1, CUP B31I23000810004

Sommario

Keywords

diritto penale – discorsi d'odio online – comparazione – libertà di espressione – democrazia plurale

"Il connubio di odio e di tecnologia è il massimo pericolo che sovrasti l'umanità. E non mi riferisco alla sola grande tecnologia della bomba atomica, mi riferisco anche alla piccola tecnologia della vita di ogni giorno: conosco persone che stanno per ore davanti al televisore perché hanno disimparato a comunicare tra di loro"

Simon Wiesenthal

1. Introduzione: il contrasto all'*online hate speech*

La comunicazione al tempo di Internet ha fortemente ridisegnato le coordinate del discorso pubblico: sulle piattaforme digitali ogni individuo diffonde il proprio pensiero a priori dall'appartenenza ad una classe ovvero etnia, in piena rispondenza al paradigma strettamente democratico che permea l'ecosistema digitale. Tuttavia, la fondamentale libertà di manifestazione del pensiero e d'informazione e la tutela del pluralismo e del diritto alla non discriminazione trovano massima frizione nel contesto delle piattaforme online, in quanto luoghi forieri di nuovi e gravi pregiudizi, sia individuali che collettivi[1].

Per detta ragione, si è ritenuto di cruciale importanza esaminare, quale asse portante del contributo, la *vexata quaestio* dell'opportunità di incriminare i discorsi d'odio online, che costituiscono una *species* del più ampio *genus* degli *hate crimes*[2], dai quali ereditano le riflessioni in tema di oggettività giuridica tutelata, oltre che i punti di contatto e di

[1] P. Stanzione, *Introduzione*, in P. Stanzione (a cura di), *I "poteri privati" delle piattaforme*, Torino, 2022, 9.

[2] L'Organizzazione per la Sicurezza e la Cooperazione in Europa (OCSE) ha definito i crimini d'odio come «*violent manifestations of intolerance and have a deep impact on not only the immediate victim but the group with which that victim identifies him or herself. They affect community cohesion and social stability. A vigorous response is therefore, important both for individual and communal security*», così OCSE, in *Hate Crime Laws. A Practical Guide*, Odihr, Varsavia, 2009, 11.

tensione con le fondamentali libertà sopra menzionate[3].

Nonostante nel panorama internazionale non esista una definizione unitaria di discorso d'odio, gli Stati che si sono interessati al tema ne hanno individuato alcuni tratti comuni: l'incitamento all'odio, la volontà che l'odio non rimanga in uno stato di potenza, bensì venga posto in azione nei confronti del gruppo che ne è destinatario e il fatto che per quest'ultimo sussista anche soltanto il rischio imminente che vengano realizzati atti di violenza e di discriminazione[4].

Tali elementi sono presenti, altresì, nella loro massima evoluzione, nella recentissima Raccomandazione CM/Rec (2022) del Comitato dei Ministri del Consiglio d'Europa, ove si afferma che con discorso d'odio «si intende qualsiasi forma di espressione mirante a stimolare, promuovere, diffondere o giustificare la violenza, l'odio o la discriminazione nei confronti di una persona o un gruppo di persone, o a denigrare una persona o un gruppo di persone per motivi legati alle loro caratteristiche o situazioni personali, reali o presunte, quali la "razza", il colore della pelle, la lingua, la religione, la nazionalità o l'origine nazionale o etnica, l'età, la disabilità, il sesso, l'identità di genere e l'orientamento sessuale»[5].

Da tempo la dottrina penalistica si interroga sull'opportunità di sanzionare penalmente la manifestazione di un pensiero "pericoloso", qualora esso contrasti con valori di rango superiore alla libertà di espressione[6]. Quest'ultima, infatti, a una prima lettura, sembra incompatibile con qualsiasi forma di limitazione. Appare, dunque, legittimo chiedersi per quale motivo alcuni ordinamenti contemporanei, tra cui il nostro, abbiano attribuito rilevanza penale ai discorsi d'odio[7].

In quest'ottica, si cercherà di investigare le problematiche insite nell'utilizzo della rete, che a tutt'oggi è in grado di mettere a rischio la sopravvivenza di società eterogenee e multiculturali, data la spiccata attitudine lesiva nei confronti dei valori fondamentali su

---

[3] Per un maggiore approfondimento in ordine ai diversi modelli di incriminazione dei discorsi d'odio che si sono diffusi a livello europeo, distinguibili in modelli istigatori in contrapposizione a modelli "ampi", si rinvia a A. Spena, *La parola(-)odio. Sovraesposizione, criminalizzazione e interpretazione dello hate speech*, in *Criminalia*, 2017, 589-593.

[4] G. Ziccardi, *Il contrasto dell'odio online: possibili rimedi*, in *Lessico di etica pubblica*, 9,1, 2018, 39 – 40; A. Portaru, *Freedom of Expression Online: the Code of Conduct on Countering Illegal Hate Speech*, Revista Romana de Drept European, 4, 2017, 82-83.

[5] Raccomandazione CM/Rec (2022) 16, del Comitato dei Ministri agli Stati membri sulla lotta contro i discorsi d'odio, adottata dal Comitato dei Ministri il 20 maggio 2022, in occasione della 132esima Sessione del Comitato dei Ministri d'Europa.

[6] M. Pelissero, *Discriminazione, razzismo e diritto penale fragile,* in *Diritto penale e procedura*, 2020, 1017 ss.; R. Bartoli-M. Pelissero-S. Seminara, *Diritto penale. Lineamenti di parte speciale,* Torino, 2022, 25; A. Cadoppi-S. Canestrari-A. Manna-M. Papa, *Diritto penale. Tomo terzo. I delitti contro l'onore e la libertà individuale*, Vicenza, 2022, 6101 ss.; P. Tanzarella, *Discriminare parlando. Il pluralismo democratico messo alla prova dai discorsi d'odio razziale,* Torino, 2020, 19-46; M. Caputo, *La menzogna di Auschwitz, le verità del diritto penale. La criminalizzazione del c.d. negazionismo tra ordine pubblico, dignità e senso di umanità,* in AA.VV. *Verità del precetto e della sanzione penale alla prova del processo*, a cura di G. Forti-G. Varraso-M. Caputo, Napoli, 2014, 309; sul tema si veda anche G. Pavich-A. Bonomi, *Reati in tema di discriminazione: il punto sull'evoluzione normativa recente, sui principi e valori in gioco, sulle prospettive legislative e sulla possibilità di interpretare in senso conforme a Costituzione la norma vigente,* in *Penale Contemporaneo*, 13.10.2014; A Galluccio, *Punire la parola pericolosa?*, Milano, 2020, 319 ss.

[7] Per un maggior approfondimento in relazione a tale profilo si rimanda all'analisi che si svolgerà nei paragrafi 3 e 4 del presente contributo.

cui si basano le moderne democrazie plurali[8].

Finalità della disamina è quella di individuare i principali profili di criticità della reintroduzione, all'interno del nostro ordinamento (e non solo), dei c.d. "reati di opinione"[9], idonei a turbare i valori morali sovra-individuali riconducibili ad un'intera collettività[10]. Al riguardo, la Corte costituzionale ha precisato che il bilanciamento degli interessi è il modello che meglio si adatta alla risoluzione della questione, in quanto consente di considerare i rapporti tra la libera manifestazione del pensiero e il cruciale controvalore che emerge, condizionato dal contesto sociopolitico di riferimento. Tale bilanciamento è chiamato a fronteggiare nuove criticità in considerazione del fatto che, con la pubblicazione online sempre più massiccia di contenuti da parte degli utenti, i fornitori di servizi digitali sono chiamati a svolgere attività di moderazione. L'attuale necessità è quella di trovare il miglior equilibrio possibile tra l'esigenza di proteggere i beni giuridici direttamente colpiti dalla diffusione del contenuto d'odio online e il diritto alla libertà di espressione, su cui si regge altresì l'odierno Stato costituzionale di diritto[11].

In aggiunta, alcune caratteristiche dei social network impongono la necessità di cooperazione tra gli Stati e le giurisdizioni: basti pensare alla velocità con cui i messaggi si diffondono, alla possibilità di raggiungere un numero indeterminato di destinatari in maniera pressoché simultanea, alla capacità del contenuto offensivo di sopravvivere per un lungo arco di tempo dalla sua immissione (anche in parti del web diverse da quelle in cui era stato originariamente inserito), alla possibilità di un "ritorno imprevedibile", all'anonimato e alla natura transnazionale degli intermediari informatici[12].

Per fornire una trattazione il più possibile esaustiva, si è deciso di adottare un metodo di tipo comparatistico, al fine di analizzare con maggiore consapevolezza l'opportunità della scelta di politica criminale effettuata dal legislatore italiano, mettendola a confronto con la legislazione radicalmente diversa vigente negli Stati Uniti d'America. Inoltre, si prenderà in considerazione l'esperienza maturata nel continente europeo, dove, passando attraverso l'esperienza giuridica francese, tedesca e spagnola, si cercheranno di analizzare le vicende legislative e giurisprudenziali che hanno segnato il cammino del ricorso alla tecnica incriminatrice in riferimento alle condotte di *online hate speech* e che appaiono, quindi, in grado di spiegarne l'attuale fisionomia.

Infine, ci si è interrogati circa l'efficacia di ricorrere a strumenti diversi dalla tecnica di stampo penalistico per contrastare (e prevenire) il fenomeno in parola, indagando

---

[8] In relazione ai rapporti con i crimini d'odio, si rimanda a L. D'Amico, *Le forme dell'odio. Un possibile bilanciamento tra irrilevanza penale e repressione*, in *La Legislazione penale,* 17.6.2020, 2 ss.

[9] Sul tema si rinvia a M. Pelissero, *La parola pericolosa. Il confine incerto del controllo penale del dissenso*, in *Questione Giustizia*, 2015, 38 e L. Alesiani, I *reati di opinione. Una rilettura in chiave costituzionale*, Milano, 2006, 141 ss.

[10] A. Spena, *Libertà di espressione e reati di opinione*, in *Rivista italiana di diritto e procedura penale*, 2, 2007, 692 ss.

[11] Quali quelli alla libera manifestazione del pensiero e alla libera informazione (ex artt. 10 e 11 Carta di Nizza). Per un approfondimento in merito, si segnala: B. Sander, *Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to content moderation,* in *Fordham International Law Journal*, 2020, 939 ss.

[12] G. Ziccardi, *Il contrasto dell'odio online: possibili rimedi*, cit., 42.

il ruolo dei meccanismi di moderazione algoritmica nella *governance* del discorso pubblico online.
Attraverso l'approfondimento delle direttive di indagine sopra individuate, l'obiettivo finale della trattazione sarà comprendere se la repressione dell'*online hate speech* possa pregiudicare l'esercizio della libertà di espressione, componente insostituibile di ogni democrazia eterogenea.

2. Il panorama giuridico europeo: la mancata inclusione dei discorsi d'odio online tra i c.d. "eurocrimini"

Si è già sottolineato come i discorsi d'odio si pongono in un rapporto di *species* rispetto al più ampio *genus* dei crimini d'odio. Essi si connotano, al pari di questi ultimi, per la sussistenza di un c.d. *bias motive*, ossia il motivo di pregiudizio che li supporta, mentre difettano, rispetto ai primi, della c.d. *criminal offence*, ossia la commissione di un reato[13]. Di talché anche l'*hate speech*, al pari degli *hate crimes,* può essere definito un crimine sorretto e motivato dal pregiudizio[14], trattandosi, secondo la definizione resa dal Comitato dei Ministri del Consiglio d'Europa nel 1997, di «discorsi suscettibili di produrre l'effetto di legittimare, diffondere o promuovere l'odio razziale, la xenofobia, l'antisemitismo o altre forme di discriminazione o odio basate sull'intolleranza»[15].
Al fine di comprendere compiutamente la portata lesiva di queste condotte, è, altresì, opportuno evidenziare come, secondo la definizione maggiormente condivisa di crimini d'odio fornita dall'Organizzazione per la Sicurezza e la Cooperazione in Europa nel 2009, essi si sostanziano nell'esternalizzazione di «violente manifestazioni di intolleranza dotate di un profondo impatto non solo sulla vittima diretta bensì anche sul gruppo con cui la vittima si identifica. Essi colpiscono la coesione della comunità e la stabilità sociale. Pertanto, una risposta vigorosa è importante sia per la sicurezza individuale che per quella comune»[16].
Da tale enunciazione emerge con chiara evidenzia l'insita duplicità che connota i cri-

[13] Ocse, *Hate Crime Laws. A practical Guide,* Adhir, Varsavia, 2009, 16, come evidenziato da L. D'Amico, *Le forme dell'odio. Un possibile bilanciamento tra irrilevanza penale e repressione*, in *La Legislazione Penale*, 2020, 6, secondo la quale proprio tale profilo rende impellente e di cruciale importanza il dibattito innestatosi in riferimento all'elemento soggettivo del reato.

[14] L. Goisis, *Crimini d'odio. Discriminazioni e giustizia penale*, Napoli, 2019, 30. Di seguito alcune delle definizioni anche ricordate da L. D'Amico, *Le forme dell'odio,* cit., 2, che si sono succedute in materia nel corso degli anni: B. Perry «il crimine d'odio [...] comporta atti violenti ed intimidatori, generalmente diretti verso gruppi già oggetto di marginalizzazione e stigmatizzazione. Così inteso, è un meccanismo di potere e di oppressione, teso a riaffermare le precarie gerarchie che caratterizzano un dato ordine sociale [...]» in B. Perry, *In the Name of Hate: Understanding Hate Crimes*, Londra 2001, 1 e 10; N. Chakraborti i crimini d'odio vanno individuati in quegli «atti di violenza, intimidazione e ostilità diretti verso persone a causa della loro identità o della loro percepita diversità» in N. Chakraborti-J. Garland, *Hate Crime. Impact, Causes, and Responses*, Los Angeles-Londra 2015, 5; F. Lawrence definisce l'hate crime – o meglio il bias crime – come un «crimine commesso per un motivo di pregiudizio (bias) contro una "caratteristica protetta", propria di un gruppo» in F. M. Lawrence, *Punishing Hate. Bias Crimes under American Law*, Cambridge 1999, 9.

[15] Comitato dei Ministri del Consiglio d'Europa, *Raccomandazione n. R (97)20.*

[16] Ocse, *Hate Crime Laws. A practical Guide*, cit., 11.

mini d'odio e che si riverbera, ineludibilmente, anche sui discorsi d'odio. Innanzitutto, si rileva una duplicità quanto ai destinatari delle condotte, volte a ledere non soltanto il singolo individualmente inteso, bensì concepito alla luce della sua riconducibilità ad una collettività. Tale profilo si riflette sulle ragioni che spingono l'autore all'azione: l'*animus* non si rivolge tanto alla vittima intesa *uti singuli*, quanto piuttosto quale mero *nuncius* del gruppo di appartenenza a cui è realmente indirizzato il contenuto del messaggio d'odio[17]. Ne deriva che tali condotte, definite dalla stessa Ocse "simboliche"[18], non arrecano pregiudizio soltanto alla sfera personale, bensì sono in grado di danneggiare la coesione e la stabilità di una società democratica, eterogenea e multiculturale, minando la sopravvivenza dell'uguaglianza e della pari dignità dei suoi componenti.
Quale corollario, i discorsi d'odio online, rinvenendo la propria matrice negli *hate crimes* e, più nello specifico, nell'*hate speech*, non costituiscono nuovi illeciti, ma configurano unicamente illeciti diversi quanto alle modalità di manifestazione di fattispecie già esistenti. Essi presentano un *quid pluris*, considerato che essi si manifestano nella dimensione della rete, che li ha trasformati e implementati sotto il profilo della permanenza, della viralità, del ritorno imprevedibile (considerato che si tratta di messaggi itineranti), dell'anonimato e del carattere transazionale che assumono le condotte[19]. Per questo motivo, si è posta la necessità di procedere con una normazione *ad hoc* per contrastare fenomeni già noti e inizialmente concepiti a livello normativo nella sola dimensione reale[20].
Nel corso dell'ultimo ventennio, la domanda di regolamentazione è aumentata in modo significativo[21], sia nella dimensione offline sia in quella online, e l'incremento dei discorsi d'odio anche sul web ha messo in luce le lacune normative riguardanti la commissione dei crimini in commento anche nel mondo reale. È stato osservato come Internet, da luogo definito, alle origini, intrinsecamente democratico, si è rivelato, invece, in grado di porre in pericolo il fondamentale presidio di garanzia proprio dello Stato inteso come collettività, considerata l'irrefrenabile e quasi incontrollabile capacità di diffondere disinformazione e di incentivare ostilità[22].
Conseguentemente, al pari dei legislatori nazionali, anche il legislatore europeo si è interessato ai rapporti tra diritto e realtà digitale, altresì, in considerazione della progressiva crescita di potere delle piattaforme digitali e delle criticità legate all'esercizio della libertà di espressione degli utenti nella nuova società dell'informazione. Nell'ecosistema del web, i discorsi d'odio si diffondono tramite i gestori di dati, che ne rafforzano

[17] Per un maggiore approfondimento di tali profili, vedasi L. D'Amico, *Le forme dell'odio*, cit., 3.

[18] Ocse, *Hate Crime Laws. A practical Guide*, cit., 17.

[19] G. Ziccardi, *Il contrasto dell'odio online: possibili rimedi*, cit. 42-44-45.

[20] Si rimanda, ai fini di un maggiore approfondimento della questione, a Oliveri F., *Diritti degli internauti, obblighi degli Stati, responsabilità delle piattaforme digitali: problemi regolativi in materia di odio online*, in *Teoria e Critica della Regolazione Sociale/Theory and Criticism of Social Regulation*, 2, 23, 2021, 116. Per una analisi dettagliata dei caratteri distintivi del discorso d'odio online vedasi UNESCO, *Countering online hate speech*, Parigi, 2015.

[21] In senso difforme, altri rilevano la non necessarietà di una legislazione specifica, atteso che i medesimi interessi possono essere perimenti protetti dalla legislazione di carattere generale: F. Easterbrook, *Cyberspace and the Law of the Horse*, in *University of Chicago Legal Forum*, 1, 1996, 207 e 208.

[22] Oliveri F., *Diritti degli internauti, obblighi degli Stati, responsabilità delle piattaforme digitali*, cit., 106.

notevolmente la portata, tanto dei comportamenti quanto dei gestori stessi. Di tale fenomeno hanno beneficiato le principali aziende tecnologiche, che sono divenute attori di primaria grandezza nelle dinamiche sociali, politiche e culturali e hanno acquisito un potere senza precedenti, la cui regolamentazione assume rilevanza pubblicistica[23].
Il Trattato sul funzionamento dell'Unione europea contempla, agli artt. 9 e 10 - tra gli obiettivi in esso annoverati - la promozione della lotta contro ogni tipo di discriminazione e all'art. 19 assegna chiara rilevanza all'esigenza di contrastare ogni condotta di tal sorta[24].
L'intervento legislativo europeo in questa materia trova fondamento giuridico in una duplicità di disposizioni. In primo luogo, si segnala la rilevanza dell'art. 83, par. 1, TFUE[25], che attribuisce alle istituzioni europee una competenza penale "diretta"[26]. Ne deriva che l'Unione europea detiene il potere di richiedere agli Stati membri l'adozione di norme incriminatrici ogniqualvolta vengano in rilievo sfere di criminalità particolarmente grave e che presentano una dimensione transnazionale, tra le quali il Trattato annovera anche le forme di "criminalità informatica". Questa nozione, al pari di quella di "cybercrime", lungi dal rivenire una definizione giuridica univoca, abbraccia una molteplicità di comportamenti lesivi di interessi penalmente rilevanti riconducibili alla categoria dei "reati informatici". Nel dettaglio, la criminalità informatica include non soltanto i "reati informatici in senso stretto", ma anche i "reati informatici in senso lato". Se, da un lato, i primi presentano elementi di tipizzazione connessi a procedimenti di automatizzazione dei dati o a una connotazione squisitamente tecnologica, dall'altro, i secondi si sostanziano in figure criminose che, pur non recando elementi di tipizzazione legislativa tecnologicamente caratterizzati, possono essere applicate a fatti commessi tramite la tecnologia. Tali reati potrebbero essere commessi anche offline, al di fuori del cyberspazio; tuttavia, ricevono un contributo rilevante, in termini di capacità lesiva, dall'utilizzo di dati, dispositivi o strutture informatiche[27].

[23] Ivi, 105-106.

[24] P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet e i social media = Regulatory analysis in the field of contrast to hate speech on the Internet and social media*, in *H-ermes. Journal of Communication*, 23, 2023, 29.

[25] In forza del quale «Il Parlamento europeo e il Consiglio, deliberando mediante direttive secondo la procedura legislativa ordinaria, possono stabilire norme minime relative alla definizione dei reati e delle sanzioni in sfere di criminalità particolarmente grave che presentano una dimensione transnazionale derivante dal carattere o dalle implicazioni di tali reati o da una particolare necessità di combatterli su basi comuni. Dette sfere di criminalità sono le seguenti: terrorismo, tratta degli esseri umani e sfruttamento sessuale delle donne e dei minori, traffico illecito di stupefacenti, traffico illecito di armi, riciclaggio di denaro, corruzione, contraffazione di mezzi di pagamento, criminalità informatica e criminalità organizzata».

[26] L. Picotti, *La nozione di "criminalità informatica" e la sua rilevanza per le competenze penali europee*, in *Rivista trimestrale di diritto penale dell'economia*, 4, 2011, 827.

[27] A. Mattarella, *Il cybercrime nell'ordinamento italiano e le nuove prospettive dell'Unione europea e delle Nazioni Unite*, in *Diritto penale e Processo*, 6, 2022, 809-810, che, alla nota 2, in riferimento agli interventi legislativi operati in ambito europeo in relazione a profili di diritto penale sostanziale concernenti la lotta alla criminalità informatica, rinvia alle seguenti direttive europee: direttiva 95/46/CE sulla tutela dei dati personali, direttive in materia di protezione dei diritti d'autore e, in particolare, direttiva 2001/29/CE; la direttiva 2000/31/CE sul commercio elettronico, oltre che alle decisioni quadro contro gli attacchi informatici (2005/222/GAI), contro lo sfruttamento sessuale di minori e la pedopornografia (2004/68/GAI), contro il terrorismo (2002/475/GAI, parzialmente riformata dalla decisione 2008/919/GAI).

Questa impostazione teorica trova riscontro, tra gli altri, nell'art. 14 della "Convenzione sul Cybercrime", le cui disposizioni processuali si applicano a tutti gli illeciti commessi attraverso sistemi informatici e a quelli per i quali è necessaria la raccolta della prova elettronica[28], nonché negli artt. 171, lett. a)-bis e 171-ter, comma 2, lett. a-bis), l. 633/1941 (c.d. "Legge sul diritto d'autore"), che sanzionano la diffusione abusiva realizzata mediante l'immissione in un sistema di reti telematiche di un'opera dell'ingegno protetta[29].

Alla luce di quanto detto e in virtù dell'ampiezza della nozione di "criminalità informatica", la stessa è in grado di abbracciare tutte le fattispecie commesse mediante l'apporto della tecnologia, che non deve costituirne il mezzo essenziale, essendo sufficiente che venga utilizzata incidentalmente per la commissione del reato. Di conseguenza, anche i discorsi d'odio, ogniqualvolta realizzati nell'ambiente digitale, ricadono nella categoria dei c.d. "eurocrimini" [30], in relazione ai quali, l'Unione europea detiene una potestà legislativa "diretta" ai sensi dell'art. 83, par. 1, TFUE[31].

Secondariamente, molte volte la Commissione europea ha proposto di ampliare l'elenco dei c.d. "reati europei" ai crimini d'odio e all'incitamento all'odio, anche online[32], mediante il ricorso all'art. 83, par. 2, TFUE[33], che prevede gli ambiti di competenza penale "indiretta" dell'Unione europea[34], trattandosi di forme di criminalità particolarmente grave, la cui commissione dà luogo ad un forte impatto sulla persona e sulla

Sul piano processuale, rinvia, invece, alle direttive sul mandato d'arresto europeo (2002/584/GAI) e sull'applicazione del principio del reciproco riconoscimento delle decisioni di confisca (2006/783/GAI), che includono la "criminalità informatica" nelle liste di reati per cui si prescinde, in conformità con il principio del mutuo riconoscimento, dal requisito della doppia incriminazione per l'esecuzione diretta dei provvedimenti emessi dall'autorità giudiziaria dello Stato richiedente.

[28] Art. 14 della Convenzione sulla criminalità informatica del Consiglio d'Europa, c.d. "Convenzione sul Cybercrime", Budapest, 23 novembre 2001, per un cui maggiore approfondimento a livello internazionale si rinvia a K. Ghimire, *International Perspective of Cyber Law: Specially Focused on Cybercrime Convention*, in *NJA Law Journal*, 2021, 307 ss.

[29] R. Flor, *Tutela penale e autotutela tecnologica dei diritti d'autore nell'epoca di internet. Un'indagine comparata in prospettiva europea ed internazionale*, Cedam, 2010.

[30] Per tali intendendosi tra le «sfere di criminalità particolarmente grave che presentano una dimensione transnazionale» ai sensi dell'art. 83, §1, co. 2 del Trattato sul Funzionamento dell'Unione europea, come precisato da N. Cardinale, *Il Parlamento europeo chiede l'inserimento della violenza di genere tra i c.d. eurocrimini ai sensi dell'art. 83, § 1, co. 2 del TFUE* in sistemapenale.it, 07 dicembre 2021.

[31] A conferma delle considerazioni sopra svolte, si segnala la rilevanza della direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio, del 14 maggio 2024, sulla lotta alla violenza contro le donne e alla violenza domestica, adottata ai sensi dell'art. 83, par. 1, TFUE.

[32] P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet e i social media*, cit., 34 - 35; Per un maggior approfondimento vedasi anche *Study to support the preparation of the European Commission's initiative to extend the list of EU crimes in Article 83 of the Treaty on the Functioning of the EU to hate speech and hate crime*, Commissione europea, novembre 2021, 32 ss.

[33] In forza del quale «Allorché il ravvicinamento delle disposizioni legislative e regolamentari degli Stati membri in materia penale si rivela indispensabile per garantire l'attuazione efficace di una politica dell'Unione in un settore che è stato oggetto di misure di armonizzazione, norme minime relative alla definizione dei reati e delle sanzioni nel settore in questione possono essere stabilite tramite direttive. Tali direttive sono adottate secondo la stessa procedura legislativa ordinaria o speciale utilizzata per l'adozione delle misure di armonizzazione in questione, fatto salvo l'articolo 76».

[34] L. Picotti, *La nozione di "criminalità informatica" e la sua rilevanza per le competenze penali europee*, cit., 829.

comunità di appartenenza[35], da cui deriva una seria minaccia per i valori democratici. Tuttavia, nessuna di tali basi giuridiche è stata finora impiegata per criminalizzare le condotte di *online hate speech*. L'approccio europeo è, difatti, storicamente risultato timido, sostanziandosi nell'adozione, a alternanza, di strumenti di *hard law*, i cui destinatari sono apparsi (con esito deludente, anche se, per alcuni profili, comprensibilmente, per le ragioni sopraccitate) i prestatori di servizi informativi (d'ora in poi, PSI), insinuando il dubbio che si trattasse di provvedimenti principalmente funzionali alla tutela del mercato interno anziché dei diritti fondamentali e, nella specifica materia penalistica, di strumenti di *soft law*, ossia atti di cooperazione giudiziaria[36].

Ne è risultata l'adozione di una metodologia minimalista[37], innestata sulla regolamentazione del mondo digitale esclusivamente nei suoi tratti essenziali (in particolare, la responsabilità dei PSI) e non a mezzo di "scelte forti"[38] di incriminazione delle condotte in parola, la cui diffusione viene contrastata a livello europeo in un'ottica preventiva e non repressiva[39], cui corrisponde, quale contraltare, anche la rilevanza strategica dei soggetti appena menzionati nella circolazione di contenuti nel cyberspazio[40].

Al contrario, è stato rilevato che è la potenza stessa della tecnologia ad esigere la messa a punto di misure di *governanc*e di Internet mediante l'adozione di strumenti di *hard law*, atteso che quelli di *soft law*, seppure utili e recanti propositi assai meritevoli di considerazione, non risultano sufficienti ad assicurare una adeguata salvaguardia dei diritti degli utenti[41].

I primi strumenti di *hard law* maggiormente risalenti nel tempo erano stati concepiti a livello di diritto derivato: il riferimento è alla direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000 volta all'incentivazione del commercio elettronico[42], che è stata definita «l'unico atto di respiro generale specificamente dedicato ai servizi digitali»[43]. Essa, infatti, ha costituito lo strumento principale che ha

---

35 Commissione europea, *Un'Europa più inclusive e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio*, COM (2021)777 final.

36 I. Anrò, *Online hate speech: la prospettiva dell'Unione europea tra regolamentazione della condotta dei prestatori di servizi intermediari e ricorso al diritto penale*, in *Osservatorio sulle fonti*, 16, 1, 2023, 14-15.

37 A riguardo vedasi O. Lynskey, *Regulating Platforms Power*, in *LSE Law, Society and Economy Working Papers*, 1, 2017.

38 Alle quali ha fatto riferimento, con specifico riguardo alla regolazione di Internet, M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di comparazione*, cit., 73.

39 Vedasi, sul punto, quanto si dirà più compiutamente infra in relazione agli strumenti di moderazione algoritmica nel par. 7.

40 M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di comparazione*, cit., 73.

41 C. Confortini, *Diffamazione e discorso d'odio in internet Note a margine di Cass., [ord.] 16.5.2023, n. 13411*, in *Persona e mercato*, 4, 2023, 703, che richiama sul punto L. Floridi, *The End of an Era: from Self-Regulation to Hard Law for the Digital Industry*, in *Philosophy & Technology*, 34, 2021, 621.

42 Direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000 *relativa a taluni aspetti giuridici dei servizi della società dell'informazione, in particolare il commercio elettronico, nel mercato interno* (Direttiva sul commercio elettronico), in GUUE L 178 del 17 luglio 2000, p. 1 ss. Per un maggior approfondimento in relazione all'evoluzione della normativa europea vedasi I. Anrò, *Online hate speech*, cit., 16 ss. e G. Morgese, *Moderazione e rimozione dei contenuti illegali online nel diritto dell'Unione europea*, in *federalismi.it*, 12 gennaio 2022, 80 ss.

43 M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di*

legiferato in merito agli spazi di autonomia e ai profili di responsabilità dei gestori delle piattaforme digitali. È d'obbligo rilevare come, poiché è trascorso oltre un ventennio dalla pubblicazione, detta direttiva non può considerarsi (e non si poteva considerarla nemmeno allora) capace di disciplinare efficacemente le diverse condotte d'odio e discriminatorie che si realizzano a mezzo di Internet[44].

Inoltre, la scarna e datata disciplina ivi contenuta si connota per essere stata «mobilizzata da più di un decennio di giurisprudenza quantomeno creativa»[45] della Corte di Giustizia dell'Unione europea, la quale ha rivoluzionato il proprio orientamento, spingendosi a riconoscere potere sempre maggiore ai PSI, tenuti, dietro ordine dello Stato membro, ad eliminare i contenuti oggetto di segnalazione. Da tale circostanza è anche derivata la configurabilità, in capo ai medesimi, di un dovere di sorveglianza sul materiale circolante sul *web*, oltre che un certo margine di discrezionalità quanto alla qualificazione del medesimo[46].

Il ricorso alla tecnica incriminatrice propria del diritto penale è stato oggetto di un tentativo di utilizzo da parte dell'Unione europea mediante la Decisione quadro 2008/913/GAI del Consiglio del 28 novembre 2008[47], la quale obbliga gli Stati membri a punire, mediante l'adozione di sanzioni effettive, proporzionate e dissuasive, le condotte di istigazione pubblica alla violenza o all'odio, indipendentemente dal fatto che esse si configurino online oppure offline, in ragione della razza, della religione e dell'ascendenza ovvero dell'origine nazionale o etnica[48]. La Decisione quadro mira a realizzare un'opera di armonizzazione minima delle legislazioni nazionali degli Stati membri, i quali possono decidere di ampliare la sfera dei motivi di odio che giustificano una repressione penale. Le difficoltà riscontrate nella sua lunga gestazione, sono riconducibili, principalmente, all'esigenza di realizzare un bilanciamento tra la libertà di espressione e la lotta contro il razzismo e la xenofobia nel complesso quadro delle diverse cornici giuridiche nazionali[49], che presentano tratti differenziali anche in

comparazione, cit., 71.

[44] Per un approfondimento circa i punti salienti della legislazione europea in commento, vedasi P. Faletta, *Controlli e responsabilità dei social network sui discorsi d'odio online*, cit., 150-151. A ciò aggiungasi che, la detta direttiva si connota per una contraddizione in termini: se, da una parte, l'art. 15 statuisce l'assenza di un obbligo di sorveglianza in capo ai PSI, dall'altra parte si prevedono, nelle *Disposizioni finali*, obblighi di rimozione dei contenuti a fronte di ordine in tal senso delle autorità pubbliche ed oneri di notifica (c.d. «*notice and take down*») che consentono alle dette autorità di individuare e prevenire attività illecite ai sensi dell'art. 21. Si è, così, consentita una certa espansione della libertà di espressione, creando significative esenzioni di responsabilità in capo ai PSI e riconoscendo in capo ai medesimi un obbligo di rimozione a posteriori, una volta appurata l'illiceità dei contenuti ospitati.

[45] M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di comparazione*, in questa *Rivista*, 2, 2021, 71.

[46] P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet ei social media*, cit., 30-31.

[47] Decisione quadro 2008/913/GAI del Consiglio del 28 novembre 2008 sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale, in GUUE, L. 328 del 6 dicembre 2008.

[48] P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet ei social media*, cit., 29, che evidenzia che, successivamente, il Parlamento europeo, con la risoluzione del 14 marzo 2013 ha sottolineato la necessità di includere anche le manifestazioni di antisemitismo, intolleranza religiosa, antiziganismo, omofobia e transfobia nell'ambito applicativo della Decisione quadro in esame.

[49] Difatti, numerose e serie sono state le difficoltà che si sono frapposte all'adozione della Decisione; a

termini sociali e culturali. Difatti, è stato rilevato che uno dei principali profili di problematicità e che meriterebbe maggiore attenzione, riguarda la dimensione culturale della questione. Si ravvisa la necessità di sensibilizzare sia le popolazioni sia le stesse piattaforme digitali, al fine di aumentare la consapevolezza delle potenzialità lesive insite nelle conversazioni d'odio online[50].

Più di recente, le istituzioni europee hanno adottato un approccio di autoregolamentazione e co-regolamentazione, in un'ottica di proficua collaborazione con le piattaforme online, a cui è stato attribuito il potere di censurare i contenuti illeciti[51], nonostante trattasi di comunità virtuali, che perseguono finalità commerciali[52], aliene dall'interferenza dei poteri statali[53] e a qualsiasi circuito democratico[54].

Come anticipato, una delle principali criticità emerse ha riguardato l'assimilazione dei PSI ai caratteri della pubblicità, in quanto diventati centri di straordinaria concentrazione di potere, conoscenza e ricchezza[55].

Su questa scia si è posto anche il Regolamento 2022/2065/UE del Parlamento europeo e del Consiglio del 19 ottobre 2022[56], noto come "Digital Services Act" o "DSA", con efficacia dal 17 febbraio 2024. L'intervento legislativo europeo si è radicato nell'esigenza improrogabile di garantire uniformità tra le legislazioni: la presenza di quadri giuridici differenti è apparsa non più sostenibile per regolamentare un fenomeno tipicamente transfrontaliero e che incide negativamente anche sul mercato interno. Sono così state ideate innovative forme di collaborazione tra le istituzioni e i PSI, tenendo conto del loro ambito di azione e della possibilità che si verifichino condotte d'odio, ovvero di discriminazione, all'interno delle piattaforme[57].

Il Regolamento ha disegnato una commistione tra *private* e *public governance* del discorso pubblico online, mediante il consolidamento della cooptazione tra poteri pubblici e

ciò aggiungasi il consistente ritardo riscontrato da numerosi Stati Membri in sede di trasposizione: basti pensare che l'Italia vi ha provveduto soltanto nel 2017 e che altri, quali la Romania e l'Estonia, sono stati destinatari di una lettera di messa in mora ex art. 258 TFUE dal parte della Commissione europea in ragione del perdurante mancato recepimento, a conferma della delicatezza del bilanciamento da operarsi alla luce di differenti tradizioni culturali, sociali e giuridiche, così, I. Anrò, *Online hate speech*, cit., 31-32.

50 G. Ziccardi, *Il contrasto dell'odio online: possibili rimedi*, cit. 42-45.

51 P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet ei social media*, cit., 30.

52 Nell'era digitale, si parla, infatti, di compenetrazione con effetti orizzontali della libertà di espressione, così M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di comparazione*, cit., 68 e 70 che richiama S. Gardbaum, *The "Horizontal Effect" of Constitutional Rights*, in *Michigan Law Review*, 102, 2003, 387 ss.; nel panorama italiano, vedasi, sul punto, anche gli scritti di A. Gentili, *Diritti fondamentali e rapporti contrattuali. Sulla efficacia orizzontale della Convenzione europea dei diritti dell'uomo*, in *Nuova giurisprudenza civile commentata*, 1, 2016, 183 ss.; A. Zoppini, *Il diritto privato e le "libertà fondamentali" (Principi e problemi della Drittwirkung nel mercato unico)*, in *Rivista di diritto civile*, 3, 2016, 712 ss.

53 M. Bassini, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati" Spunti di comparazione*, cit., 71.

54 C. Confortini, *Diffamazione e discorso d'odio in internet*, cit., 699.

55 *Ibid.*

56 Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), in GUUE L 277 del 27 ottobre 2022, 1 ss.

57 P. Falletta, *Controlli e responsabilità dei social network sui discorsi d'odio online*, in questa *Rivista,* 1, 2020, 147.

privati[58] e ciò in ragione dell'ampiezza acquisita dal mercato digitale, essenziale strumento di informazione e di partecipazione al dibattito politico e culturale. L'approccio adottato è stato definito "risk-based", in quanto vengono fissate regole di condotta e meccanismi di *enforcement* diversificati in funzione dell'attività svolta dai *providers*, della loro dimensione e della loro capacità di incidere sui valori fondamentali dell'Unione. Il legislatore europeo ha, quindi, optato per il superamento di una disciplina unitaria e per l'adozione di paradigmi regolatori differenziati, allo scopo di assicurare la legalità delle comunicazioni che hanno luogo sul web[59].

Nel complesso, gli Stati membri, attraverso l'introduzione delle normative descritte, hanno cercato di contrastare «l'opacità dello spazio digitale e garantire la costruzione di una solida democrazia digitale»[60], tentando di colmare il rapido processo di erosione delle di loro sovranità che le società private stavano realizzando. Nondimeno, l'azione legislativa si è concentrata eccessivamente sui profili strettamente economici legati al corretto funzionamento del mercato interno e del mercato digitale, anziché sulla repressione delle condotte in grado di porre in pericolo la sopravvivenza delle democrazie europee. Ne è seguita una stretta connessione tra i provvedimenti legislativi riguardanti l'ecosistema digitale e il concetto di sovranità[61].

Quanto al formante giurisprudenziale europeo, si evidenzia come la Corte di Giustizia abbia tratto ispirazione dai risultati della Corte Europea dei diritti dell'Uomo, che storicamente hanno manifestato diffidenza nei confronti del *cyberspace*[62]. Nel solco dei propri precedenti, i giudici di Strasburgo hanno confermato che l'adozione di misure di contrasto ai discorsi d'odio mediante il ricorso allo strumento del diritto penale deve essere considerata priva di profili di illiceità, apponendo, tuttavia, quale limite, la circostanza che detti discorsi presentino un maggior grado di offensività tale da poterli inquadrare nelle forme di *hate speech*, le cui manifestazioni più gravi costituiscono fattispecie di abuso dell'art. 10[63].

Pur essendo evidente l'impegno profuso dalle istituzioni europee e, in particolare, dalla Commissione, nel contrasto ai discorsi d'odio online, tale azione presenta tuttora delle criticità, alla luce della mancata adozione di direttive idonee a estendere l'ambito dei c.d. "eurocrimini" e ricondurre tali reati informatici alle materie di legislazione penale europea in forza dell'art. 83, par. 1, TFUE, ovvero, in subordine, dell'art. 83, par. 2, TFUE[64].

Nel settembre 2020, è stato annunciato, ad opera della Presidente della Commissione

---

58 C. Confortini, *Diffamazione e discorso d'odio in internet*, cit., 702-703.

59 S. Braschi, *Il nuovo Regolamento sui servizi digitali: quale futuro per la responsabilità degli Internet Service Provider?,* in *Diritto Penale E Processo*, 3, 2023, 368-369.

60 P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet ei social media*, cit., 36.

61 Per un maggior approfondimento della questione, vedasi S. Torregiani, *Il Data Act: una versione europea del Data Nationalism?*, in *Rivista italiana di informatica e diritto*, 5, 2, 2023, 135.

62 CEDU, *Sanchez c. Francia*, ric. 45581/15 (2021).

63 Come nel caso del contesto politico, così P. Dunn, *Carattere eccezionale dell'"hate speech" e nuove forme di responsabilità per contenuti di terzi nella giurisprudenza EDU. Nota a C.edu, Sanchez c. Francia, 15 maggio 2023*, in *Osservatorio costituzionale*, 6, 2023, 243-244.

64 Per un maggior approfondimento a riguardo vedasi I. Anrò, *Online hate speech*, cit., 32 ss.

Ursula Von Der Leyen, l'obiettivo di inserire le condotte di *hate speech* e *hate crime*[65] all'interno di quest'ultimo novero, proposta cui è seguita, in data 18 gennaio 2024, l'approvazione da parte del Parlamento europeo di una Risoluzione volta a «Estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio»[66].

La Risoluzione evidenzia che tutte le manifestazioni di odio e di intolleranza, inclusi l'incitamento e i reati generati dall'odio, sono incompatibili con i valori dell'Unione, quali la dignità, la libertà, la democrazia, l'uguaglianza, lo Stato di diritto e il rispetto dei diritti umani, compresi i diritti delle persone appartenenti a minoranze, sanciti dall'art. 2 TUE[67]. Oltre a ciò, è stato osservato come la libertà di espressione costituisca un valore fondamentale delle società democratiche e non debba essere limitata in modo ingiustificato; qualsiasi legislazione a riguardo dovrebbe, pertanto, basarsi sui principi di necessità e proporzionalità, così da garantirne l'esercizio in conformità all'art. 11 della Carta dei diritti fondamentali dell'Unione europea, che non potrebbe mai essere utilizzato come scudo per la commissione di reati[68].

La Risoluzione è stata trasmessa alla Commissione e al Consiglio, esortando quest'ultimo a adottare una Decisione che ampli l'elenco dei c.d. "reati europei", affinché la Commissione possa avviare la seconda fase della procedura di modifica del Trattato sul funzionamento dell'Unione europea. La progettazione di un'iniziativa comune a livello europeo e la realizzazione di un solido quadro unionale, mediante l'adozione di misure globali, rappresenterebbe la risposta più efficace alle sfide di cui sopra[69], cui, però, osta la necessità che la Decisione venga adottata dal Consiglio all'unanimità, soglia assai difficile da raggiungere, attesa la titolarità, in capo a ciascuno Stato, del potere di veto e in considerazione della diversità dei contesti sociali e culturali che fanno da sfondo ai quadri giuridici degli Stati membri. All'interno di un quadro democraticamente ordinato, plurale e eterogeneo quale quello europeo, un epilogo da ritenersi necessario e imprescindibile viene precluso dall'agire democratico stesso del sistema di funzionamento, che assegna cruciale rilevanza alla voce di ciascuno Stato membro. Per tale ragione, il Parlamento ha raccomandato al Consiglio la modifica dell'art. 83 TFUE, al fine di renderlo soggetto ad una maggioranza qualificata rafforzata, anziché all'attuale unanimità, attraverso l'attivazione della c.d. "clausola passerella"[70] di cui

[65] Comunicazione della Commissione del 9 dicembre 2021 dal titolo «Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio» (COM (2021)0777), citata in I. Anrò, *Online hate speech*, cit., 32 ss.

[66] Risoluzione del Parlamento europeo del 18 gennaio 2024 (P9_TA (2024)0044), «Estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio» [2023/2068(INI)] (A9-0377/2023).

[67] Ivi, 3.

[68] Ivi, 8.

[69] Come confermato, oltre che dalla Risoluzione in commento, anche dalla Comunicazione della Commissione al Parlamento Europeo e al Consiglio «Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio» COM/2021/777 final.

[70] Risoluzione del Parlamento europeo del 18 gennaio 2024 (P9_TA (2024)0044), 8.

all'art. 48, par. 2, TUE[71].

In conclusione, il Parlamento europeo deplora con forza il fatto che sono trascorsi quasi due anni dalla pubblicazione della Comunicazione e che il Consiglio non ha compiuto alcun progresso in merito, sebbene sia stato in grado di estendere rapidamente l'elenco dei c.d. "eurocrimini" per altri fini. Tale inazione risulta ancor più riprovevole alla luce dell'aumento statistico dell'incitamento all'odio e dei reati generati dall'odio[72].

Alla luce del quadro sopra delineato, emerge con chiarezza che il problema sottostante al mancato avvio di un intervento legislativo mediante lo strumento della direttiva, sulla base dell'art. 83, par. 1 (e, eventualmente, par. 2), TFUE, sia da ricondursi a una volontà di stampo politico. A tal proposito, è stata rilevata la mancanza di una preparazione socioculturale adeguata riguardo il grado di lesività insito nelle condotte in esame. Nel proseguo si approfondirà come diversi Stati europei abbiano criminalizzato i discorso d'odio online, ciascuno secondo la propria sensibilità nazionale. Tuttavia, per un'efficace neutralizzazione non può dirsi sufficiente la sola incriminazione, altresì, considerato che il ricorso alla "clava penale"[73] segna il fallimento delle democrazie contemporanee nel loro impegno a educare seriamente e in modo effettivo i propri cittadini al rispetto e all'uguaglianza[74].

La natura globale di Internet rende impossibile una regolamentazione giuridica onnicomprensiva del cyberspazio e il coinvolgimento di tutti gli attori coinvolti nella catena di propagazione dei contenuti si rivela indefettibile. Nella sua duplice veste di trasmettitore e destinatario, è necessario configurare in capo all'utente sia una specifica responsabilità, sia una specifica forma di tutela[75]. Dunque, è indifferibile cercare delle alternative per limitare la pubblicazione dei discorsi d'odio e minimizzare i pregiudizi arrecati dagli stessi, mediante uno sforzo coordinato che veda la cooperazione delle componenti legislativa, tecnologica e socioculturale. Questa soluzione permetterebbe di mitigare la diffusione senza influire negativamente sul libero flusso di conoscenze, idee e informazioni all'interno della rete[76].

Per risolvere in maniera definitiva le problematiche connesse alle conversazioni d'odio online, che nascono altresì da fenomeni di emarginazione sociale, diventa cruciale

---

[71] La relativa attivazione dà luogo ad un procedimento di revisione dei Trattati in forma semplificata tramite il quale è possibile disporre il passaggio dalla procedura legislativa speciale a quella ordinaria, ovvero disporre la sostituzione della regola dell'unanimità con quella della maggioranza qualificata per l'assunzione delle delibere da parte del Consiglio. Difatti, il potere deliberativo è attribuito al Consiglio europeo, che lo esercita mediante l'adozione di una decisione assunta all'unanimità, a condizione della mancata opposizione dei Parlamenti nazionali e previa approvazione del Parlamento europeo, che si pronuncia a maggioranza dei suoi membri.

[72] Risoluzione del Parlamento europeo del 18 gennaio 2024 (P9_TA (2024)0044), 7.

[73] A. Pugiotto, *Le parole sono pietre? I discorsi di odio e la libertà di espressione nel diritto costituzionale,* in *Diritto penale contemporaneo Rivista trimestrale,* 3, 2013, 5.

[74] G. Giorgini Pignatiello, *Profili Comparati e problemi costituzionali della legislazione contro l'omotransfobia. Il caso spagnolo e quello italiano*, in *Diritto pubblico comparato ed europeo*, 4, 2020, 1022.

[75] *Avis relatif à la lutte contre la haine en ligne (A - 2021 - 9), NOR: CDHX2122366V, JORF n°0170 du 24 juillet 2021, Texte n° 79*, in legifrance.gouv.fr.

[76] J. Banks, *Regulating hate speech online*, in *International Review of Law, Computers & Technology*, 24, 3, 2010, 238-239.

ricorrere a strumenti quali l'educazione digitale, anche già in ambito scolastico, la formazione di tutta la popolazione e, infine, il ruolo della politica legislativa, in grado di sensibilizzare l'opinione pubblica[77].
Soltanto promuovendo, a livello sia nazionale sia europeo, l'integrazione sociale e le sue potenzialità, nel lungo termine si potrà garantire coesione e sicurezza, impedire la marginalizzazione delle singole appartenenze politiche[78] e radicare la cultura del contrasto ai discorsi d'odio online.
Il primo baluardo contro la fitta trama di ostilità reciproche che pervade il mondo digitale[79] è l'educazione e la consapevolezza del digitale stesso. Tale assunto può aiutare a spiegare per quale ragione lo strumento forte del diritto penale, seppure importante, non è sufficiente per contrastare efficacemente le condotte d'odio, obiettivo essenziale di ogni democrazia plurale.

3. Il quadro normativo e giurisprudenziale italiano

Volendo ora indagare quale sia il ruolo legittimamente ascrivibile al diritto penale nel contrasto alle condotte d'odio online, ci si concentrerà sull'analisi normativa e giurisprudenziale interna.
In Italia non esiste uno strumento normativo che riconosce e sanziona in modo specifico l'*hate speech* (online); tale condotta è perseguita (seppur con difficoltà) quando ricade sotto il cappello di fattispecie penali esistenti, come quelle previste dagli artt. 604-bis c.p. (propaganda e istigazione a delinquere per motivi di discriminazione razziale etnica e religiosa) ovverosia, 595 c.p. (diffamazione), aggravata ai sensi dell'art. 604-ter c.p.. Di tali disposizioni parleremo più diffusamente infra, ritenendosi ora preliminarmente opportuno contestualizzare l'evoluzione normativa con riferimento alla criminalizzazione delle condotte d'odio genericamente intese, la quale ha condotto all'inserimento delle citate fattispecie entro il perimento del Codice penale.
Nel 1975 l'Italia con la l. 654/1975[80], c.d. legge Reale, ha dato esecuzione alla Convenzione ONU «sulla eliminazione di tutte le forme di discriminazione razziale», firmata a New York nel 1966 (G.U. n. 337/1975)[81]. Tale, ha introdotto nel nostro ordinamento

[77] *Avis relatif à la lutte contre la haine en ligne (A - 2021 - 9),* cit.

[78] D. Piccione, *L'espressione del pensiero ostile alla democrazia, tra diritto penale dell'emotività e psicologia collettiva*, in questa *Rivista*, 3, 2018, 87-88.

[79] A. Spena, *La parola (-) odio*, cit., 577-578.

[80] Prima dell'entrata in vigore della c.d. L. Reale, con la l. 645/1952 (c.d. L. Scelba) l'Italia ha dato attuazione alla XII disposizione transitoria e finale della Costituzione imponendo il divieto di riorganizzazione del partito fascista. Successivamente, con la L. 962 del 9 ottobre 1967, nota come legge «per la prevenzione e repressione del delitto di genocidio», l'Italia ha punito (art. 8) «chiunque pubblicamente istiga a commettere alcuno dei delitti preveduti negli articoli da 1 a 5, è punito per il solo fatto dell'istigazione, con la reclusione da tre a dodici anni. La stessa pena si applica a chiunque pubblicamente fa l'apologia di alcuno dei delitti preveduti nel comma precedente».

[81] L'art. 3 della L. in esame incriminava chiunque diffondesse, in qualsiasi modo, idee fondate sulla superiorità o sull'odio razziale o incitasse in qualsiasi maniera alla discriminazione o alla commissione di atti di violenza o provocazione alla violenza nei confronti di certe persone in quanto appartenenti ad un gruppo nazionale, etnico o razziale.

autonome fattispecie di reato caratterizzate dalla matrice razzista: la propaganda razzista, l'incitamento alla discriminazione razziale e agli atti di violenza nei confronti dei medesimi soggetti e, infine, la costituzione di associazioni e di organizzazioni con scopo di incitamento all'odio o alla discriminazione razziale. Condotte tutte riconducibili alla nozione di *hate speech* e di *hate crime*, in rapporto di specie a genere[82].

Successivamente, la l. 205/1993, c.d. legge Mancino, ha introdotto tra i motivi di incitamento alla commissione di atti discriminatori (art. 3 della l. 654/1975[83]) sia il fattore etnico, idoneo a connotare la propaganda, che quello religioso. Tale normativa è altresì nota per aver inserito, sempre nell'ambito dell'art. 3 della legge Reale, l'aggravante dell'odio razziale (etnico, nazionale, religioso)[84].

In questo complesso impianto antidiscriminatorio si è innestata, nel 2006, la l. 85/2006, recante «modifiche al Codice penale in materia di reati di opinione», la quale ha nuovamente modificato il citato art. 3, sostituendo il termine «diffusione» con quello di «propaganda» ed il termine «incitamento» con quello di «istigazione»[85]. La modifica non è di poco conto se si considera che la qualificazione del reato deve oggi corrispondere a condotte di maggiore gravità (propaganda e istigazione in luogo della diffusione e dell'incitamento).

Nel 2008, poi, la Decisione Quadro 2008/913/GA[86] (attuata in Italia tramite la l. 115/2016) ha modificato nuovamente l'art. 3, introducendo un successivo comma 3-bis*,* attraverso il quale è stata prevista la misura della reclusione da due a sei anni nei casi in cui la propaganda, l'istigazione e l'incitamento, comportino il concreto pericolo di diffusione e si fondino «in tutto o in parte sulla negazione della Shoah o dei crimini di genocidio, dei crimini contro l'umanità e dei crimini di guerra, come definiti dagli articoli 6, 7 e 8 dello statuto della Corte penale internazionale, ratificato ai sensi della legge 12 luglio 1999, n. 232».

Solo un anno dopo, nel 2017 (l. 167/2017), si è proceduto ad un'ultima modifica del comma 3-bis attraverso l'introduzione, dopo il riferimento alla negazione, della «minimizzazione in modo grave o l'apologia», con il chiaro fine di omologarsi ulteriormente

---

[82] Il testo originario dell'art. 3 della l. in esame puniva: «chi incita in qualsiasi modo alla discriminazione, o incita a commettere o commette atti di violenza o di provocazione alla violenza, ne confronti di persone perché appartenenti a un gruppo nazionale, etnico o razziale». Cfr. paragrafo 2 del presente scritto.

[83] Il nuovo art. 3 della l. 654/1975, come novellato dalla c.d. legge Mancino puniva: «a) con la reclusione fino a tre anni chi diffonde in qualsiasi modo idee fondate sulla superiorità o sull'odio razziale o etnico, ovvero incita a commettere o commette atti di discriminazione per motivi razziale, etnici, nazionali o religiosi; b) con la reclusione da sei mesi a quattro anni che, in qualsiasi modo, incita a commettere o commette violenza o atti di provocazione alla violenza per motivi razziali, etnici, nazionali o religiosi».

[84] L. Goisis*, Crimini d'odio,* cit.13.

[85] Viene punito non più chi «diffonde in qualsiasi modo» ma chi «propaganda idee fondate sulla superiorità o sull'odio razziale o etnico»; non più chi «incita», ma chi «istiga a commettere o commette atti di discriminazione per motivi razziali, etnici, nazionali o religiosi»; nonché chi «istiga» anziché chi «incita» a commettere o commette atti di violenza o atti di provocazione alla violenza per motivi razziali, etnici o religiosi. Si segnala che la Corte di cassazione con la pronuncia n. 34713 del 2016 ha sostenuto che vi è continuità normativa tra il testo anteriore alle modifiche *ex* l. 85/2006 e quello successivo (per un'analisi completa sul punto si segnala il commetto alla citata pronuncia, in *Rivista Penale,* 10, 2016, 895 ss.).

[86] Decisione Quadro del Consiglio del 28 novembre 2008.

alle indicazioni contenute nella citata Decisione Quadro.
L'approdo di tale evoluzione normativa è rappresentato dal d.lgs. 21/2018, il quale ha trasferito la disciplina entro il perimento del Codice penale (titolo XII «dei delitti contro la persona», capo III, «dei delitti contro la libertà individuale», sezione I-bis, dedicata ai «delitti contro l'uguaglianza»[87]), introducendo gli artt. 604-bis c.p., rubricato «propaganda e istigazione a delinquere per motivi di discriminazione razziale etnica e religiosa» e 604-ter c.p., il quale disciplina un'autonoma circostanza aggravante, per i reati commessi per finalità di discriminazione o di odio etnico, nazionale, razziale o religioso, ovvero al fine di agevolare l'attività di organizzazioni, associazioni, movimenti o gruppi che hanno tra i loro scopi le medesime finalità.
Le citate figure delittuose sono comunemente definite crimini d'odio, estrinsecandosi non solo nella loro componente fattiva, intesa come compimento materiale di un atto discriminatorio, bensì anche in quella dialettica ed epistemica, nota come *hate speech*.
L'inserimento nel Codice penale di tali disposizioni[88] deve essere apprezzato in forza del maggior rilievo attribuito ai relativi precetti, i quali mirano a tutelare l'essere umano come individuo, in quanto tale non discriminabile per motivi razziali, etnici, nazionali o religiosi[89]. Appare dunque inevitabile il contrasto che potrebbe verificarsi tra il diritto alla libera manifestazione del pensiero[90] (disciplinato a livello sovranazionale e interno, dagli artt. 10 CEDU e 21 Cost.), il divieto di abuso del diritto (art. 17 CEDU) e la pari dignità dei consociati (art. 3 Cost). E proprio nella tutela della pari dignità sociale deve essere individuato il perno essenziale della normativa in esame.
L'individuazione del bene giuridico tutelato dalle disposizioni sopra richiamate ha assunto, e continua ad assumere, una rilevanza decisiva. Fino a quando queste disposizioni sono state interpretate come finalizzate esclusivamente alla tutela dell'ordine pubblico[91], la giurisprudenza ha incontrato difficoltà nel riconoscere i presupposti per una loro concreta applicazione[92]. Tuttavia, il quadro interpretativo è mutato con

[87] Sui delitti contro l'uguaglianza, nella manualistica, cfr.: F. Bacco, *Norme antidiscriminatorie,* in D. Pulitanò, (a cura di), *Diritto penale. Parte speciale, I, Tutela penale della persona,* Torino, 2019, 403 ss.; a livello monografico, L. Goisis, *Crimini d'odio,* cit., 263 ss.

[88] Si segnala che le disposizioni in oggetto hanno abrogato l'art. 3 della l. 654 del 1975 e l'art. 3 del d.l. 122 del 1993, convertito nella c.d. legge Mancino, pur mantenendone inalterato il testo.

[89] Cfr. M. Pelissero, *Discriminazione, razzismo e diritto penale fragile,* in *Diritto penale e procedura*, 2020, 1017 ss.

[90] Cfr. G. de Vero, *La giurisprudenza della Corte di Strasburgo,* in *Delitti e pene nella giurisprudenza delle Corti europee,* (a cura di) G. de vero, Torino, 2007, 46 ss.

[91] La sentenza spartiacque, da questo punto di vista, è rappresentata dalla n. 341/2001, secondo la quale: «il diritto alla libera manifestazione del pensiero non può dilatarsi sino a comprendere la diffusione di idee ed altre condotte che neghino la personalità e la dignità dell'uomo, valori questi che sono affermati dalla Costituzione come principi fondamentali e non tollerano alcuna forma di gerarchia fondata sull'appartenenza ad un gruppo etnico, nazionalità e razza »; »il diritto alla libera manifestazione del pensiero, tutelato dall'art. 21 della Costituzione, non può essere esteso fino alla giustificazione di atti o comportamenti che, pur estrinsecandosi in una esternazione delle proprie convinzioni, ledano tuttavia altri principi di rilevanza costituzionale ed i valori tutelati dall'ordinamento giuridico interno ed internazionale». In altre parole, per la Corte di cassazione l'esternazione delle convinzioni equivale ad atti e comportamenti che ledono la pari dignità sociale, a prescindere fagli effetti esplicati in concreto (cfr.: i motivi della decisione).

[92] Per una disamina dei beni giuridici in gioco, si veda, P. Tanzarella, *Discriminare parlando. Il pluralismo democratico messo alla prova dai discorsi d'odio razziale,* Torino, 2020, 12; G. Puglisi, *La parola acuminata.*

l'affermarsi della dottrina maggioritaria, che ha visto in esse una tutela plurioffensiva, adottando una lettura personalista e riconoscendo, altresì, la protezione dei beni giuridici dell'uguaglianza e della pari dignità dei cittadini[93].

L'art. 604-bis c.p.[94] comma 1, lettera a), limita la sanzionabilità alla propaganda, all'istigazione a commettere e alla commissione di atti discriminatori. Ai fini della configurazione della condotta di propaganda, secondo il prevalente orientamento dottrinale, non basterebbe una generica diffusione di idee, rendendosi piuttosto necessario che la dichiarazione esternata sia idonea a raccogliere un considerevole numero di consensi in ordine al pensiero manifestato[95].

Le condotte di istigazione a commettere e di commissione di atti discriminatori non richiederebbero, invece, la pubblicità del fatto. L'istigazione può verificarsi nei confronti di una sola persona, in deroga alla disposizione generale di cui all'art. 115 c.p.; d'altro canto, gli atti discriminatori possono essere realizzati anche in privato. Inoltre, non è neppure richiesto che l'atto discriminatorio configuri un illecito penale, essendo sufficiente il suo carattere discriminatorio per motivi razziali o etnici (come per la condotta di propaganda) o nazionali o religiosi. Da questo punto di vista le condotte di propaganda e di istigazione devono necessariamente incidere sull'altrui volontà, ovverosia devono essere concretamente idonee a «provocare l'immediata esecuzione di delitti o, quantomeno, la probabilità che essi vengano commessi in un futuro più o meno prossimo»[96]. Il momento consumativo si individua nell'istante in cui la propaganda e l'istigazione vengono esternate, diventando percepibili e conoscibili, senza che sia necessario che siano anche accolte dai destinatari né che si verifichi un ulteriore evento materiale ed esterno.

Ai fini della configurazione è fondamentale accertare in concreto la pericolosità delle stesse, tenendo in considerazione il contesto in cui si è svolto il fatto, il ruolo dell'agente e le attitudini dei destinatari[97].

Il richiamo alla concreta idoneità della condotta, combinato con l'anticipazione san-

Contributo allo studio dei delitti contro l'eguaglianza, tra aporie strutturali ed alternative alla pena detentiva, in Rivista Italiana di Diritto e Procedura penale, 2018, 1331 ss.

[93] La dignità è stata ben definita come «insieme delle condizioni necessarie a uno sviluppo della persona che le consenta di vivere un'esistenza piena», in M. Caputo, *La menzogna di Auschwitz, le verità del diritto penale. La criminalizzazione del c.d. negazionismo tra ordine pubblico, dignità e senso di umanità,* in AA.VV. *Verità del precetto e della sanzione penale alla prova del processo*, a cura di G. Forti-G. Varraso-M. Caputo, Napoli, 2014, 309; sul tema si veda anche G. Pavich-A. Bonomi, *Reati in tema di discriminazione: il punto sull'evoluzione normativa recente, sui principi e valori in gioco, sulle prospettive legislative e sulla possibilità di interpretare in senso conforme a Costituzione la norma vigente,* in Penale Contemporaneo, 13.10.2014, 12 ss.; L. Scaffardi, *Oltre i confini della libertà di espressione. L'istigazione all'odio razziale,* Padova, 2009, 239.

[94] Art. 604-bis c.p., c. 1: «Salvo che il fatto costituisca più grave reato, è punito: a) con la reclusione fino ad un anno e sei mesi o con la multa fino a 6.000 euro chi propaganda idee fondate sulla superiorità o sull'odio razziale o etnico, ovvero istiga a commettere o commette atti di discriminazione per motivi razziali, etnici, nazionali o religiosi».

[95] F. Bacco, *Norme antidiscriminatorie* cit., 406.

[96] Cass. pen., sez. I, 7 ottobre 2009, n. 40552, in *Dejure*.

[97] In tal senso, giurisprudenza consolidata afferma che l'odio razziale o etnico è integrato da un sentimento idoneo a determinare il concreto pericolo di comportamenti discriminatori, e non da qualsiasi sentimento di generica antipatia, insofferenza o rifiuto riconducibile a motivazioni attinenti alla razza, alla nazionalità o alla religione, cfr: Cass. pen., sez. V, 22 luglio 2019, n. 32862.

zionatoria caratteristica delle fattispecie analizzate, le quali non richiedono l'effettiva attuazione dell'istigazione o della propaganda, consente di inquadrare il reato tra quelli a pericolo concreto[98]. Tale orientamento, del resto, consente di evitare il rischio di un possibile conflitto con il principio di offensività, il quale potrebbe essere compromesso da una sanzione penale nei confronti di un soggetto autore di condotta propagandistica concretamente inidonea, per i mezzi, le modalità e i destinatari, a ledere il bene giuridico tutelato. In un campo, come quello che ci occupa, in cui il bilanciamento dei beni giuridici in gioco è così delicato, inquadrare il reato nel novero di quelli a pericolo concreto permette di garantire il rispetto della Costituzione[99]. È dunque necessario svolgere un accertamento in concreto sulla pericolosità della propaganda e dell'istigazione, tenendo conto del contesto, del ruolo dell'agente e delle attitudini dei destinatari[100].

La ricostruzione in termini di pericolo concreto è stata accolta anche nella risalente (seppur estremamente significativa) pronuncia della Corte costituzionale in materia di reati di opinione[101] oltreché, in modo analogo, anche dalla Corte EDU nelle sue numerose sentenze in materia di negazionismo[102].

Con riferimento alla previsione di cui alla lett. b) [103] dell'art. 604-bis c.p., comma 1[104], essa concerne l'istigazione a commettere o la commissione di atti di violenza o di pro-

[98] In tal senso A. Cadoppi-S. Canestrari-A. Manna-M. Papa, *Diritto penale. Tomo terzo. I delitti contro l'onore e la libertà individuale,* Vicenza, 2022, 6105 ss.; L. Goisis, *Crimini d'odio,* cit., 283, dove parla di «reati di pericolo concreto (implicito)»; *contra* A. Galluccio, *Punire la parola pericolosa?*, Milano, 2020, 416.

[99] R. Bartoli-M. Pelissero-S. Seminara, *Diritto penale,* cit., 26.

[100] Cfr.: Cass. pen., sez. V, 29 gennaio 2020, n. 3722, in *Dejure*; Cass. pen., sez. I, 16 gennaio 2020, n. 1602, in *Dejure*; Cass. pen., sez. V, 30 luglio 2019, n. 34815, in *Dejure*; Cass. pen., sez. V, 22 luglio 2019, n. 32862, in *Dejure*.

[101] Corte cost., 4 maggio, 1970, nella quale si afferma che l'apologia punibile ai sensi dell'art. 414, ultimo comma c.p., non sarebbe la manifestazione di pensiero pura e semplice, bensì quella che per le sue modalità integra un comportamento concretamente idoneo a provocare la commissione di delitti. Invero, la libertà di manifestazione del pensiero, garantita dall'art. 21, primo comma della Costituzione troverebbe i suoi limiti non soltanto nella tutela del buon costume, ma anche nella necessità di proteggere altri beni di rilievo costituzionale e nell'esigenza di prevenire e far cessare turbamenti della sicurezza pubblica, la cui tutela costituisce una finalità immanente del sistema.

[102] Si segnalano le sentenze della Corte EDU nelle quali si è fatta applicazione di questi due criteri: *Perincekin* (2015), un caso di negazionismo del genocidio armeno, giudicato non concretamente pericoloso. La Corte ha, in tale occasione, ribadito che le limitazioni della libertà di espressione sono giustificate se le dichiarazioni di odio si innestano in un clima di tensione politica e sociale e se le dichiarazioni incitano in modo diretto o indiretto alla violenza o giustificano la violenza, l'odio o l'intolleranza. Si segnala altresì, con esiti opposti, *Dieudonnè* (2015), dove uno spettacolo pubblico, camuffato da manifestazione artistica, ha integrato una presa di posizione d'odio antisemita, capace di integrare l'abuso del diritto di manifestazione del pensiero *ex* art. 17 CEDU. Per un ulteriore approfondimento di tali pronunce si veda anche G. Forti, *Le tinte forti del dissenso nel tempo dell'ipercomunicazione pulviscolare. Quale compito per il diritto penale?* in *Rivista Italiana di Diritto e Procedura Penale*, 4, 2018.

[103] Art. 604 bis, comma 1, lett. b): «con la reclusione da sei mesi a quattro anni chi, in qualsiasi modo, istiga a commettere o commette violenza o atti di provocazione alla violenza per motivi razziali, etnici, nazionali o religiosi».

[104] Per ordine di completezza si segnala che il comma 2 dell'articolo in esame viete le organizzazioni, le associazioni, i movimenti o i gruppi aventi tra i propri scopi l'incitamento alla discriminazione o alla violenza per motivi razziali, etnici, nazionali o religiosi. In seno a questa condotta associativa, parte della dottrina ravvisa un residuo di tutela del bene giuridico dell'ordine pubblico (L. Goisis, *Crimini d'odio,* cit. 285).

vocazione alla violenza per scopi discriminatori. La violenza può essere esercitata sia direttamente sulle persone, che anche sulle cose dalle stesse possedute.
I motivi razziali, etnici, nazionali o religiosi devono emergere in maniera univoca dalle idee, mentre risulta essere privo di rilievo l'eventuale, diverso, intento del soggetto attivo. A differenza di quanto previsto per le condotte di cui alla lett. a), il riferimento agli atti violenti sembra integrare, di per sé, un comportamento intrinsecamente illecito.
Una questione particolarmente dibattuta tra dottrina e giurisprudenza sembra riguardare l'elemento soggettivo. La giurisprudenza prevalente ritiene che la propaganda di idee discriminatorie e l'istigazione alla commissione di atti discriminatori (prima parte, lett. a) integrerebbero un'ipotesi di dolo generico; diversamente, la commissione di atti di discriminazione (seconda parte, lett. a) e la commissione di violenza o di atti di provocazione alla violenza sarebbero sorrette dal dolo specifico, in quanto «in tali ultime ipotesi il motivo ispiratore eccede la condotta discriminatoria o violenta, mentre nel caso della propaganda e dell'istigazione tale motivo è incluso nelle idee propagandate o negli atti discriminatori istigati»[105]. La dottrina, d'altro canto, ritiene che entrambe le condotte (lettere a e b) debbano essere punite a titolo di dolo generico[106].
Un ultimo aspetto che merita di essere esaminato è quello concernente il trattamento sanzionatorio[107].
Rispetto alle condotte disciplinate alla lettera a) dell'art. 604-bis c.p., appare incongruo omologare condotte dotate di un differente disvalore, ponendo sul medesimo piano la propaganda, l'istigazione e la materiale realizzazione di atti discriminatori (reclusione fino a un anno e sei mesi o con la multa). Peraltro, l'alternatività della pena indebolisce l'efficacia del precetto, ammettendo una rilevanza, per così dire, "bagatellare", di fatti pur offensivi di beni di rango così elevato. Con riferimento alla graduazione della sanzione prevista alle lettere a) e b) dell'art. 604-bis c.p., essa si considera proporzionata: la propaganda e l'istigazione (lettera a) sono punite meno gravemente (fino a un anno e sei mesi di reclusione o con la multa fino a 6.000 euro) rispetto all'istigazione e agli atti violenti (reclusione da sei mesi a quattro anni) disciplinati alla lettera b).
Un discorso diverso deve essere fatto per la fattispecie di cui all'art. 604-ter c.p., ossia l'aggravante di natura teleologica[108]. Nello specifico, essa prevede un aumento di pena sino alla metà per i reati puniti con la pena diversa da quella dell'ergastolo, quando sono commessi per finalità di discriminazione o di odio etnico, nazionale, razziale o religioso, ovvero al fine di agevolare organizzazioni aventi i medesimi scopi.

---

[105] Cass. pen., sez. III, 7 maggio 2008, n. 37581, in *Dejure*.

[106] Cfr.: G. Fiandaca – E. Musco, *Diritto Penale. Parte speciale,* vol. II, tomo I, Bologna, 2024, 243; L Picotti, cit., 145; G. Forti-S. Riondato-S. Seminara, *Commentario breve al Codice Penale,* VII ed., Milano, 2024, 2381.

[107] La delicatezza della materia oggetto d'esame e la tensione che questa può comportare con il diritto alla libera manifestazione del pensiero impongono una puntuale attenzione al momento sanzionatorio. Tale necessità è stata peraltro avvertita dalla Corte europee dei diritti dell'uomo, la quale ha avuto sin da subito la necessità di specificare che ogni sanzione imposta in questo ambito deve essere proporzionata allo scopo legittimo perseguito (cfr.: Corte EDU, *Handyside c. UK,* par. 49,18 del 7 dicembre 1976).

[108] L'aspetto peculiare risiede nella sottrazione di siffatta circostanza aggravante al sistema di bilanciamento di circostanze di cui all'art. 69 c.p., eccetto che nel caso di attenuante per la minore età di cui all'art. 98 c.p. Questa impostazione è eloquente del notevole disvalore che connota tali condotte, tanto da imporre un trattamento punitivo differenziato e meno favorevole per il reo.

La finalità di discriminazione o di odio non deve essere rigidamente intesa come movente psicologico: essa sussiste sia quando l'azione concreta è finalizzata a rendere percepibile all'esterno la sua causale, sia quando evidenzia un pregiudizio di tipo etnico, nazionale, razziale o religioso. Non sembra inoltre essere necessario uno specifico scopo di incitamento[109].

Il fondamento di tale disposizione non dà luogo a problemi interpretativi quando il reato base presenta una specifica e autonoma direzione lesiva, alla quale si aggiungono le finalità tipizzate dalla norma.

Problemi di diversa natura sorgono con particolare riferimento al reato di diffamazione, in quanto il confine tra gli artt. 604-bis e 595 c.p. nella sua forma aggravata è assai labile: le condotte di propaganda e istigazione recano solitamente in sé una nota diffamatoria nei confronti dei soggetti appartenenti alle razze, etnie, nazioni o religioni.

Da diversi anni pendono in Parlamento proposte di legge tese a estendere le incriminazioni contenute negli artt. 604-bis e ter c.p. alle discriminazioni fondate su motivi sessuali. Il contrasto vede, da un lato, coloro i quali si oppongono al sacrificio della libertà di manifestazione del pensiero; da altro lato, quelli che ravvisano nella negazione di un riconoscimento all'individuo diverso, un'offesa intollerabile di un diritto inviolabile proiettato sull'eguaglianza. Rispetto a questa contrapposizione, è difficile discostarsi dal contenuto della pronuncia della Corte EDU *Gunduz v. Turchia*, con la quale si è affermato che: «la tolleranza e il rispetto per l'eguale dignità di tutti gli esseri umani costituiscono il fondamento di una società democratica e pluralistica. Ne deriva che nelle società democratiche può ritenersi necessario sanzionare, e così prevenire, ogni forma di espressione che diffonda, istighi, promuova o giustifichi l'odio basato sull'intolleranza»[110]. Si segnala che l'Italia risulta essere uno dei pochi paesi in Europa che non dispone di specifiche incriminazioni dei reati d'odio in ambiti legati alla sfera sessuale, in palese contrasto con gli atti sovranazionali[111].

---

[109] Cass. pen., sez. V, 22 luglio 2019, n. 32862, in *Dejure,* per quanto concerne i riferimenti alla giurisprudenza della Corte EDU. Si esclude, ai fini della configurazione, l'esigenza che l'azione risulti intenzionalmente diretta a dare luogo, in futuro o nell'immediato, al concreto pericolo di comportamento discriminatorio.

[110] Corte EDU, sez. I, 4 dicembre 2003, n. 35071/97.

[111] Si segnala, per completezza, che il disegno di legge noto come d.d.l Zan, recante «Misure di prevenzione e contrasto della discriminazione e della violenza per motivi fondati sul sesso, sul genere, sull'orientamento sessuale, sull'identità di genere e sulla disabilità», naufragato in Senato, perseguiva l'obiettivo di modificare l'attuale assetto dei motivi discriminatori delle norme esaminate aggiungendo dopo le parole «per motivi razziali, etnici, nazionali o religiosi» l'inciso «oppure fondati sul sesso, sul genere, sull'orientamento sessuale, sull'identità di genere o sulla disabilità». Tale estensione avrebbe dovuto essere riferita esclusivamente alle condotte di istigazione al compimento o al compimento di atti discriminatori (art. 604-bis c.p. comma 1 lett. b), nonché alla condotta associativa di cui al secondo comma dell'art. 604-bis c.p. ed alla circostanza aggravante di cui all'art. 604-ter c.p..

Con riferimento agli atti sovranazionali, si segnalano: il Trattato sul funzionamento dell'Unione Europea che stabilisce che l'Unione mira in tutte le sue attività a combattere la discriminazione sulla base di diversi fattori, incluso l'orientamento sessuale; l'art. 21 della Carta dei Diritti Fondamentali; la direttiva 2000/78/CE, che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro, ribadendo il diritto universale di tutti all'uguaglianza davanti alla legge e alla protezione contro le discriminazioni quale diritto universale; la Raccomandazione Rec CM2010(05) del Comitato dei Ministri del Consiglio d'Europa agli Stati membri sulle misure dirette a combattere la discriminazione fondata sull'orientamento sessuale o l'identità di genere (adottata il 31 marzo 2010); le Linee guida per la promozione e la tutela dell'esercizio di tutti i diritti umani da parte di lesbiche, gay,

Le problematiche relative ai discorsi d'odio online vengono affrontate principalmente nelle aule di giustizia, a causa della difficoltà, del legislatore italiano e di quello europeo, di rispondere in maniera adeguata e repentina alle sfide poste dall'incessante evoluzione tecnologica[112]. Di conseguenza, la risposta di un singolo tribunale a un caso specifico tende a configurarsi come un riferimento generale per l'intera materia, contribuendo alla sua definizione normativa. Secondo i dati diffusi dall'allora ministro della Giustizia Marta Cartabia nel 2022[113], l'introduzione di queste fattispecie penali non avrebbe assolto all'auspicata funzione di deterrenza.

Nonostante un esponenziale aumento dei discorsi d'odio[114], si è riscontrato un numero esiguo di procedimenti giudiziari. Tra il 2016 e il primo semestre del 2021, i procedimenti iscritti (prevalentemente nel nord Italia, con le percentuali maggiori nelle grandi di città di Roma con il 12,62%, e Milano con il 4,85%) non hanno superato le 300 unità, sia con riferimento alla forma di manifestazione della propaganda e dell'istigazione, quanto alla configurazione dell'aggravante, la cui applicazione risulta più diffusa, a discapito del delitto di cui all'art. 604-bis c.p.[115].

Rispetto ai flussi dei procedimenti definiti dalle sezioni G.i.p./ G.u.p. e dibattimentali, si è potuto osservare che nell'80 % dei casi l'iscrizione delle ipotesi di reato viene poi archiviata; nei casi di rinvio a giudizio, relativamente scarsi, prevale la condanna (circa il 40%). Solo la metà dei procedimenti penali riguardanti ipotesi delittuose aggravate ai sensi dell'art. 604-ter c.p si conclude con una condanna aggravata.

Si segnala, inoltre, che in alcuni dei casi in cui il rinvio a giudizio avviene ai sensi dell'art. 595 c.p., aggravato *ex* art. 604-ter c.p., la configurazione del reato viene successivamente esclusa a causa dell'impossibilità di individuare la persona destinataria dell'offesa[116],

bisessuali, transgender e intersessuali, adottati dal Consiglio europeo il 24 giugno 2013; la Risoluzione del 14 febbraio 2019 adottata dal Parlamento europeo, con il quale si invita la Commissione a garantire priorità ai diritti delle persone LGBT+; la Risoluzione del 1 marzo 2021 del Parlamento europeo.

[112] Si ritiene opportuno segnalare che negli anni Novanta le sentenze penali aventi ad oggetto il reato di *hate speech* sono state piuttosto rare. Nello specifico, la lett. a), art. 3 legge Reale, così come modificato dalla c.d. legge Mancino, non ha mai trovato applicazione ai fini della valutazione del reato di diffusione di idee fondate sulla superiorità o sull'odio razziale o etnico. In alcune (rare) circostanze ha trovato invece applicazione il reato di incitamento a commettere atti di discriminazione, sia con riferimento alla violazione dell'art. 3 lett. a), sia in relazione all'aggravante *ex* art. 3 della legge Mancino. Interessante, al riguardo, è osservare come inizialmente i Giudici abbiano inteso interpretare tali fattispecie delittuose alla luce del criterio del pericolo concreto, sulla scorta di un precedente che aveva dimostrato riluttanza a offrire una tutela anticipata per la mera diffusione di frasi solo ipoteticamente foriere di atti violenti (cfr.: Cass. pen., sez. III, 24 novembre 1998, n. 434.

[113] Dati forniti dall'allora Ministra Marta Cartabia, in seno alla Commissione straordinaria per il contrasto dei fenomeni di intolleranza, razzismo, antisemitismo e istigazione all'odio e alla violenza, presieduta dalla Sen. Segre, l'8 febbraio 2022.

[114] Il Barometro dell'odio 2023-2024 di Amnesty International, rivela un aumento significativo dei contenuti problematici o di *hate speech*: dal 2019 a luglio 2024 il tasso di discorsi offensivi, discriminatori o che incitano all'odio è passato dal 10 per cento al 15 per cento. In particolare, i contenuti che incitano alla discriminazione e alla violenza sono triplicati, superando il tre per cento del *corpus* analizzato. Nell'ambito della ricerca è stato inoltre condotto un sondaggio, in collaborazione con Ipsos, per rilevare quale fosse la posizione dell'opinione pubblica rispetto all'attivismo e alle varie forme di protesta. È emerso che il 48 per cento delle persone intervistate vede le manifestazioni come un passatempo o una moda, mentre il 17 per cento non crede che tutti, in Italia, dovrebbero avere il diritto di manifestare.

[115] Dati forniti dall'allora Ministra Marta Cartabia nel 2022.

[116] Cfr.: Cass. pen., sez. V, 22 luglio 2019, n. 32862, in *Dejure*.

con la conseguente disapplicazione della circostanza aggravante in esame. Un'ulteriore ipotesi problematica riguarda il caso di rinvio a giudizio per l'ipotesi di reato di cui all'art. 604-bis c.p., successivamente riqualificato in diffamazione aggravata. Il rischio, in questo caso, è rappresentato dal proscioglimento per difetto di querela[117].
L'analisi svolta fino a questo momento evidenzia come lo strumento del diritto penale possa rivelarsi utile, in quanto idoneo a stigmatizzare comportamenti che ledono beni giuridici fondamentali per il funzionamento della democrazia. Tuttavia, stante le criticità emerse nel corso della trattazione, non sembra essere sufficiente a contrastare in modo efficace la diffusione dell'odio online.

4. Esperienze giuridiche di incriminazione dell'*online hate speech* a confronto: Francia, Germania e Spagna

Volendo ora rivolgere l'attenzione ad altri sistemi giuridici che hanno operato scelte di politica criminale analoghe a quelle effettuate dal legislatore italiano, merita di essere menzionato l'ordinamento francese, che si contraddistingue per la previsione di un complesso sistema di contrasto all'odio online articolato su tre livelli: i contenuti vietati, i contenitori - ossia i soggetti che ospitano tali contenuti - e i controllori esterni (magistratura, autorità amministrativa, autorità indipendente, ecc.). In Francia, Internet non è un'area esente dalla legge[118].
Più nel dettaglio, l'odio online viene punito ricorrendo alla sussunzione sia sotto i reati previsti dal Codice penale, come l'apologia di terrorismo o di minaccia di morte, sia sotto i reati previsti dalla legge sulla libertà di stampa del 29 luglio 1881[119], che sanziona gli attacchi alla presunzione di innocenza, alla dignità delle vittime, il negazionismo, l'insulto, la diffamazione e la provocazione, in modo graduale a seconda della loro portata più o meno pubblica o del loro carattere discriminatorio, nonché l'apologia di alcuni reati[120].
La suddetta legge punisce l'espressione di qualsivoglia commento che possa assumere una dimensione pubblicistica, vale a dire "penetrare in un modo o nell'altro nel pubblico dominio", e di conseguenza si applica anche alle manifestazioni pubblicate su Internet[121].
Tuttavia, è importante ricordare che la legge sulla stampa fu adottata per proteggere la

[117] Dati forniti dall'allora Ministra Marta Cartabia nel 2022 *Norme antidiscriminatorie*.

[118] L. Marguet, *La complexité du dispositif juridique de lutte contre la haine en ligne*, in *Les sciences sociales et les défis éthiques de la recherche*, 16, 2023, 1-5.

[119] *Loi du 29 juillet 1881 sur la liberté de la presse.*

[120] Vedasi riguardo a detti reati gli artt. 35 ter, 35 quater, 24 bis, 33, 29, 23 e 24 della l. sulla stampa del 1881, citati da L. Marguet, *La complexité du dispositif juridique de lutte contre la haine en ligne*, cit., 10, la quale rimanda, sul punto, per un maggiore approfondimento a L. Marguet, *La répression de la provocation, de la diffamation et des injures non publiques représentant un caractère raciste ou discriminatoire en France*, in *Reveu des droits et libertes fondamentaux*, 20, 2019.

[121] Art. 23 della legge sulla libertà di stampa del 1881, modificata dalla l. 2004-575 del 21 giugno 2004 per la fiducia nell'economia digitale (c.d. "LCEN"), citata da L. Marguet, *La complexité du dispositif juridique de lutte contre la haine en ligne*, cit., 16.

libertà di espressione. Ne consegue che, quando un discorso d'odio rientra nel suo ambito di applicazione, all'autore vengono garantite una serie di tutele procedurali, quali un termine di prescrizione molto breve (di tre mesi), l'inammissibilità della custodia cautelare, della comparizione con rito immediato e del pronunciamento anticipato di colpevolezza. Inoltre, il giudice istruttore e il tribunale di primo grado sono vincolati alla qualificazione dei fatti come descritti negli atti che hanno dato origine all'azione penale[122].

Questo quadro giuridico è attualmente in evoluzione. Alcune iniziative legislative mirano a escludere talune condotte dal campo di operatività della legge sulla stampa. Tra queste, la l. 1109 del 24 agosto 2021, la cosiddetta "legge di rafforzamento dei principi della Repubblica", ha introdotto una serie di disposizioni relative alla lotta contro i discorsi d'odio e i contenuti illeciti nella sfera digitale[123].

La principale novità di interesse è costituita dall'introduzione, ad opera dell'art. 36, di una nuova fattispecie criminosa all'art. 233-1-1 del *Code pénal*, rubricata "messa in pericolo della vita altrui tramite la diffusione di informazioni relative alla vita privata, familiare o professionale"[124]. Tale disposizione criminalizza il *doxing*[125], ossia la rivelazione di informazioni relative alla vita privata, familiare o personale di una persona che consentano di identificarla o di localizzarla, allo scopo di esporla a un rischio diretto di danno, nella misura in cui il responsabile della rivelazione non avrebbe potuto ignorare il verificarsi di un pregiudizio alla persona medesima o a una cosa[126].

La figura incriminatrice è stata concepita facendo ricorso alla forma di imputazione

[122] Artt. 50 e 51 della l. sulla libertà di stampa del 1881. In riferimento ai profili procedurali, si rinvia a R. Evan, *La procédure pénale en droit de la presse*, Gazette du Palais, 2019.

[123] *Loi n° 2021-1109 du 24 août 2021 confortant le respect des principes de la République*, entrata in vigore a seguito dell'attentato a Samuel Paty, decapitato all'uscita della scuola dove insegnava, per aver mostrato immagini di Maometto, intesa principalmente a contrastare i c.d. "separatismi religiosi", in particolare quelli di matrice islamista, che secondo il legislatore francese – e alla luce degli attentati che la Francia ha dovuto sopportare nell'ultimo decennio – rappresentano una minaccia per i valori della *République*. È stata così introdotta una normativa restrittiva che, secondo alcuni commentatori e gli stessi vescovi francesi, pone forti limiti all'esercizio della libertà religiosa, così E. Tawil, *La recente legge francese sulla laicità dello Stato contro i separatismi religiosi*, intervento reso durante una lezione di Diritto pubblico presso l'Université Paris II Panthéon-Assas, lumsa.it, 21 novembre 2022.

[124] La formulazione dell'art. 233-3-1 del Codice penale francese è la seguente: «L'atto di rivelare, diffondere o trasmettere, con qualsiasi mezzo, informazioni relative alla vita privata, familiare o professionale di una persona che ne permettano l'identificazione o la localizzazione al fine di esporre tale persona o i membri della sua famiglia a un rischio diretto di danno alla loro persona o ai loro beni di cui l'autore non poteva non essere a conoscenza, è punito con tre anni di reclusione e una multa di 45.000 euro. Quando i fatti sono commessi a danno di un pubblico ufficiale, di una persona incaricata di un pubblico servizio o che ricopre una carica elettiva pubblica o di un giornalista, ai sensi c. 2 dell'art. 2 della l. del 29 luglio 1881 sulla libertà di stampa, le pene sono aumentate a cinque anni di reclusione e a 75.000 euro di multa. Quando i fatti sono commessi a danno di un minore, le pene sono aumentate a cinque anni di reclusione e 75.000 euro di multa. Quando i fatti sono commessi a danno di una persona la cui particolare vulnerabilità, dovuta all'età, alla malattia, all'infermità, alla disabilità fisica o mentale o alla gravidanza, è evidente o nota all'autore del reato, le pene sono aumentate a cinque anni di reclusione e 75.000 euro di multa. Nel caso in cui i reati siano commessi attraverso la stampa scritta o audiovisiva o attraverso la comunicazione pubblica online, si applicano le disposizioni specifiche delle leggi che regolano queste materie per quanto riguarda la determinazione della responsabilità».

[125] L. Marguet, *La complexité du dispositif juridique de lutte contre la haine en ligne*, cit., 17.

[126] P. Januel, *Séparatisme: les principales dispositions de la loi*, in dalloz-actualite.fr, 07 settembre 2021.

soggettiva della responsabilità introdotta dal Codice penale del 1994 all'art. 121, cc. 2 e 3, che si colloca in una posizione intermedia tra i concetti che, nella scienza penalistica italiana, assumono la qualifica di "dolo eventuale" e "colpa cosciente". Ai fini della configurazione della *mise en danger*, è necessario che l'autore abbia coscienza del pericolo per l'interesse tutelato, presunto dalle modalità di aggressione[127]. È evidente che il legislatore francese ha attribuito rilevanza al solo stato di pericolo dell'oggettività giuridica tutelata oltre che alle caratteristiche con cui si esplica il comportamento, anticipando la soglia della rilevanza penale della condotta.

Questa soluzione si inserisce nel solco della giurisprudenza costituzionale che ha posto l'attenzione non soltanto sull'importanza del mantenimento dell'ordine pubblico e dei diritti dei destinatari delle manifestazioni d'odio, ma anche sui canoni di proporzionalità e necessità che devono guidare l'accertamento concreto dell'illiceità del fatto. A tale proposito, numerose pronunce del Consiglio costituzionale francese hanno affrontato il tema dell'impatto delle norme penali sulla libertà di espressione, definita come «una condizione della democrazia e una delle garanzie del rispetto di altri diritti e libertà»[128]. In linea generale, il Consiglio non ha sollevato obiezioni nei confronti delle scelte del legislatore di incriminare condotte di abuso della libertà di cui si tratta, a condizione che le stesse, in ossequio all'art. 34 Cost., compromettano «l'ordine pubblico e i diritti di terzi»[129] e superino positivamente un triplice vaglio, che ruota intorno ai canoni di necessità, adeguatezza e proporzionalità [130]. Tale valutazione viene operata mediante un approccio di tipo casistico e avendo particolare riguardo alla certezza dell'illiceità del messaggio o, al contrario, all'incertezza di tale illiceità. Ne deriva che il rischio che la violazione possa essere qualificata come proporzionata è tanto maggiore quanto più la qualificazione giuridica dei comportamenti è suscettibile di dare luogo a interpretazioni differenti o contrastanti[131].

Le successive disposizioni della l. 1109 del 24 agosto 2021 (sino all'art. 48), attraverso la combinazione di nuove previsioni normative e l'estensione ai gestori di siti web delle responsabilità già sancite dalle leggi sulla stampa, tentano di disciplinare il tema spinoso della diffusione dei messaggi d'odio, d'incitamento alla violenza o d'induzione alla commissione di reati, in tutti gli aspetti, talora sfuggenti, che concorrono a delineare il quadro sociologico dell'odierno radicalismo politico di ispirazione islamica affermatosi in Francia[132].

In quest'ottica, l'ordinamento francese punisce anche le molestie informatiche: l'art. 222-33-2-2 del Codice penale precisa che il reato di molestie si configura anche quando

[127] G. Fornasari - A. Menghini, *Percorsi europei di diritto penale*, Cedam, 2005, 84.

[128] Consiglio cost. francese, 10 giugno 2009, n. 580 DC.

[129] Art. 34 Cost. francese. Per un commento sul punto, vedasi Consiglio cost. francese, d28 febbraio 2012, n. 2012-647 DC, par. 5, citata in *Commento alla decisione 2020-801 del 18 giugno 2020*, cit., 18.

[130] Consiglio cost. francese, 20 maggio 2011, n. 2011-131 QPC, Consiglio cost. francese, 15 dicembre 2017, n. 2017-682 QPC, par. 4; Consiglio cost. francese, 8 gennaio 2016, n. 2015-512 QPC par. 5-8, citate in *Commento alla decisione 2020-801 del 18 giugno 2020*, cit.

[131] *Ibid.*, come risulta dal *Commento alla decisione 2020-801 del 18 giugno 2020*, cit., 10.

[132] Così, A. Tira, *La legge francese n. 1109 del 24 agosto 2021 sul "rafforzamento del rispetto dei principi della Repubblica"*, in *Stato, Chiese e Pluralismo confessionale*, 14, 2021, 25. Si veda *ex multis*, per questo delicato profilo, il saggio di F. Benslama, *Un furieux désir de sacrifice. Le surmusulman*, in *Media Diffusion*, 2016.

gli atti sono stati commessi attraverso un servizio di comunicazione online (*scapel*[133]). Dal 2018 il reato si configura anche «quando tali parole o comportamenti vengono imposti alla stessa vittima da più persone, in modo concertato o su istigazione di una di esse, anche se ciascuna di queste persone ha agito ripetutamente»[134]. Tale precisazione mira a sanzionare le c.d. "molestie da branco".

La densità del sistema disegnato dal legislatore francese, pone, tuttavia, numerosi interrogativi: la quantità di norme applicabili non è sempre garanzia di qualità o di efficacia, bensì, al contrario, può essere sinonimo di complessità, che ostacola l'intelligibilità del sistema[135].

Analogamente a quanto avvenuto in Francia, anche l'ordinamento giuridico tedesco ha introdotto una legislazione finalizzata a contrastare, in maniera quanto più complessiva e sistemica, l'incitamento all'odio online e la commissione di reati sui social network, con la *Network Enforcement Law* (c.d. NetzDG)[136] (c.d. NetzDG), entrata in vigore il 1° ottobre 2017 e successivamente riformata dalla l. del 3 giugno 2021.

È bene premettere che la Costituzione tedesca[137], all'art. 5, par. 1, tutela il diritto individuale di «esprimere e diffondere [...] opinioni con parole, scritti e immagini». Il concetto di opinione deve essere inteso in senso ampio e non può essere limitato, a meno che l'espressione non si configuri come una "critica abusiva", concetto che segna la sottile linea di demarcazione tra l'opinione critica (che merita protezione) e la condotta illecita[138].

Il libero esercizio del diritto di opinione non è soggetto a riserve di natura contenutistica: il suo ambito di applicazione è esteso a qualunque contenuto[139] indipendentemente dalla sua forma di manifestazione[140].

A tutt'oggi, l'ordinamento tedesco non stabilisce una definizione di discorso d'odio[141]: neppure la *NetzDG* ha introdotto una definizione giuridica univoca. Ciononostante, il § 1, c. 3, della *NetzDG* fornisce un'importante direttiva ermeneutica per attribuire rilevanza penale a queste condotte: il legislatore ha operato un rinvio al reato di cui al § 130 del Codice penale tedesco (*Strafgesetzbuch*, in forma abbreviata *StGB*), il cui c. 1 punisce l'"istigazione all'odio" nei confronti di un gruppo nazionale, razziale, religioso

133 Cass. francese, 10 marzo 2020, n. 19-81026; Cass. francese, 17 settembre 2019, n. 18-834.

134 L. 703 del 3 agosto 2018 sul rafforzamento della lotta alla violenza sessuale e di genere.

135 L. Marguet, *La complexité du dispositif juridique de lutte contre la haine en ligne*, cit., 5.

136 L. del 03 giugno 2021 *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* (Legge per migliorare l'applicazione della legge nei social network) che prevede alcuni obblighi per i social network, tra i quali la rimozione entro 24 ore dei contenuti illegali diffusi, come quelli che incitano all'odio o alla diffamazione. Per un maggiore approfondimento, si rinvia a O. Pollicino – G. De Gregorio, *Hate speech: una prospettiva di diritto comparato*, in *Giornale di diritto amministrativo*, 4, 01 luglio 2019, 433.

137 Più propriamente la Legge fondamentale della Repubblica federale di Germania, c.d. *Grundgesetz*.

138 V. Claussen, *Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation*, in *Media Laws*, 3, 3, 2018, 121.

139 Per un maggior approfondimento vedasi M. Lothar, *Die wehrhafte Demokratie als verfassungsimmanente Schranke der Meinungsfreiheit,* in *Zeitschrift für das Juristische Studium*, 2, 2010, 155.

140 A. Espinoza – J. Rivas Alberty, *Limitaciones Inmanentes de la libertad de expresiòn. El discurso de odio en Alemania, España y el TEDH*, in *Revista de Direito Pùblico Contemporaneo*, 2, 2020, 224-225.

141 J. Rinceanu, *Verso una forma di polizia privata nello spazio digitale? L'inedito ruolo dei provider nella disciplina tedesca dei social network*, in *Sistema penale, Studi in onore di Antonio Fiorella: Volume 1. Tre press*, 2021, 4.

o di un gruppo determinato dall'origine etnica, contro parti della popolazione o contro un individuo, a causa della sua appartenenza a uno dei suddetti gruppi o a una parte della popolazione[142]. Dunque, il discorso d'odio diventa illecito e non costituzionalmente tutelato, ogni qualvolta esso raggiunga l'intensità dell'istigazione[143]. La soglia della punibilità individuata è paragonabile a quella prevista dall'art. 604-bis del Codice penale italiano[144].

Ferma restando l'applicazione della disposizione sopra esaminata, anche altre fattispecie vengono attualmente impiegate per perseguire i discorsi d'odio. È questo il caso dell'ingiuria (§ 185 *StGB*), della diffamazione (§ 186 *StGB*), della menzogna diffamatoria (§ 187 *StGB*), della diffamazione e menzogna diffamatoria contro personaggi della vita politica (§ 188 *StGB*) e dell'incitamento pubblico a commettere reati (§ 111 *StGB*)[145].

Al fine di garantire il principio della certezza del diritto e della legalità in materia penale, il legislatore tedesco ha tassativamente elencato i contenuti da ritenersi illeciti al § 1, par. 3, della NetzDG, positivizzando un vero e proprio catalogo dei discorsi d'odio penalmente rilevanti[146]: § 86 StGB (Diffusione di mezzi di propaganda di organizzazioni incostituzionali), § 86a StGB (Uso di contrassegni di organizzazioni incostituzionali), § 89 StGB (Preparazione di un grave atto di violenza che mette in pericolo lo Stato), § 91 StGB (Istruzioni per commettere un grave atto di violenza che mette in pericolo lo Stato), § 100a StGB (Falsità a scopo di tradimento della Patria), § 111 StGB (Pubblico incitamento a commettere reati), § 126 StGB (Turbativa della pace pubblica mediante la minaccia di reati), §§ 129-129b StGB (Formazione di associazioni criminali, Formazione di associazioni terroristiche, Associazioni criminali e terroristiche all'estero), § 130 StGB (Istigazione all'odio), § 131 StGB (Apologia della violenza), § 140 StGB (Remunerazione ed apologia di reati), § 166 StGB (Vilipendio di culti, comunità religiose e associazioni ideologiche), § 184b StGB i.V.m. § 184d StGB (Distribuzione, acquisizione e possesso di scritti pedopornografici in relazione alla messa a disposizione di contenuti pornografici per mezzo di trasmissioni radiotelevisive o telematiche), §§ 185-187 StGB (Ingiuria, Diffamazione, Menzogna diffamatoria), § 201a StGB (Violazione della sfera intima e dei diritti della personalità attraverso ripresa di immagini), § 241 StGB (Minaccia) o § 269 StGB (Falsificazione di dati probatori)[147].

Quanto al formante giurisprudenziale, la Corte costituzionale federale tedesca (*Bundesverfassungsgericht*, per brevità BVerfG) ha precisato che in tutte le ipotesi riconducibili al § 130 si ha una violazione della dignità umana, che costituisce il nucleo essenziale di ciascun diritto fondamentale e non può essere oggetto di bilanciamento con nessuno di essi. Pertanto, ogni qualvolta la dignità umana venga lesa, non potranno essere presi

142 § 130, c. 1, *StGB*.

143 J. Rinceanu, *Verso una forma di polizia privata nello spazio digitale?,* cit., 4.

144 Rubricata "*Propaganda e istigazione a delinquere per motivi di discriminazione razziale etnica o religiosa*" per il cui approfondimento si rimanda al par. 3 del presente contributo.

145 J. Rinceanu, *Verso una forma di polizia privata nello spazio digitale?,* cit., 5.

146 V. Claussen, *Fighting hate speech and fake news*, cit., 118.

147 O. Pollicino – G. De Gregorio, *Hate speech: una prospettiva di diritto comparato*, cit., nota 59.

in considerazione gli interessi della libertà di espressione[148]. Poiché, secondo la Corte costituzionale federale tedesca, tutti i diritti umani costituiscono una concretizzazione del principio di dignità, è necessaria una motivazione particolarmente accurata per affermare che l'esercizio di un diritto fondamentale, quale la libertà di espressione, possa giustificare una violazione della dignità umana, per sua natura stessa inviolabile[149]. Ne deriva che anche questo ordinamento giuridico concepisce la proporzionalità come strumento di controllo della conformità delle fattispecie incriminatrici in commento ai diritti fondamentali[150]. Più nel dettaglio, la giurisprudenza costituzionale tedesca richiede che esse perseguano «uno scopo legittimo e [siano] idonee, necessarie e, in senso stretto, proporzionate al raggiungimento di tale scopo»[151].

Nel modello tedesco, il concetto di dignità umana vincola il diritto al rispetto della persona. Per questo motivo, tale nozione viene spesso concepita in termini di violazione, che può consistere nella denigrazione, nella stigmatizzazione, nella persecuzione ovvero in qualsiasi altra condotta che metta in dubbio la qualità della persona che la subisce[152].

Infine, si precisa che, secondo la Corte Suprema federale di giustizia (*Bundesgerichtshof*, in forma abbreviata *BGH*), l'odio rilevante ai sensi del § 130 *StGB*, si sostanzia in un comportamento che colpisce i sentimenti, ovvero l'intelletto, di un'altra persona, oggettivamente idoneo oppure soggettivamente determinato a creare, ovvero rafforzare, un atteggiamento emotivamente ostile nei confronti di parte della popolazione, che vada oltre il semplice rifiuto o disprezzo[153]. È evidente che il carattere doppiamente lesivo dei crimini in questione[154] venga sottolineato anche dalla giurisprudenza tedesca.

Volgendo, infine, lo sguardo all'ordinamento spagnolo, la l. organica n. 1 del 30 marzo 2015[155], in linea con l'evoluzione del diritto internazionale e, in particolare, della giurisprudenza della Corte europea dei diritti dell'uomo, ha rafforzato le politiche criminali antidiscriminatorie[156], introducendo l'art. 510 del Codice penale, rubricato "reato di incitamento all'odio e alla violenza". Successivamente, il novero dei crimini d'odio è stato ulteriormente ampliato dalla riforma introdotta dalla l. organica n. 6 del 2022[157].

---

148 BVerfG, 06 settembre 2000, n. 1056, par. 43.

149 A. Espinoza – J. Rivas Alberty, *Limitaciones Inmanentes de la libertad de expresiòn*, cit., 228.

150 Ivi, 223.

151 BVerfGE, 24 febbraio 1961, n. 12, 205.

152 BVerfG, 04 febbraio 2010, n. 369, par. 32.

153 Concetti per i quali J. Rinceanu, *Verso una forma di polizia privata nello spazio digitale?*, cit., 4 rimanda, quanto a formante giurisprudenziale, a BGH, Urt. v. 27.7.2017, 3 StR 172/17, BGH, Urt. v. 3.4.2008, 3 StR 394/07 e, quanto a dottrina, a Sternberg-Lieben, Schittenhelm, *§ 130, Rd. 5a*, in *Schönke, Schröder, StGB – Strafgesetzbuch Kommentar*, Aufl. 2019, 30.

154 Per un cui maggiore approfondimento si rinvia al par. 2 del presente contributo.

155 Che modifica la l. organica 10/1995 del 23 novembre 1995 sul Codice penale.

156 G. Giorgini Pignatiello, *Profili Comparati e problemi costituzionali della legislazione contro l'omotransfobia,* cit., 1006.

157 L. complementare alla l. 15/2022 per la parità di trattamento e la non discriminazione, che modifica la l. organica n. 23/1995, e ha introdotto altre due fattispecie, l'aporofobia e l'antiziganismo. Si segnaa che l'introduzione di questi reati è stata rafforzata con l'elaborazione del *Protocollo d'azione per le forze di sicurezza e i corpi d'armata per i crimini d'odio e i comportamenti che violano le norme giuridiche sulla discriminazione* e la creazione di 53 uffici di Procura specializzati, un Procuratore di Sezione presso la Corte Suprema e 52

Merita una menzione il fatto che, secondo un approccio da molti definito "olistico", il *Codigo penal* disciplini in modo dettagliato gli atti di discriminazione basati sull'orientamento sessuale e l'identità di genere dall'art. 510 all'art. 521[158]. Nello specifico, l'art. 510[159] sanziona la condotta di chiunque incoraggi, promuova ovvero inciti pubblicamente discriminazione, odio, ostilità, oppure violenza, direttamente o indirettamente, o leda la dignità delle persone attraverso atti che comportano umiliazione, disprezzo o discredito di un gruppo, ovvero di un'associazione, ovvero di una parte di esso o contro una persona determinata, a causa della sua appartenenza a esso, per motivi razzisti, antisemiti o per altri motivi riferiti all'ideologia, alla religione o alle convinzioni, alla situazione familiare, all'appartenenza dei suoi membri ad un'etnia, una razza o una nazione, la loro origine nazionale, il loro sesso, orientamento o identità sessuale, per ragioni di genere, malattia o disabilità[160]. Il c. 3 costituisce il punto nevralgico per la criminalizzazione del *ciberodio* oppure *odio online*: esso stabilisce che la disposizione in commento si applica anche quando «i fatti si sono svolti attraverso un social media, via Internet o attraverso l'uso di tecnologie informatiche, in modo da renderli accessibili a un gran numero di persone»[161].

In sintesi, anche la Spagna, al pari del legislatore tedesco, ha anticipato la soglia della rilevanza penale delle condotte, incriminando quelle che siano da sole idonee a incitare la commissione di *hate speech*. Inoltre, come il legislatore francese, ha previsto un apposito comma per punire la realizzazione del reato sul web.

È bene precisare che a differenza dell'art. 510, che criminalizza le condotte di *hate speech*, l'art. 22, al c. 4, del Codice penale prevede un'aggravante generica per motivi discriminatori, la cui sussistenza darebbe luogo a un crimine d'odio in senso stretto[162]. Considerato che il combinato disposto di tale aggravante con il reato di provocazione e di apologia alla commissione di crimini d'odio, punito all'art. 18[163], potrebbe supplire

procuratori provinciali per il servizio di crimini d'odio e discriminazione, così R.V. Candalija, *Tratamiento jurídico del discurso y los delitos de odio en Irlanda y España: una visión penal comparativa*, in *Revista General de Derecho Penal*, 39, 2023, 22. Per un maggiore approfondimento circa la norma citata vedasi E. Nuñez Castaño, *Libertad de Expresión y Derecho Penal: la criminalización de los discursos extremos*, Aranzadi/Civitas, 2022, 238-241.

158 L. Goisis, *La violenza di genere in ottica comparata. La recente novella spagnola. Verso la progressiva affermazione di un modello consesualistico*, in *Genius Rivista di studi giuridici sull'orientamento sessuale e l'identità di genere,* del 16 maggio 2023, 11.

159 Per un approfondimento dei delitti d'odio nel quadro giuridico spagnolo sussiste un ampio numero di contributi ed opere di riferimento, come R. Alcacer Guirao, *La libertad de Odio. Discurso intolerante y protección penal de las minorías*, Marcial Pons, Madrid, 2020; M. Díaz Y García Conlledó, *El discurso del odio y el delito de odio de los artículos 510 y 510 bis del Código Penal: Necesidad de limitar*, en *Boletín Límites a la Libertad de Expresión, Juezas y Jueces para la Democracia*, 5, 2018; A. Galán Muñoz, *¿Juntos o revueltos? Algunas consideraciones y propuestas sobre la cuestionable fundamentación y distinción de los delitos de odio y el discurso del odio*, in J.León Alapont, *Temas Claves de Derecho Penal. Presente y futuro de la política criminal* en España, Bosch Editor, Barcelona, 2021.

160 Art. 510, c. 1, Codice penale spagnolo.

161 I. G. Benito, *Ciberodio: Un estudio de derecho penal comparado*, in *Cuadernos de RES PUBLICA en derecho y criminología*, 3, 2024, 16.

162 D. P. Herradón, *La responsabilidad de las entidades deportivas como consecuencia de los delitos de odio por razones religiosas que se producen en sus instalaciones*", in S. Meseguer Velasco-E. García Antón Palacios, *Deporte, diversidad religiosa y Derecho*, Aranzadi, Cizur Menor, 2020, 221.

163 J. Bernal Del Castillo, *La justificación y enaltecimiento del genocidio en la Reforma del Código Penal de 2015*, in

alla punizione anche dei discorsi d'odio[164], ciò ha indotto a dubitare circa la necessità della fattispecie autonoma di cui all'art. 510[165]. Difatti, tra le criticità riscontrate nella formulazione, si menziona la portata apparentemente priva di limiti, che secondo alcuni condurrebbe all'estensione del suo ambito applicativo sino al "parossismo"[166]. A causa della mancata specificazione degli elementi tipici, l'art. 510 si riferirebbe a ogni condotta "in procinto" di raggiungere un livello di eccezionale gravità. Secondo un orientamento, la via dell'azione penale intermedia, che predilige la tecnica dell'aggravamento dei reati di espressione attraverso l'art. 22, c. 4, per colmare l'indeterminatezza concettuale dell'art. 510, sarebbe più equilibrata in termini sanzionatori. Tale approccio predilige la creazione di un'ampia rete dell'azione penale[167].

Ciò nonostante, l'applicazione dell'art. 22, c. 4, porta con sé il rischio di snaturare l'essenza stessa dei discorsi d'odio, quali manifestazioni del pensiero proiettate verso l'esterno e capaci di valicare i confini dell'individuo, raggiungendo il gruppo di riferimento. Il filone interpretativo sopra menzionato non terrebbe debitamente in considerazione che, così ragionando, l'attenzione si concentrerebbe sul "microconflitto"[168] anziché sulla sua rilevanza super-individuale. In altre parole, si perderebbe di vista l'impatto sul gruppo che il discorso d'odio è in grado di generare e, quindi, la sua dimensione strutturale. L'unica via percorribile sembra quella di distinguere con nettezza gli ambiti applicativi dell'art. 510, strumento marcatamente eccezionale di tutela avanzata[169], sia nell'ambito dell'istigazione che in quello del discorso diffamatorio, e dell'art. 22, c. 4, quale mera circostanza aggravante. Di conseguenza, una corretta interpretazione dell'art. 510 richiede una lettura particolarmente restrittiva; di qui, l'innalzamento della soglia di gravità richiesta ai fini della rilevanza penale della condotta attraverso la necessaria connotazione collettiva del discorso d'odio, a prescindere dal mezzo utilizzato[170]. L'applicazione della circostanza aggravante esaminata dovrebbe limitarsi ai commenti offensivi indirizzati a un individuo in quanto tale e non alla sua collettività di appartenenza (come invece accade nei discorsi d'odio)[171].

In linea generale, l'intervento legislativo spagnolo si è caratterizzato per un approccio sistematico mediante l'adozione di un ricco impianto normativo, che assomma diverse tecniche incriminatrici[172] e che, al pari di quello francese, ha suscitato profonde critiche

In Dret Penal. Revista para el análisis del Derecho, 4, 2015, 9-12.

164 I. G. Benito, *Ciberodio: Un estudio de derecho penal comparado*, cit., 29.

165 A. Galán Muñoz, *"¿Juntos o revueltos? Algunas consideraciones y propuestas sobre la cuestionable fundamentación y distinción de los delitos de odio y el discurso del odio*", in J. León Alapont, *Temas Claves de Derecho Penal. Presente y futuro de la política criminal en España*, Bosch Editor, Barcelona, 2021, 327-328.

166 Così, R. A. Guirao, *Dimensiones del discurso de odio*, in *Derecho penal y orden constitucional: límites de los derechos políticos y reformas pendientes*, Tirant lo Blanch, 2022*,* 41.

167 Di tale orientamento dà atto I. G. Benito, *Ciberodio: Un estudio de derecho penal comparado*, cit., 30.

168 *Ibid.*

169 I. G. Benito, *Ciberodio: Un estudio de derecho penal comparado*, cit., 25.

170 Ivi, 30.

171 Ivi, 31.

172 G. Giorgini Pignatiello, *Profili Comparati e problemi costituzionali della legislazione contro l'omotransfobia*, cit., 1010.

in ragione della sua notevole forza espansiva[173]. Già prima della riforma del 2022, si riteneva che le condotte tipizzate dagli interventi legislativi del 1995[174] e del 2015[175] potessero difficilmente conciliarsi con i principi fondamentali che governano un diritto penale democratico, in quanto ritenuti capaci di oltrepassare i limiti di un intervento punitivo minimo[176].

In riferimento alla criminalizzazione dei discorsi d'odio anche nelle manifestazioni indirette, il formante giurisprudenziale ha mantenuto un orientamento ondivago[177]: se negli 90'[178] ha riconosciuto la punibilità delle stesse, nei primi anni 2000[179] si è orientato in senso restrittivo, tornando ad abbracciare soltanto nel 2016 l'orientamento estensivo che ne aveva caratterizzato l'operato alle origini, altresì, allineandosi alla giurisprudenza della Corte EDU[180].

Il Tribunale costituzionale spagnolo ha precisato che un discorso d'odio che incitava alla violenza elogiando l'autore di attività terroristiche non può rientrare nel contenuto costituzionalmente protetto del diritto alla libertà di espressione[181]. La giurisprudenza costituzionale è giunta a conclusioni simili in relazione a espressioni o campagne di carattere razzista o xenofobo, in quanto la Costituzione non garantisce il diritto di diffondere un determinato significato della storia o una concezione del mondo con l'intento deliberato di sminuire e discriminare persone o gruppi per ragioni legate a qualsiasi condizione personale, etnica o sociale[182].

Per determinare se una condotta costituisce una manifestazione punibile, la giurisprudenza costituzionale osserva il seguente giudizio sommario: in primo luogo, valuta se si tratta di un'espressione d'odio basato sull'intolleranza, formulata in modo aggressivo e caratterizzata da ostilità inequivoca nei confronti di altri individui. Secondariamente, verifica se la diffusione mediante i mezzi di comunicazione produce un effetto equiparabile a quello dei mezzi di diffusione tradizionali, come i giornali e i notiziari televisivi. Tuttavia, tale giudizio non può ritenersi esaustivo. L'impatto potenziale, la lesione della dignità e il pericolo per il tessuto sociale sono presi in esame come elementi di carattere assoluto, senza considerare altri criteri, come l'interesse generale o il contributo

---

173 M. Revenga Sánchez, *El discurso del odio: entre la trivialización y la hiper-penalización*, in *Liber Amicorum per Pasquale Costanzo*, 13-2-2019, 1-14, giurcost.org.

174 L. organica n. 10/1995 sul Codice penale.

175 L. organica n. 1/2015 sul Codice penale.

176 G. Giorgini Pignatiello, *Profili Comparati e problemi costituzionali della legislazione contro l'omotransfobia*, cit., 1008-1009.

177 Cfr. M. Iacometti, *Il Tribunale costituzionale spagnolo verso l'ipertrofia del concetto di "discurso del odio" e la eccessiva compressione della libertà di espressione?*, in *Rivista Associazione Italiana dei Costituzionalisti*, 1, 2017.

178 Ricostruzione effettuata da G. Giorgini Pignatiello, *Profili Comparati e problemi costituzionali della legislazione contro l'omotransfobia*, cit., nota 60, mediante il riferimento a Tribunale cost. spagnolo, n. 214, 11 novembre 1991.

179 Tribunale cost. spagnolo, n. 235, 07 novembre 2007.

180 Tribunale cost. spagnolo, n. 112, 20 giugno 2016.

181 *Ibid.*, richiamato da A. Espinoza – J. Rivas Alberty, *Limitaciones Inmanentes de la libertad de expresiòn*, cit., 246-247.

182 Tribunale cost. spagnolo, n. 214, 11 novembre 1991.

alla formazione di un'opinione pubblica libera e pluralista[183]. Seguendo questa logica, gli organi giudiziari arriverebbero a sostenere che determinate condotte, pur non godendo di tutela costituzionale, non potrebbero essere punite. Tale risultato verrebbe raggiunto mediante l'oggettivizzazione delle garanzie proprie del diritto penale che derivano dall'effetto di irradiazione dei diritti fondamentali operanti nel caso specifico. L'influenza di tali diritti sul piano penale comporta il rafforzamento del principio di legalità. Il diritto fondamentale può, pertanto, operare come causa di esclusione dell'antigiuridicità o come causa di giustificazione dell'eccesso nell'esercizio di un diritto, a condizione che non ne venga snaturato il contenuto[184].

Il quadro comparativo sopra tracciato mostra come gli ordinamenti penalistici esaminati abbiano previsto l'incriminazione delle condotte d'odio, anche online, mediante forme di anticipazione della soglia della rilevanza penale, al fine di incriminare quelle condotte che possano anche soltanto mettere a repentaglio l'esistenza, o il godimento, dei beni giuridici tutelati. Con la precisazione che tali comportamenti devono essere rivolti all'individuo in quanto membro del suo gruppo di appartenenza. A tal fine, le scelte di politica criminale analizzate si sono caratterizzate per il ricorso alla tecnica dell'istigazione o della messa in pericolo.

In Francia, è stato sottolineato come la libertà di manifestazione del pensiero costituisca una condizione essenziale per la sopravvivenza stessa della democrazia e l'incriminazione è stata operata mediante la creazione di una figura criminosa basata sulla forma di imputazione soggettiva della "*mise en danger*".

Per quanto riguarda il quadro tedesco, è importante sottolineare l'impegno legislativo nel tentativo di definire con rigore le condotte d'odio, mediante l'introduzione di un elenco tassativo di contenuti da qualificare come illeciti. Inoltre, la fattispecie di discorso d'odio è individuata tramite il rinvio al reato di istigazione all'odio, soluzione che trova fondamento nella prevalenza del principio della dignità umana su ogni altro diritto fondamentale dell'individuo, inclusa la libertà di espressione.

In Spagna, si è posta la questione del mancato utilizzo della specifica figura incriminatrice appositamente introdotta, costruita sul modello istigatorio, al pari di quanto occorso anche nell'ordinamento tedesco. Conseguentemente, nella prassi talvolta si ricorre al combinato disposto degli artt. 18 e 22, c. 4, e, quindi, alla fattispecie generica di istigazione alla commissione di delitti aggravata dai motivi discriminatori, a causa della ritenuta indeterminatezza concettuale dell'art. 510 del Codice penale. Tale corollario costituisce uno dei principali motivi delle critiche rivolte a un così ricco impianto normativo, ritenuto in grado di superare i limiti di un intervento punitivo minimo.

In sede di accertamento, principalmente secondo un giudizio scrupoloso presidiato dai principi di proporzionalità, adeguatezza e necessarietà, è demandato al formante giurisprudenziale il compito di verificare lo stato di pericolo della libertà di espressione dell'autore del discorso d'odio e la capacità di sopravvivenza del nucleo di individui cui le parole pericolose si rivolgono.

Dall'analisi delle diverse esperienze giuridiche emerge un'identità di vedute anche in merito all'oggetto della tutela giuridica: il discorso d'odio deve porre in pericolo l'indi-

[183] A. Espinoza – J. Rivas Alberty, *Limitaciones Inmanentes de la libertad de expresiòn*, cit., 248.

[184] Ivi, 249.

viduo non come singolo, bensì come componente di una collettività e, in ultima analisi, dell'agire democratico plurale, multiculturale e eterogeneo.
In conclusione, si ritiene di concordare circa la preferibilità dell'utilizzo di strumenti di tutela penale avanzata, atteso che l'elevata soglia di gravità varcata da questi crimini giustifica la specifica incriminazione dei discorsi d'odio, anche online, nell'ottica di una salvaguardia quanto più effettiva del bene giuridico super-individuale tutelato, pilastro dello Stato di diritto democratico.

5. Lo stato dell'arte oltre Oceano

L'orientamento nordamericano, basato sul Primo emendamento della Costituzione[185], non tollera alcuna interferenza da parte dei poteri pubblici nell'esercizio della libertà di espressione e non prevede limitazioni con riguardo ai contenuti espressi.
Questa ideologia liberale ha per molto tempo giustificato l'astensionismo dalla regolamentazione dell'*hate speech*; in tale contesto, la libertà di parola del singolo può subire limitazioni solo in presenza di altri diritti costituzionalmente tutelati, e sempre entro margini molto ristretti. Per distinguere le condotte coperte dal Primo emendamento la giurisprudenza della Corte Suprema ha fatto ricorso ai criteri del *clear and present danger* [186] e delle *fighting words*, il cui comune assunto è quello di ritenere che la manifestazione del pensiero possa essere limitata non tanto in relazione al contenuto del messaggio, bensì in quanto idonea a tradursi in atto violento. In altre parole, secondo tali teorie, la limitazione del pensiero può avvenire tutte le volte in cui i discorsi d'odio producono conseguenze tangibili e dannose per l'ordine pubblico.
Una prima formulazione del concetto di *clear and present danger* è stata fornita dal giudice Holmes[187] nel caso *Schenk v. United States* del 1919 e in *Abrams v. United States*[188]. In tale ultima pronuncia egli scrisse: «the Unite States constitutionally may punish speech that produces or is intended to produce a clear and imminent danger». Con queste parole la Corte Suprema scelse di non indicare i criteri in base ai quali graduare la pericolosità del discorso, al fine di permettere allo Stato di scegliere di vietare le opinioni apparentemente minacciose, analizzando caso per caso e tenendo in considerazione il contesto e la natura delle parole utilizzate.
In una pronuncia successiva degli anni Cinquanta, *Dennis v. United States*[189], la Corte

[185] «Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances».

[186] Il primo a teorizzare questa dottrina fu il Giudice della Corte Suprema Holmes, nel caso *Schenk v. United States*, nel 1919; successivamente sono risultate rilevanti al riguardo anche le pronunce *Chaplinsky v. New Hampshire,* del 1942 e *Brandeburg v. Ohio,* del 1969. Si segnala altresì la lettura di A. Fricano, *Prove tecniche di resilienza costituzionale: l'assalto a Capitol Hill e la censura mediatica di Donald Trump,* in *Rivista del Gruppo Pisa*, fascicolo speciale monografico, 3, 2021, 742-743.

[187] Si segnala A. Lewis, *Freedom for thought that we hate,* New York, 2007, 11-20, per una rassegna delle opinioni del Giudice Holmes nelle sentenze della Corte Suprema.

[188] *Abram v. United States,* 250 U.S. (1919).

[189] *Dennis v. United States,* 345 U.S. 494 (1951), inerente ad un caso di presunta incostituzionalità dello *Smith Act* del 1940, il quale stabilita l'illegalità di costituire organizzazioni atte a sovvertire qualsiasi

Suprema, conscia dell'impossibilità di determinare la probabilità che una minaccia sovversiva si traducesse in azione, attribuì agli aggettivi *clear and present danger* portata preventiva: «the words cannot mean that, before the Government may act, it must wait until the putsch is about to be executed». Con tali parole si inaugurò un nuovo filone destinato a far scuola nel futuro diritto penale c.d. «prepressivo»[190]. In particolare, la Corte rimodulò il criterio della punibilità alla stregua di un pericolo astratto ed elaborò una nuova teoria del rischio presunto: se il pregiudizio è davvero grave, ai fini della punibilità dell'autore della condotta non è necessario che l'istigazione sia idonea a indurre qualcuno, in un intervallo temporale ristretto, alla sua commissione.

Successivamente[191], a partire dal caso *Brandenburg v. Ohio*[192], il test del *clear and present danger* iniziò ad essere applicato in modo più rigoroso. In particolare, la Corte Suprema dichiarò incostituzionale l'*Ohio criminal Syndicalism Statute,* approvato nel 1919, che prevedeva la punizione dell'istigazione alla violenza finalizzata alla promozione di riforme industriali o politiche. In tale occasione, recuperando la necessaria carica di offensività della fattispecie di istigazione, la Corte rovesciò la condanna e dichiarò protetto dal Primo emendamento qualunque forma di incitamento a usare la forza o a violare la legge (ad eccezione dell'ipotesi di produzione di un'azione illegale imminente e idonea al raggiungimento dello scopo). Nel passaggio chiave della sentenza si legge che un discorso può essere punito se si verificano due condizioni: i) intenzionalità di adottare un comportamento violento e ii) l'idoneità delle parole utilizzate a produrre un pericolo. L'idoneità e l'imminenza, quali requisiti necessari per violare i margini del Primo emendamento, sono stati successivamente affinati in *Hess v. Indiana*[193] e in *N.A.A.P. v. Claiborne Hardware Co*[194], pur rimanendo ambigui.

Analizzando ora il criterio delle cosiddette *fighting words*, è opportuno precisare che tale concetto comprende tutte quelle espressioni verbali e non verbali che manifestano un

governo negli Stati Uniti mediante la forza e la violenza. Il ricorrente, Eugene Dennis, fu condannato per aver tentato di costituire un partito di matrice comunista.

190 Per il significato di questo neologismo e per un'analisi della recente diffusione globale della giustizia penale «prepressiva», si rinvia a E. Grande, *Il dispositivo penale della paura. Diffusione di un'ideologia,* in *Democrazia e Diritto*, 3, 2014,17 e ss.

191 In costanza del periodo della guerra del Vietnam.

192 Un leader del Ku Klux Klan venne condannato per il reato di istigazione a un illecito penale o alla violenza (previsto dal legislatore dell'Ohio) per aver organizzato una manifestazione razzista in cui aveva paventato il ricorso alla vendetta nei confronti delle istituzioni politico-giuridiche degli Stati Uniti (Presidente, Congresso e Corte Suprema), che a suo giudizio non avrebbero tutelato abbastanza i bianchi nei confronti dei neri e degli ebrei. Alla manifestazione, cui molti parteciparono armati, il leader dell'organizzazione razzista aveva invitato un cameraman della televisione per farsi filmare. Una volta trasmesso il servizio egli fu processato e successivamente condannato.

193 414 U.S. 105 (1973). Il caso riguardava un manifestante che dopo essere stato disperso dalla polizia aveva gridato «we'll take the fucking street», senza indicare quando. Tale istigazione fu ritenuta non capace di condurre a un'azione violenta imminente. Non c'era prova, sostenne la Corte, che le «parole fossero volte a produrre e fossero idonee a produrre un disordine imminente».

194 458 U.S. 886 (1982*)*. In questa vicenda, inerente il boicottaggio di alcuni negozi i cui titolari erano stati accusati di discriminare persone nere, la Corte ritenne che la seguente affermazione rivolta alla gente: «se becchiamo uno di voi entrare in questi negozi razzisti vi rompiamo il collo», pronunciata da uno dei responsabili dell'associazione nazionale per i diritti della gente di colore, fosse protetta dal Primo emendamento, in quanto «mere advocay of the use of force or violence does not remove speech from the protection of the first Amendment», ivi, 928.

chiaro intento di incitare alla violenza da parte di chi le pronuncia. Si tratta, dunque, di dichiarazioni che sono immediatamente in grado di generare disordine nel contesto pubblico, in quanto finalizzate a tradursi in comportamenti idonei a suscitare reazioni violente[195].

I capisaldi giurisprudenziali di questo criterio sono enucleati dalla seguente casistica.

Nel caso *Chaplinsky v. New Hampshire*[196], un testimone di Geova, che si era issato su un banchetto collocato sul marciapiede nel centro di Rochester, New Hampshire, inveiva contro le organizzazioni religiose. Invitato dallo sceriffo a interrompere la sua condotta, egli reagiva insultandolo con frasi come «you are a god damned racketeer» e a «damned Fascist». Arrestato, fu condannato in base a una legge nazionale che puniva la condotta di chi offendeva intenzionalmente una persona in un luogo pubblico. In merito, la Corte Suprema, investita della questione per verificarne l'incostituzionalità, all'unanimità la dichiarò infondata, ritenendo che le espressioni usate avrebbero dovuto essere ricomprese nella categoria delle *fighting words,* in quanto non idonee ad apportare alcun valore aggiunto al dibattito pubblico[197]. Nel caso specifico, la Suprema Corte individuò nella carica aggressiva delle espressioni utilizzate il discrimine per distinguere le parole violente da quelle non violente, concludendo di non poterle considerare, in quella circostanza, mera espressione di un punto di vista e, quindi, libera manifestazione del pensiero[198].

Questo primo approdo venne successivamente testato nel famosissimo caso di *hate speech* americano, ovverosia *Village of Skokie v. National Socialist Party of America*[199]. La vicenda trae origine dal divieto imposto a un partito filonazista di sfilare in corteo in una cittadina dell'Illinois, prevalentemente abitata da appartenenti alla comunità ebraica. I partecipanti alla manifestazione avrebbero, come in altre occasioni, voluto sfilare indossando uniformi naziste e mostrando cartelli con scritte del seguente tenore: «free speech for the White man» e «Free speech for White America». La Corte Suprema, chiamata a valutare la legittimità del divieto imposto, affermò che non era possibile escludere dal dibattito pubblico le espressioni ovvero l'esposizione di simboli che, solo potenzialmente, avrebbero potuto causare rabbia o risentimento nel pubblico[200]. A fronte di tale posizione, anche la Corte dell'Illinois stabilì che la marcia avrebbe semplicemente rappresentato una forma simbolica libertà di espressione.

Successivamente con *R.A.V. v. City of St. Paul* [201], un gruppo di ragazzi adolescenti venne condannato per aver collocato una croce fiammeggiante davanti al giardino di proprietà di una famiglia afroamericana. La condanna si fondava su un'ordinanza della città di St. Paul, nel Minnesota, la quale vietava di occupare il suolo pubblico o privato con simboli, oggetti, graffiti o altro «*including, but not limited to, a burning cross or Nazi swastika, which one knows or has reasonable grounds to know arouses anger, alarm or resentment in*

195 P. Tanzarella, *Discriminare parlando,* cit., p. 57 e ss.

196 *Chaplinsky v. New Hampshire*, 568, U.S.315 (1942).

197 *Chaplinsky v. New Hampshire*, 568, U.S.315 (1942).

198 *Ibid.*: «*are not essential part of any exposition of ideas*».

199 432 U.S. 43 (1977).

200 *Ibid.*

201 505, U.S. 377 (1992).

others on the basis of race, color, creed, religion or gender». La Suprema Corte del Minnesota, interessata della questione, sottopose alla Corte Suprema federale un dubbio di costituzionalità con riferimento al Primo emendamento. La sentenza fu redatta principalmente dal giudice Scalia, secondo il quale sarebbe stato un errore includere nella categoria delle *fighting words* le idee, facendo esclusivo riferimento al loro contenuto e non, invece, alla loro effettiva pericolosità. In effetti, ai fini dell'applicazione del criterio in esame, sarebbe rilevante solo il modo in cui l'idea odiosa viene espressa[202]. Sulla base di tale ragionamento si spiega, quindi, il motivo della decisione di incostituzionalità dell'ordinanza della città di St. Paul: essa non era diretta a punire un modo – pericoloso e offensivo – di espressione[203], bensì esclusivamente, a selezionare i messaggi intolleranti da condannare, ovvero quelli riguardanti la razza, il genere e la religione, trascurando altri tipi di discorsi che ben avrebbero potuto urtare la sensibilità di altre categorie di persone (come, per esempio, quelli riguardanti l'orientamento sessuale, non inclusi nel divieto)[204].

I principi sanciti nella sentenza *R.A.V.*[205] sono stati ulteriormente ribaditi in *Black v. Virginia*[206], dove la croce fiammeggiante era stata innalzata nel giardino di una famiglia afroamericana. In questo caso venne sollevato un dubbio di costituzionalità sulla legge della Virginia volta a condannare «*any person [...], with the intent of intimidating any person or group [...], to burn a cross on the property of another, a highway or other public place», specificando inoltre che «[a]ny such burning [...] shall be prima facie evidence of an intent to intimidate a person or group*»[207]. Sulla scia della giurisprudenza *R.A.V.*, la Corte Suprema dichiarò l'incostituzionalità di tale disposizione, affermando che le leggi volte a punire le condotte descritte, negando la possibilità di indagare sui motivi che hanno condotto a compierle, devono necessariamente ritenersi incostituzionali. La valutazione del caso concreto deve essere sempre essere ammessa, in quanto solo attraverso di essa si potrebbe stabilire la vera natura dell'atto, se intimidatorio o meramente espressivo (come ritenuto nel caso specifico) di un dissenso, privo dell'intenzione di minacciare qualcuno[208].

---

202 *Ibid.,* così il Giudice Scalia: «*the reason why fighting words are categorically excluded from the protection of the First Amendment is not that their content communicates any particular idea, but that their content embodies a particularly intolerable (and socially unnecessary) mode of expressing whatever idea the speaker wishes to convey*».

203 *Ibid.,* «*has not singled out an especially offensive mode of expression – it has not, for example, selected for prohibition only those fighting words that communicate ideas in a threatening (as opposed to a merely obnoxious)*».

204 *Ibid.,* «*the ordinance applies only to 'fighting words' that insult, or provoke violence, 'on the basis of race, color, creed, religion, or gender'. Displays containing abuse invective, no matter how vicious or severe, are permissible unless they are addressed to one of the specified disfavored topics. Those who wish to use 'fighting words' in connection with other ideas – to express hostility, for example, on the basis of political affiliation, union membership, or homosexuality – are not covered*».

205 Si segnala che i giudici statali hanno riscontrato difficoltà applicative dei principi sanciti nella citata sentenza, sul punto si richiama l'analisi di V. M. Manetti, *L'incitamento all'odio razziale tra realizzazione dell'eguaglianza e difesa dello Stato*, Torino, 2005, 103 – 137, la quale ricorda come il Giudice Stevens avrebbe ribadito, nella sua opinione dissenziente, che il caso *RAV* avrebbe assicurato una tutela maggiore al pensiero razzista rispetto a quella conferita al *commercial speech*.

206 538, U.S. 343 (2003).

207 Va. Code Ann. § 18.2-423 (1996).

208 Questa argomentazione è stata criticata dai sostenitori della teoria critica della razza c.d. *critical race theory*. Per questi ultimi, i discorsi d'odio non costituiscono semplici opinioni fuori dal coro ma, essendo diretti contro minoranze vulnerabili (come la comunità afroamericana negli Stati Uniti), rappresentano atti di violenza verbale volti ad avvilire le vittime, a ridurle al silenzio e a emarginarle sul piano sociale e

Posta in questi termini la teoria delle *fighting words*, così come quella del *clear and present danger,* risulta relegata ai casi limite, facendo apparire gli Stati Uniti come i massimi garanti della libertà di manifestazione del pensiero[209]. Entrambe, tuttavia, mostrano chiaramente che la Corte Suprema non sia mai riuscita (o non abbia mai voluto) classificare in maniera tassativa le ipotesi di punibilità dei discorsi d'odio in relazione al pericolo concreto prodotto, lasciando ai giudici ampio spazio di apprezzamento a seconda delle singole circostanze. Ciò, diversamente rispetto a quanto avviene nelle Corti Europee ove, come visto[210], sulla base dell'art. 10 CEDU e dell'art. 11 della Carta dei diritti fondamentali dell'Unione europea, pur proclamando la libertà di espressione come diritto fondamentale, si delineano le situazioni giuridiche nelle quali essa trova espressione[211].
Nel contesto attuale dell'esperienza giuridica oltreoceano, sono le modalità espressive a giocare un ruolo centrale, costituendo l'unico criterio per valutare l'applicazione del Primo Emendamento. In particolare, per quanto riguarda i mezzi di informazione, le differenze nelle loro caratteristiche giustificherebbero una differenziazione negli standard applicativi dello stesso Emendamento[212].
Con specifico riferimento alla libertà di espressione online, a partire dalla nota sentenza *Reno v. ACLU*, del 1997, relativa all'incostituzionalità del *Communication Decency Act*

politico. Ci si può anche chiedere se intimidazioni come la pratica delle croci infiammate – tipica del *Ku Klux Clan* – siano forme di espressione ammesse in un dibattito di idee degno di questo nome. Infine, in un'ottica europea, la sentenza colpisce anche per le sue considerazioni sbrigative sui diritti delle minoranze: su questo tema, la Corte si limitò infatti ad affermare che, per proteggerli, il legislatore non deve per forza ricorrere a misure che restringono la libertà di espressione ma dispone di alternative.

[209] Per un esame ulteriore dei casi di *hate speech* rispetto ai quali hanno trovato applicazione le dottrine esaminate si rimanda a E. Grande, *Il dispositivo penale della paura*, cit. 47 ss.

[210] Cfr., paragrafo 2.

[211] O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione nell'era di Internet,* in *Rivista di Diritto dei Media,* 1/2018, cit. 3 ss.; M. Bassini, *Internet e libertà di espressione. Prospettive costituzionali e sovranazionali,* Roma, 2019, 163 ss.

[212] Il *leading* case è rappresentato dalla decisione *Red Lion Broad. Co. v. FCC*, del 1969, nell'ambito della quale la Corte Suprema ha valutato la legittimità della c.d. *fairness doctrine*, una policy della *Federal Communications Commission (FCC)* (fondata sul «Radio Act» del 1927 e sulla sez. 307 (licenze) del Titolo 47 (radiodiffusione) dello *US Code*), introdotta nel 1949, in base alla quale la Commissione, in considerazione della scarsità delle risorse nella diffusione radiofonica e televisiva, subordinava la concessione della licenza a trasmettere all'impegno da parte dei titolari delle relative licenze a fornire un servizio di informazione pubblica orientato al pluralismo e alla correttezza. Il tema della scarsità delle risorse d'accesso al mezzo televisivo viene poi ripreso anche in altre decisioni successive ove la Corte sancisce la legittimità costituzionale di alcune disposizioni c.d. *must carry* finalizzate a garantire il pluralismo delle emittenti radiotelevisive imponendo, ad esempio, che una percentuale delle reti via cavo fosse riservata all'emittenza locale. La giurisprudenza della Corte Suprema in relazione alla limitazione della libertà di espressione in ambito radiotelevisivo si spinge sino a pronunciarsi anche sugli stretti contenuti veicolati tramite tali mezzi. In particolare, nella decisione *FCC v. Pacifica Foundation* del 1978, dove conferma il potere della *FCC* di regolare, in circostanze limitate (stabilendo orari e condizioni di trasmissione), programmi che pur non osceni fossero qualificabili come indecenti. A fianco della motivazione concernente la scarsità del bene, nella decisione in esame uno dei perni in relazione al quale il relatore per la Corte ritiene di differenziare il regime di applicabilità del Primo emendamento al mezzo radiotelevisivo è legato alla passività dell'utenza ed alla sua estrema diffusione presso le abitazioni private. Questa giurisprudenza restrittiva, inaugurata per i mass media radiotelevisivi, non si è poi estesa alla carta stampata la quale, anche dopo tale decisione, ha continuato a godere nelle sentenze della Corte Suprema di un'ampia libertà di espressione.

del 1996[213], non hanno trovato applicazione i precedenti giurisprudenziali aventi ad oggetto la diffusione del discorso tramite l'emittenza radiotelevisiva. La Corte, nella persona del Giudice Stevens, operò un netto distinguo, affermando che, Internet non dovesse essere considerato altrettanto invasivo dei *mass media* tradizionali, in quanto l'accesso alle informazioni online richiederebbe un'attività di ricerca da parte dell'utente, in ragione della quale, esso, non potrebbe qualificarsi passivo[214].

Circa vent'anni dopo il caso *Reno,* la Corte suprema con la sentenza *Packingham v. North Carolina* del 19 giugno 2017[215] ebbe l'opportunità di tornare a ragionare sul diritto di Internet[216], seppur in un caso non strettamente attinente ai discorsi di odio online[217].

L'esordio della parte motivazionale della sentenza è significativo nella misura in cui evidenzia che il Primo emendamento trova piena attuazione se chiunque ha «accesso a luoghi in cui possa parlare ed ascoltare, e poi, dopo riflessione, parlare ed ascoltare ancora»[218].

Un altro passaggio rilevante della sentenza è quello in cui specifica che, se in passato potevano nutrirsi perplessità riguardo al luogo più importante per liberamente esprimere la propria opinione, oggi (nel 2017) non vi sono più dubbi: tali sono il *cyber spazio* e i *social media*[219]. Se, quindi, anche in rete possono essere commessi alcuni reati, secondo la Corte non sono giustificati divieti volti a colpire in maniera integrale l'esercizio dei diritti di cui al Primo emendamento sul web «tessuto connettivo della nostra moderna società e cultura»[220].

L'attuale modello statunitense non ha raggiunto un soddisfacente equilibrio tra la libertà di espressione e gli altri interessi di pari livello, non riuscendo a concretizzare in maniera efficace la tutela dalle condotte di *online hate speech* a causa della difficoltà di anticipare la tutela ad un momento in cui altri diritti fondamentali non hanno ancora subito una lesione apprezzabile[221]. Un simile approccio sembra costituire un terreno fertile per i siti Internet che promuovono contenuti discriminatori

---

213 Atto normativo adottato dal Congresso nel 1996 a protezione dei minori dalla diffusione di materiale osceno o pornografico.

214 521 U.S. (1997). Una traduzione italiana della sentenza citata può essere consultata in *Foro Italiano*, 198, VI, 23.

215 582 U.S. (2017), No. 15-1194.

216 V. Neri, *Il lato oscuro della Rete e l'esigenza di una legislazione responsabile. Due recenti casi emblematici,* in *Federalismi.it*, 22 novembre 2017, 22, 2017.

217 Nel 2008 il legislatore del North Carolina introdusse una nuova fattispecie di reato consistente nel prevedere una pena per coloro i quali, iscritti nel registro dei criminali sessuali, accedevano ad un *commercial social networking web site* pur consapevoli del fatto che il sito avrebbe permesso a soggetti minorenni di diventare membri della community o di creare e mantenere pagine personali sul sito: «*Offense. - It is unlawful for a sex offender who si registered in accordance with Article 27° of Chapter 14 of the General Statutes of access a commercial social networking Web site where the sex offender knows that the site permits minor children to become members or to create or maintain personal Web pages or have commercial social networking Web site*» (North Carolina General Statutes, 14-202.5(a)).

218 Corte Suprema federale, *Packingham v. North Carolina*, 582 U.S. (2017), slip opinion, 4.

219 *Ibid.,* 5.

220 *Ibid.,*10.

221 M. Lamanuzzi, *Il "lato oscuro della rete": odio e pornografia non consensuale. Ruolo e responsabilità dei gestori delle iattaforme social oltre la net neutrality",* in *La Legislazione Penale*, 24 maggio 2021, 12 e ss.

Negli anni più recenti, con l'obiettivo di trovare un equilibrio tra la salvaguardia del libero mercato delle idee e il principio di uguaglianza, i legislatori statali[222] hanno cercato di dotarsi di atti normativi anti *hate speech* per tutelare la dignità dei propri cittadini appartenenti alle cosiddette categorie deboli, quali gli immigrati di nuova generazione, gli afro-americani, gli omosessuali e le minoranze religiose, che reclamano una maggior protezione dalle istituzioni, proprio in ragione dell'aumento della frequenza di fenomeni di intolleranza nei loro confronti.

6. Rimedi alternativi: quale ruolo per gli strumenti di "moderazione algoritmica"?

In questo panorama complesso, in cui le pubblicazioni online sono sempre più numerose e provengono da categorie anche molto diverse tra loro, individuare la soluzione atta a contrastare efficacemente il fenomeno in esame significa trovare il giusto coordinamento tra la legge, la sanzione penale e altre possibili misure tecnico-giuridiche. In tale ottica, a partire dalla prima decade degli anni 2000, le piattaforme digitali hanno iniziato a utilizzare gli strumenti di moderazione e a redigere policy dei contenuti diffondibili in rete[223], con l'obiettivo di impedire la pubblicazione, ovvero consentire la rimozione di messaggi, immagini o espressioni ritenuti deplorevoli o inappropriati, sulla base di criteri stabiliti unilateralmente dalle stesse piattaforme[224] e, quindi, altamente discrezionali.

A partire dal 2010 l'attenzione delle scelte legislative e politiche con riferimento alle

[222] Al riguardo si può far riferimento al report diffuso dall'Anti-Defamation League del 2012, con il quale si è documentato che 43 Stati e il distretto di Washington D.C. hanno approvato leggi sui crimini d'odio. Si segnala inoltre che sono parecchie le vie attraverso le quali gli Stati hanno cercato di disciplinare tale fenomeno, soprattutto a seguito dell'attentato alle Torri Gemelle di New York. In particolare, non sono mancate le prese di posizione di alcuni professori universitari licenziati a seguito delle rispettive posizioni antipatriottiche e di condanna alle politiche estere degli Stati Uniti. Per una panoramica completa sul punto si richiama E. Grande*, Il dispositivo penale della paura*, cit., 60-61.

[223] M. Monti, *Privatizzazione della censura e Internet Platforms: la libertà di espressione e i nuovi censori dell'agorà digitale,* in *Rivista Italiana di informatica e diritto*, 1, 2019, 37.

[224] Può essere di interesse, a conferma dell'elevata discrezionalità della classificazione dei contenuti e della conseguente individuazione di quelli vietati, richiamare alcuni recenti interventi dell'*Independent Oversight Board di Facebook*, relativi a contenuti rimossi dalla piattaforma perché ritenuti di incitamento all'odio e riabilitati in quanto riconducibili alla critica politica. È il caso del post contraddistinto da un'immagine di una serie tv turca raffigurante un combattente, con la didascalia «se la lingua del Kafir (non musulmano) si scaglia contro il Profeta, allora la spada deve essere tirata fuori dal fodero» e le connesse definizioni del presidente francese Emmanuel Macron come «il diavolo». Il Comitato, nel caso in esame aveva revocato (decisione 2020-007-FB-FBR) la cancellazione del post originariamente disposta da Facebook che lo aveva ritenuto di incitamento all'odio, qualificandolo come critica politica alla risposta di Macron nei confronti della violenza di matrice religiosa. Un altro caso è quello della pubblicazione di una (presunta) citazione di Goebbels, ministro della propaganda del regime nazista, secondo cui «invece di appellarsi agli intellettuali, le discussioni dovrebbero fare leva sulle emozioni e sugli istinti». Questa frase, in un primo momento rimossa da Facebook in quanto ritenuta una forma di elogio e di supporto al regime nazista, era poi stata nuovamente pubblicata a seguito della decisione del citato Comitato (decisione 2020-005-FB-UA), in quanto volta a creare un parallelo tra il concetto espresso dalla citazione e la presidenza di Donald Trump e quindi riconducile alla critica politica nei confronti del Presidente.

tecniche di regolazione della libertà di espressione è costantemente aumentata, scegliendosi di intervenire nella regolazione delle infrastrutture digitali attraverso l'elaborazione di forme di responsabilità sussidiaria a carico dell'intermediario[225], evitando forme di restrizione e di sanzione che possano ledere la libertà di espressione dei singoli individui. La circostanza che fossero le piattaforme aderenti a dover effettuare una valutazione sostanziale dei contenuti da moderare (e rimuovere), facendosi quindi carico di un delicato bilanciamento fra i valori di rilievo costituzionale in gioco, prospettava una pericolosa equiparazione dei PSI agli organi giudicanti, chiamati a sindacare la legalità delle condotte espressive[226].

In Europa, la normativa relativa alla moderazione dei PSI si compone di misure legislative, non legislative e volontarie, sia di carattere generale (applicabili a tutti i contenuti) che di carattere speciale (con riferimento agli specifici obiettivi da raggiungere)[227]. Il punto di riferimento è rappresentato dal *Code of conduct on countering illegal hate speech online,* atto a carattere non vincolante, adottato il 30 maggio 2016 e immediatamente sottoscritto dai principali attori del mondo digitale. Questo strumento nasce dall'esigenza di costruire un'alleanza tra gli attori privati (PSI) e le istituzioni europee[228]. La disciplina di tale codice affida alle piattaforme il compito di eliminare contenuti qualificati come atti di incitamento all'odio.

Seppur pregevole in quanto rappresenta un tentativo concreto di coniugare i diversi strumenti utilizzati per impedire il dilagare del fenomeno di odio online, tale codice

[225] In una prima fase, la Commissione europea faceva per lo più ricorso a strumenti di auto-regolazione. Tra questi, il rimando è soprattutto al Codice di condotta dell'UE per contrastare l'illecito incitamento all'odio (2016), nonché al Codice di buone pratiche sulla disinformazione (2019). Progressivamente, l'approccio della Commissione è mutato nella direzione di una maggiore regolazione dall'alto, attraverso il ricorso sempre più diffuso a strumenti di *hard law*. Si vedano, in particolare: la direttiva (UE) 2018/1808 del Parlamento europeo e del Consiglio, del 14 novembre 2018, recante modifica della direttiva 2010/13/UE, relativa al coordinamento di determinate disposizioni legislative, regolamentari e amministrative degli Stati membri concernenti la fornitura di servizi di media audiovisivi (direttiva sui servizi di media audiovisivi), in considerazione dell'evoluzione delle realtà del mercato (2018) OJ L61/69; la direttiva (UE) 2019/790 del Parlamento europeo e del Consiglio, del 17 aprile 2019, sul diritto d'autore e sui diritti connessi nel mercato unico digi- tale e che modifica le direttive 96/9/CE e 2001/29/CE (2019) OJ L130/92; il Regolamento (UE) 2021/784 del Parlamento europeo e del Consiglio, del 29 aprile 2021, relativo al contrasto della diffusione di contenuti terroristici online (2021) OJ L172/79. Si veda, da ultimo il c.d. Digital Services Act: Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE, COM (2020) 825.

[226] G. Vasino, *Censura privata e contrasto all'hate speech nell'era delle Internet Platforms,* in *Federalismi*, 8 febbraio 2023, 130-159.

[227] Direttiva 2000/31/CE sul commercio elettronico.

[228] Commission staff working document *Countering racism and xenophobia in the EU: fostering a society where pluralism, tolerance and non-discrimination prevail*, in Brussels, 15.3.2019, SWD (2019) 110 final. Si segnala che l'elaborazione dell'atto è stata preceduta dal Colloquium *Tolerance and respect: preventing and combating anti-Semitic and anti-Muslim hatred in Europe* (internet Forum, dicembre 2015) in cui la Commissione manifestava un interesse all'individuazione di strumenti adeguati a prevenire la diffusione di discorsi discriminatori con specifico riferimento alle condizioni di alcune minoranze (promuovendo un approccio inizialmente fondato, appunto, sul *soft law*). La pubblicazione del Codice è stata seguita dalla emanazione di una comunicazione contenente importanti linee guida relative all'implementazione del Codice (si veda: *Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the Committee of the regions Tackling Illegal Content Online - Towards an enhanced responsibility of online platforms*, Bruxelles, 28 settembre 2017, COM (2017) 555).

presenta profili controversi che meritano di essere analizzati per comprendere se (e come) la moderazione possa rappresentare uno strumento di contrasto efficace.

In primo luogo, è necessario evidenziare la difficoltà di trovare, all'interno degli Stati membri, una definizione condivisa di odio online. Il Codice rimanda alla formulazione contenuta nella Decisione Quadro 2008/913/GAI[229], la quale identifica l'*hate speech* come «*incitement to hatred*» e non come «*incitement to violence, discrimination e hostility*»[230]. A seconda della giurisdizione di riferimento, le condotte ascrivibili ai discorsi d'odio penalmente rilevanti presentano delle differenze considerevoli. A loro volta le piattaforme e gli intermediari digitali tendono a definire autonomamente il concetto di *hate speech* sanzionabile ai sensi dei loro termini e condizioni d'utilizzo, adottando nozioni[231] notevolmente più ampie e aperte rispetto alle fattispecie considerate dai sistemi giuridici nazionali. Il campo di applicazione di tali standard privati rischia di risultare estremamente ampio e, per certi versi, indefinito, con il pericolo di compromettere la lotta al contrasto al fenomeno dell'*online hate speech*[232].

Secondariamente, si deve affrontare il problema della fallibilità dello strumento algoritmico come ausilio alla moderazione e il suo conseguente impatto sui diritti fondamentali degli utenti e delle categorie vulnerabili.

Le piattaforme devono necessariamente fare ricorso a strumenti di moderazione algoritmica complessi e raffinati[233], basati sull'utilizzo di sistemi di intelligenza artificiale e *machine learning*[234]. Tale soluzione, sebbene sia l'unica effettivamente praticabile, cela evidenti rischi connessi al loro utilizzo[235]. I *software* di moderazione automatica dei contenuti sono inevitabilmente soggetti a errore in quanto funzionano sulla base di modelli probabilistici. Ciò significa che, anziché operare con certezza assoluta, questi sistemi effettuano delle previsioni su quale contenuto sia appropriato o meno, con un margine di incertezza che può portare a decisioni imprecise[236]. Il problema è rappresentato dal fatto che l'errore ha un impatto sulle comunità tradizionalmente emarginate

---

229 Per una lettura critica della Decisione Quadro del 2008, si rimanda a T. Moschetta, *La decisione quadro 2008/913/GAI contro il razzismo e la xenofobia: una «occasione persa» per l'Italia?* in *Rivista di Diritto dell'Economia, dei Trasporti e dell'Ambiente*, vol. XII, 2014, 31, ss.

230 B. Bukovska, *The European Commission's Code of Conduct for Countering Illegal Hate Speech Online. An analysis of freedom of expression implications*, in *Transatlantic working group,* 4, ove afferma: «*incitement to hatred makes the proscribed outcome an emotional state or opinion, rather than the imminent and likely risk of a manifested action (discrimination, hostility or violence)*».

231 Si veda, in tal senso, R. Wilson-M. Land, *Hate Speech on social media: Content Moderation in Context,* in *Connecticut Law Review,* 52(3), 2021, 1029-1076.

232 P. Dunn, *Moderazione automatizzata e discriminazione algoritmica: il caso dell'hate speech,* in *Rivista Italiana di Informatica e Diritto,* fasc. 1-2022.

233 J. Grimmelmann, *The Virtues of Moderation*, in *Yale Journal of Law and Technology*, 17, 2015, 42-109.

234 Vi sono tre modelli di moderazione: a) umana (o manuale); b) di intelligenza artificiale (moderazione algoritmica o automatica); c) ibrida (combinazione di a) e b)), in questo caso la funzione dei sistemi di intelligenza artificiale è quella di operare una scrematura preventiva dei contenuti pubblicati dagli utenti e rimettere al moderatore umano solo i casi più ambigui). I sistemi ibridi sono quelli che hanno assunto maggior rilievo.

235 O. Pollicino, *Judicial protection of fundamental rights on the Internet*, in *Oxford*, 2021, 193-194.

236 G. De Gregorio-O. Pollicino-P. Dunn, *Digitisation and the central role of intermediaries in a post-pandemic world*, in *MediaLaws.eu*, 2021; G. Sartor-A. Loreggia, *The impact of algorithms for online content filtering or moderation*, 2020.

e discriminate[237], data la ricorrente presenza di *bias* involontari[238]. Nonostante i tentativi di perfezionamento dei dispositivi, permane un'incapacità della macchina di comprendere il metatesto dell'espressione esaminata. Gli elementi intrinseci del contenuto pubblicato quali, ad esempio, sfumature semantiche[239], o l'uso ironico o satirico di una determinata terminologia, sfuggono allo screening del sistema; in altre parole, con la moderazione online, diversamente rispetto ai discorsi d'odio offline, si rinuncerebbe alla (irrinunciabile) contestualizzazione della condotta.

In sintesi, le numerose disfunzionalità algoritmiche idonee a generare un ampio numero di falsi positivi restano numerose. Se riletto nella cornice della tutela dei diritti fondamentali dell'utente, questo aspetto delineerebbe uno scenario particolarmente grave, anche in considerazione dei plurimi studi[240] multidisciplinari che hanno dimostrato che l'errore del sistema di moderazione automatizzato impatterebbe maggiormente sulle comunità tradizionalmente marginalizzate e discriminate[241]. Gli studi citati sembrano mostrare un approccio dell'algoritmo orientato ad un criterio di uguaglianza formale, teso al mantenimento (e perseguimento) degli equilibri sociali preesistenti, per evitare di creare ulteriori disuguaglianze[242]. Ciò tuttavia, non può essere accettato,

[237] Un interessante approfondimento sul tema dei bias discriminatori negli algoritmi delle piattaforme digitali è stato fornito da S.U. Noble, nel suo libro *Algorithms of Oppression: How Search Engines Reinforce Racisms,* New York, 2018). Come ormai ampiamente documentato dalla letteratura, il margine di errore dei sistemi automatizzati utilizzati per rilevare *l'hate speech* ha un impatto particolarmente significativo sulle comunità tradizionalmente marginalizzate e discriminate. In effetti, sono sempre più numerosi gli studi che si concentrano sullo sviluppo di tecniche di *debiasing* per i moderatori automatici, ma il problema rimane lontano dalla soluzione. Per esempio, è stato osservato che i contenuti pubblicati da membri delle comunità afroamericana o LGBTQIA+ sono spesso ingiustamente penalizzati per presunta violazione delle normative contro l'hate speech o il toxic speech. Le cause di questo fenomeno sono molteplici. Ad esempio, i dataset utilizzati per addestrare gli algoritmi non sono sempre di alta qualità, soprattutto perché non riflettono adeguatamente il linguaggio e le modalità comunicative proprie di questi gruppi minoritari. Inoltre, molte volte i gruppi marginalizzati utilizzano termini ed espressioni che, seppur potenzialmente offensivi in altri contesti, sono intrinsecamente parte della loro comunicazione. L'incapacità degli algoritmi di cogliere queste sfumature e il loro significato implica un alto rischio di errori, con la possibilità di falsi positivi.

[238] G. Ziccardi, *Odio online. Violenza verbale e ossessioni in rete*, Milano, 2016, 97 ss.

[239] La non sensibilità dell'algoritmo è stata oggetto di numerosi studi, in tempi in cui ancora non si faceva largo uso dei sistemi di IA, finalizzati ad analizzare la capacità del sistema di rilevare correttamente tweet razzisti nei confronti della comunità afroamericana, cfr., ad esempio, I. Kwok-Y. Wang, *Locate the Hate: Detecting Tweets against Blacks,* in *Proceedings of the AAAI Conference on Artificial Intelligence*, Washington, DC, USA, vol. 27, 2013.

[240] P. Dunn, *Moderazione automatizzata*, cit., 137, il quale osserva come tale effetto distorsivo si rileva anche a livello di *content curation* e non solo nell'*hard moderation* in senso stretto. Più specificamente, è stato sottolineato in dottrina come il meccanismo algoritmico tenda a perseguire una generale massimizzazione dell'engagement dei gruppi di maggioranza con l'effetto di ridurre sempre di più gli spazi riservati a gruppi marginalizzati.

[241] Questa tendenza, registrata in relazione a diversi gruppi minoritari quali, ad esempio, la comunità LGBTQ+85 e la comunità afroamericana si scontrerebbe con il meccanismo di funzionamento dell'algoritmo. Le principali disfunzionalità che ancora si rilevano si sostanziano in *lexical bias* (meccanica associazione di una parola a un contenuto tossico) e *dialectal bias* (i quali derivano dall'incapacità del sistema di cogliere l'uso peculiare che si fa di un determinato lessico all'interno di una subcategoria linguistica o dialetto).

[242] Si segnalano al riguardo per completezza le interessanti riflessioni e le proposte correttive, basate sui meccanismi di *bias transfroming*, di S. Watcher-B. Mittelstadt-C. Russel, *Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law*, in *Virginia Law Review*, 123(3),

comportando altrimenti un secondario effetto discriminatorio molto superiore per alcune minoranze a base razziale o fondato sull'orientamento sessuale[243]. In definitiva, l'effetto paradossale dell'utilizzo della moderazione algoritmica, seppur indispensabile, consisterebbe nel rovesciamento del principio di uguaglianza sostanziale sotteso al percorso normativo euro-unitario, modellato sull'accrescimento delle tutele e finalizzato ad eliminare ogni forma di disparità di trattamento[244].

Gli aspetti analizzati, seppur brevemente, hanno evidenziato come la disciplina di moderazione del Codice possa avere un impatto sui diritti fondamentali degli utenti. La Commissione europea ha mostrato una crescente consapevolezza in tal senso, come documentato dal Regolamento (UE) 2021/784[245]. Anche il *Digital Services Act* contiene alcune disposizioni in tal senso, come l'articolo 12, il quale richiede agli intermediari di applicare condizioni generali dei loro servizi in modo «equo, trasparente, coerente, diligente, tempestivo, non arbitrario, non discriminatorio e proporzionato», nel pieno rispetto dei diritti e degli interessi legittimi di ogni parte coinvolta. Inoltre, l'articolo 17 impone alle piattaforme online di predisporre sistemi interni di gestione dei reclami da attuarsi «in modo tempestivo non discriminatorio, diligente e non arbitrario». Sebbene tali disposizioni rappresentino un sicuro passo in avanti, da più parti è stato rilevato come in realtà raffigurino petizioni di principio, non essendo corredate di apparati applicativi definiti e sviluppati. Il sistema introdotto dal DSA incentiverebbe un incremento dell'utilizzo su vasta scala di sistemi di moderazione automatizzati senza, tuttavia, prevedere adeguati rimedi a tutela dell'individuo[246] né, tantomeno, sufficiente attenzione con riferimento al rischio di un'iniqua rimozione dei contenuti.

Secondo l'opinione di chi scrive, il nodo cruciale della moderazione risiede nell'individuazione della soglia oltre la quale il margine di errore degli strumenti algoritmici di moderazione automatica (falso positivo) non può più essere considerato accettabile. In conclusione, sebbene tali strumenti siano validi e, anzi, indispensabili, essi da soli non sono sufficienti. La soluzione, provvisoria e necessariamente legata al contesto storico, non può che consistere nell'integrazione tra diritto e tecnologia, alla luce dei risultati emersi dalla discussione pubblica.

2021, 735 ss.

243 Si segnalano alcune analisi statistiche, le quali sembrano dimostrare come l'incapacità dello strumento di moderazione automatizzata di cogliere il contesto porti ad un maggior numero di contenuti eliminati di utenti appartenenti alla comunità delle *drag queen* rispetto a contenuti discriminatori di nazionalisti bianchi, cfr. O. Dias Thiago-D. M. Antonialli-A. Gomes, *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, in *Sexuality and Culture*, 25, 2021, 700 ss.

244 G. Vasino, *Censura privata,* cit. 146-147.

245 Il Regolamento all'art. 5 prevede che un fornitore di servizi riconosciuto come esposto a contenuti terroristici debba predisporre misure specifiche volte a contrastarne la diffusione: nell'applicare tali misure, tuttavia, dovrà tenere pienamente conto dei diritti e degli interessi legittimi degli utilizzatori (ivi inclusa la libertà di espressione e di informazione) e, allo stesso tempo, agire in maniera diligente e, soprattutto, non discriminatoria. È inoltre disposta, all'art. 10, la predisposizione di meccanismi di reclamo a tutela degli utenti i cui contenuti siano stati rimossi, con l'obbligo per il fornitore di rendere decisioni motivate e fatto salvo l'eventuale ricorso all'autorità amministrativa o giudiziaria dello Stato.

246 J. Barata, *The Digital Services Act and Its Impact on the Right to Freedom of Expression: Special Focus on Risk Mitigation Obligations*, in *DSA Observatory*, 27 July 2021.

7. I rapporti tra diritto penale, libertà di manifestazione del pensiero nell'era digitale e democraticità

Dal quadro sin qui delineato emerge chiaramente l'assunzione di una dimensione e di una rilevanza pubblicistica da parte dei PSI, che si evince chiaramente dalla definizione recentemente data ai social network di «un pezzo di infrastruttura democratica immateriale». In un Paese democratico, non è accettabile che i crimini d'odio online vengano classificati e trattati come mere vicende private, gestite a livello contrattuale tra PSI e utenti, dal momento che le dette piattaforme costituiscono strumenti di cittadinanza essenziali per la democrazia e spazi globali di confronto[247]. Limitazioni alla libertà di espressione sono ammissibili esclusivamente se imposte da un ente statale e non per ottemperare a logiche aziendali[248].

La migrazione da un "Internet delle reti" ad un "Internet delle piattaforme"[249], ha accentrato nelle mani di pochi attori privati (si badi che le più grandi piattaforme digitali occupano posizioni oligopolistiche o di monopolio sul *social marketplace of ideas*[250]) una quantità di potere ordinativo (contrattuale ed economico) che si configura come autosufficiente e autonomo. Ne deriva che il cyberspazio aspirerebbe a rendersi immune da interventi pubblicistici. Di qui, la precarietà e fragilità delle libertà degli utenti[251].

Si pone, di conseguenza, il problema di trasferire "la Costituzione nella rete"[252]. Ciò richiede di procedere a una costituzionalizzazione delle interazioni che hanno luogo sul web per garantire non solo una società reale democratica, ma anche una società digitale democratica. La rivoluzione informatica sta modificando l'idea stessa di realtà e la rete è reale prima ancora che virtuale; essa non si può concepire come semplice estensione immateriale delle dinamiche della società, bensì come luogo elettivo per la costruzione della società civile[253].

L'esigenza di strutturare e presidiare i fenomeni digitali in maniera costituzionalmente

---

247 P. Falletta, *Controlli e responsabilità dei social network sui discorsi d'odio online*, cit., 148.

248 Per un maggiore approfondimento sul tema vedasi V. Claussen, *Fighting hate speech and fake news*, cit., 135 ss.

249 C. Confortini, *Diffamazione e discorso d'odio in internet*, cit., 694, che pone in risalto la progressiva crescita di un'economia basata sui dati c.d. «data driven economy» con richiamo, altresì, alla Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni «Costruire un'economia dei dati europea», COM/2017/09.

250 Cfr. K. Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, in *Harvard Law Review*, 131, 2018, 1630. Vedasi anche F. Pasquale, *Privacy, Antitrust, and Power*, in *George Mason Law Review*, 4, 2013, 20, 1015 ss.; L.M. Khan, *Amazon's Antitrust Paradox*, in *Yale Law Journal*, 126, 2017, 710 ss. e L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico. Fake news, hate speech e profili di responsabilità dei social network,* in *Federalismi.it*, 26, 2021, 9.

251 C. Confortini, *Diffamazione e discorso d'odio in internet*, cit., 702 che evidenzia come, al fine di contemperare i beni in discussione, il *Digital Services Act* abbia fortemente inciso sul contratto stipulato tra la piattaforma digitale e l'utente, il quale, pur rimanendo espressione di un potere di tipo privatistico, viene integrato nel suo contenuto, allo scopo di tutelare un bene di rango superiore, l'ordine giuridico del mercato, nell'ambito del quale la tutela della concorrenza si coniuga (incidentalmente) con la protezione dei diritti fondamentali della persona.

252 F. Oliveri, *Diritti degli internauti, obblighi degli Stati, responsabilità delle piattaforme digitali*, cit., 116.

253 M. Ferraris, *Metafisica del Web*, lezione tenuta presso il Centro Nexa, Politecnico di Torino, in nexa.polito.it, 08.01.2020,

conforme è proporzionale al potere acquisito dalle grandi società informatiche nel plasmare l'ecosistema web, potere che richiede parimenti di essere democratizzato. Di riflesso, il diritto a essere protetti dai discorsi d'odio online deve trovare fondamento all'interno della nuova democrazia digitale e deve costituire oggetto di bilanciamento con la libertà di informazione, di espressione e di comunicazione, nonché con i diritti posti a presidio dell'ordinamento democratico plurale e multiculturale che caratterizzano la fisionomia di ogni Stato di diritto.

La stessa Corte di Strasburgo per oltre vent'anni ha seguito un approccio definito "militante"[254] che, mediante l'adozione della tecnica decisionale tipica dei Paesi di *Common Law*, ossia il metodo *case by case*[255], prevede un bilanciamento quanto più adeguato e effettivo tra la libertà di espressione e la tutela del diritto alla dignità e a non subire discriminazioni. In particolare, la Corte EDU ha tenuto conto in maniera omnicomprensiva del materiale illecito commesso, in piena ottemperanza ai principi di offensività e di materialità del diritto penale, cristallizzati nel nostro quadro positivo all'art. 25, c. 2, Cost. In tale ottica, il ricorso allo strumento repressivo forte appare giustificato al fine di presidiare i fondamentali valori democratici dell'uguaglianza e del pluralismo e deve essere finalizzato a punire qualsiasi condotta che - fermi restando gli essenziali principi sopra citati - sia in grado di arrecare pregiudizio a tali valori. I giudici di Strasburgo sono giunti, così, ad attribuire "carattere eccezionale" ai discorsi d'odio, in forza del quale il vaglio di legittimità delle misure restrittive della libertà di espressione assume maglie più larghe[256]. Di qui, seguendo una linea di severo contrasto del fenomeno, la Corte EDU ammette il ricorso al diritto penale per contrastare l'*online hate speech*.

Appare fuori discussione come, a tutt'oggi, spazi di condivisione quali i social network si concretizzino in luoghi di esercizio del diritto di manifestare liberamente il proprio pensiero[257], sancito dall'art. 10 CEDU (che, al par. 2, ne disciplina le eventuali restrizioni), dagli artt. 11 e 21 della Carta dei diritti fondamentali dell'UE (che la sancisce, nello specifico, nella accezione relativa al principio di non discriminazione)[258] e, non da ultimo, dall'art. 21 della Costituzione italiana.

Anche la nostra Corte costituzionale ha affrontato il tema della repressione penale delle c.d. "parole pericolose"[259], che sono state storicamente inquadrate nell'ambito dei reati posti a tutela dell'"ordine pubblico", come, a titolo esemplificativo, le norme volte al contrasto delle condotte discriminatorie[260]. Il nostro legislatore ha tentato di

[254] C. Caruso, *L'hate speech a Strasburgo: il pluralismo militante del sistema convenzionale*, in *Quaderni costituzionali*, 37, 4, 2017, 963-984.

[255] Su questo aspetto, cfr. V. Cinà, *Libertà di espressione e importanza del contesto: la Corte europea dei diritti dell'uomo ridefinisce il perimetro della protesta politica*, in *La Nuova giurisprudenza civile commentata,* 6, 2021, 1379 ss.

[256] P. Dunn, *Carattere eccezionale dell'"hate speech" e nuove forme di responsabilità per contenuti di terzi nella giurisprudenza EDU*, cit., 249.

[257] C. Confortini, *Diffamazione e discorso d'odio in internet*, cit., 698.

[258] P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet ei social media*, cit., 29.

[259] Fin dalla sentenza della Corte cost., 26 gennaio 1957, n. 1 avente ad oggetto la previsione della l. 645 del 1952, *Norme di attuazione della XII disposizione transitori e finale della Costituzione*, che incrimina le condotte di apologia di fascismo.

[260] Vedasi, per un maggiore approfondimento, G. Puglisi, *La parola acuminata. Contributo allo studio dei*

condizionare l'esercizio della libertà di manifestazione del pensiero sotto lo spettro del bene giuridico "ordine pubblico", a cui è stata inizialmente ispirata l'introduzione di fattispecie che incriminano forme di incitamento e di istigazione a commettere reati. È questo il caso della diffusione di ideologie fondate sull'incitamento all'odio razziale che hanno posto la questione dell'idoneità della libertà di pensiero a impedire la repressione di manifestazioni aggressive di dissenso ideologico idonee a negare i valori di fondo della convivenza civile[261].

Merita di essere menzionata la sopravvivenza dei c.d. "reati di opinione", contenuti nel Codice Rocco e concepiti dalla Consulta sotto forma di limiti esterni all'art. 21 Cost., che hanno reso necessaria l'enucleazione di una serie di principi per verificare la conformità a Costituzione degli strumenti posti a tutela di una democrazia minacciata da linguaggi discriminatori[262]. L'opera della Corte costituzionale, mossa verso la salvaguardia dell'impianto repressivo del dissenso, ha individuato gli argini irrinunciabili della libertà di espressione, sia in termini di contenuti del messaggio informativo, sia di diffusione dello stesso attraverso qualsiasi mezzo[263].

Il passaggio a un "Internet delle piattaforme" ha fatto sì che, se, in un primo momento, i social media sono stati concepiti come spazio di agevolazione del dibattito pubblico[264], essi sono, a tutt'oggi, assurti al rango di "pietra angolare dell'ordine democratico"[265].

Pertanto, sebbene, originariamente, la repressione dei discorsi d'odio sia stata giustificata in forza della nozione di "ordine pubblico", attualmente appare maggiormente corretto inquadrarli nel novero delle componenti atte a garantire la solidità della democrazia nella dimensione digitale (forse, ormai, si potrebbe dire, nell'era "post" digitale). A conferma di ciò, il parametro utilizzato dalla Consulta è divenuto il più elastico principio del pluralismo, inteso quale distinta e ulteriore libertà che trova specifica cristallizzazione nell'art. 21 Cost. e che garantisce, altresì, il diritto ad «esteriorizzare liberamente il proprio pensiero con la parola, lo scritto e ogni altro mezzo di diffusione»[266]. Peraltro, tale diritto assolve non soltanto un'esigenza di carattere individuale, ma anche una vera e propria "funzione sociale", assicurando alla collettività il contributo del pensiero di tutti i consociati, oltre che la libera discussione e il confronto tra idee, perseguendo interessi pubblici e privati diversi da quelli del solo titolare[267].

---

delitti contro l'eguaglianza, tra aporie strutturali ed alternative alla pena detentiva, in *Rivista italiana di diritto e procedura penale*, 2018, 1329 ss.

261 L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico,* cit., 10-11.

262 *Ibid.*

263 Ivi, 3.

264 P. Falletta, *Analisi normativa in tema di contrasto agli hate speech su Internet e i social media*, cit., 23.

265 Così Corte cost., 17 aprile 1969, n. 84 e 04 febbraio 1965, n. 9. Per una definizione analoga della libertà di espressione, inserita «fra i diritti primari e fondamentali» nella giurisprudenza costituzionale italiana vedasi Corte cost. 28 gennaio 1981, n. 1 che la definisce «cardine del regime democratico».

266 Così, L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico,* cit., 5, richiamando i seguenti precedenti giurisprudenziali: Corte cost., 14 luglio 1988, n. 826, 07 dicembre 1994, n. 420 e 18 ottobre 2002, n. 422.

267 L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico,* cit., 8.

Di talché, i sostenitori della necessità del diritto di cittadinanza dei "reati di opinione" nel nostro ordinamento valorizzano la protezione fornita a beni costituzionalmente rilevanti da forme di tutela anticipata, vista la pericolosità insita nell'*hate speech* di compromettere i processi di integrazione dei valori fondanti la convivenza civile[268].
In senso parzialmente difforme, la Corte EDU ha per molto tempo richiamato la "funzione didattico-terapeutica"[269] che può essere dispiegata dalla pubblica manifestazione del pensiero, seguendo l'approccio della dottrina statunitense che ammette la legittimità di ogni manifestazione, anche dichiaratamente espressiva di odio razziale o etnico, a condizione che non sia tale da produrre un pericolo chiaro ed imminente che giustifichi l'intervento repressivo dello Stato[270]. Tale logica non è molto diversa da quella seguita dalla Corte costituzionale italiana e dalle successive pronunce dei tribunali di merito nell'opera di adeguamento della portata applicativa delle diverse ipotesi di istigazione, che sono state limitate laddove, per eccessiva ampiezza e genericità, si potesse dire consentito incriminare condotte tutelate dall'art. 21 Cost., muovendo dalla distinzione fra ciò che è "pensiero puro" e ciò che, invece, è "principio in azione"[271]. Così, ad esempio, la Corte ha rilevato la contrarietà all'art. 21 dell'istigazione alla disobbedienza delle leggi di ordine pubblico e dell'odio fra le classi sociali di cui all'art. 415 c.p., nella parte in cui quest'ultima non specificava che l'istigazione, per essere punita, oltre che una pubblica esternazione, deve comportare un concreto pericolo per la pubblica incolumità[272]. Merita una menzione il fatto che, in relazione al primo reato citato, la Consulta abbia ravvisato anche la violazione dell'art. 25, c. 2, Cost. Ciò in quanto la fattispecie è indeterminata e prescinde da un qualsiasi esame circa l'effettivo verificarsi di un pericolo per l'interesse tutelato, non precisando le modalità con cui l'istigazione si potrebbe distinguere dalla diffusione della persuasione di ideologie e dottrine politiche, sociali, filosofiche o economiche, e, quindi, penalmente perseguibile, senza violare il precetto costituzionale dell'art. 21 Cost. Ne deriva che la Consulta, nel solco del c.d. "diritto penale materiale"[273], ha attribuito rilievo alle concrete e specifiche modalità con le quali è stata compiuta l'istigazione, allo scopo di indagare laddove la stessa rivesta carattere di effettiva pericolosità per l'esistenza di beni costituzionalmente protetti e integri un comportamento concretamente idoneo a promuovere discorsi d'odio. Sol-

[268] M. Manetti, *L'incitamento all'odio razziale tra realizzazione dell'eguaglianza e difesa dello Stato*, in *Scritti in onore di Gianni Ferrara*, Padova, 2006, 103 ss.

[269] L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico,* cit., 11.

[270] Vedi, sul punto, considerazioni svolte nel paragrafo 5.

[271] L. Califano, *La libertà di manifestazione del pensiero…in rete; nuove frontiere di esercizio di un diritto antico,* cit., 11 e 12 che richiama Corte cost. 16 marzo 1962, n. 19 nella parte in cui afferma che «se per turbamento dell'ordine pubblico bisogna intendere l'insorgere di un concreto ed effettivo stato di minaccia per l'ordine legale, mediante mezzi illegali idonei a scuoterlo (…) è chiaro che non possono essere considerate in contrasto con la Costituzione le disposizioni legislative che effettivamente ed in modo proporzionato siano volte a prevenire e reprimere siffatti turbamenti».

[272] Corte cost., 5 aprile 1964, n. 108 e 5 giugno 1978, n. 71.

[273] Del quale si può parlare, all'interno del nostro ordinamento penalistico, in forza del divieto di punire l'intenzione (*cogitationis poenam nemo patitur*), cristallizzato all'art. 25, c. 2, Cost.; sul punto, si riportano di seguito gli orientamenti della giurisprudenza costituzionale maggiormente rilevanti: Corte cost., 02 novembre 1996, n. 370 sull'art. 708 c.p. e 05 luglio 2010, n. 249 sulla qualità di immigrato «irregolare».

tanto quando l'istigazione è diretta a commettere gli atti che costituiscono – secondo una valutazione legislativa definita immune da irragionevolezza – un pericolo per il bene costituzionalmente protetto e non si limita a una mera critica di fatti specifici, per i quali si esercita democraticamente il controllo dell'opinione pubblica, tali condotte assumeranno rilevanza penale[274].

In sintesi, un diritto penale costituzionalmente orientato impone l'incriminazione dei discorsi d'odio nella misura in cui questi ultimi possano configurarsi come la premessa di un'azione delittuosa[275] volta a ledere la collettività di appartenenza del destinatario, compromettendone la coesione e pregiudicandone l'ordine democratico eterogeneo.

8. Riflessioni conclusive: il ricorso al diritto penale, utopia od opportunità?

Alla luce delle considerazioni di cui sopra, pare irrinunciabile svolgere alcune riflessioni conclusive circa il quadro tracciato in punto di rilevanza penale dei discorsi d'odio online.

Le condotte d'odio assumono nell'ecosistema digitale, una dimensione transfrontaliera e transnazionale che amplifica enormemente gli effetti pregiudizievoli del messaggio illecito e rende estremamente difficoltosa l'adozione di efficaci strumenti per contrastarlo. Di conseguenza, appare ancora più impellente la necessità che l'Unione europea adotti direttive in questa materia, atteso che tali condotte, in ragione della loro portata, sono in grado di assumere un elevatissimo grado di lesività per i beni giuridici coinvolti. Ne discende che i discorsi d'odio online possono essere ricondotti alla categoria della "criminalità informatica" di cui all'art. 83, par. 1, TFUE, in riferimento alla quale le istituzioni europee detengono una competenza legislativa "diretta". A ciò si aggiunga che la Commissione ha a lungo esaminato la possibilità di punire tali condotte ricorrendo alla competenza "indiretta" in materia penale dell'Unione europea, ai sensi dell'art. 83, par. 2, TFUE.

Tuttavia, nessuna di queste soluzioni di intervento legislativo è stata intrapresa e, allo stato attuale, l'inerzia del legislatore europeo nell'incriminare a livello sovranazionale questi comportamenti non è più tollerabile. L'esigenza che si impone è quella di realizzare il tanto auspicato quadro unionale di tutela giuridica, anche ai fini del ravvicinamento dei quadri degli Stati membri. L'adozione di un efficace sistema di repressione non può prescindere da un'opera di preparazione socioculturale che miri a far comprendere il grado di pericolosità insito nelle condotte in esame e che consenta di coniugare l'utilizzo della tecnica penalistica con il formante tecnologico e con quello sociale, al fine di radicare la cultura di contrasto ai discorsi d'odio online. È, quindi, necessario uno sforzo coordinato tra queste componenti per contenerne la circolazione e mitigare i fenomeni pregiudizievoli che ne derivano. Diventa, quindi, essenziale promuovere la consapevolezza e l'educazione digitale di tutte le fasce della popolazione. A tal fine, la politica legislativa, sia a livello nazionale che europeo, riveste un ruolo di primaria

[274] Corte cost., 5 giugno 1978, n. 71.

[275] L. Califano, *La libertà di manifestazione del pensiero,* cit., 12 e 13.

importanza.
Avendo riguardo al quadro normativo italiano, le fattispecie analizzate non sono in grado di stigmatizzare in modo efficace il diffondersi dei discorsi d'odio online. Tuttavia, lo strumento del diritto penale, impiegato nel doveroso rispetto dei principi di tassatività, precisione e determinatezza, corollari ineludibili del principio di legalità, continua a rappresentare un baluardo irrinunciabile per punire quei comportamenti che ledono beni giuridici fondamentali. In questo contesto, la scelta di prevedere l'incriminazione dei discorsi d'odio (anche online) mediante la tecnica incriminatrice dei reati a pericolo concreto, consente di evitare il rischio di un possibile conflitto con il principio di offensività, il quale potrebbe essere compromesso qualora si applicasse una sanzione penale a una condotta concretamente inoffensiva, per i mezzi, le modalità e i destinatari. In un ambito così sensibile, in cui è necessario bilanciare accuratamente i beni giuridici coinvolti, inquadrare il reato tra quelli a pericolo concreto contribuisce a salvaguardare il rispetto della Costituzione.
Nello specifico, l'opera della Consulta si è incentrata sulla salvaguardia dell'impianto repressivo del dissenso, al fine di preservare i beni giuridici del pluralismo, dell'uguaglianza, della dignità umana e della democrazia. Analogamente a quanto riscontrato negli ulteriori quadri giuridici analizzati, anche la nostra Corte costituzionale si è occupata di indagare la linea di demarcazione tra l'espressione di pensiero puro e una manifestazione che, al contrario, racchiude in sé non una mera divulgazione, bensì una condotta in azione, idonea a porre in pericolo l'oggettività giuridica tutelata. Soltanto quest'ultima condotta si connota per una duplicità di motivazioni e di destinatari: non è rivolta all'individuo singolarmente inteso, ma piuttosto alla collettività di appartenenza che viene colpita tramite il singolo.
Altresì dallo studio dei sistemi europei analizzati è emerso il favore verso l'utilizzo dello strumento del diritto penale, a discapito di altri possibili, allo scopo precipuo di punire specificamente le condotte di *hate speech*, anche nella dimensione online. L'elevata gravità di queste condotte, potenzialmente idonee a compromettere il valore super-individuale su cui si regge la collettività, ha indotto a far ritenere preferibile la via dell'utilizzo di strumenti eccezionali di tutela avanzata.
Tuttavia, nei Paesi che hanno effettuato scelte di incriminazione, estremamente insidiosa è apparsa la strutturazione dell'ecosistema penale all'interno della complessa realtà digitale. Ciò in quanto si è ravvisata la contestuale necessità di garantire il rispetto dei principi di matrice costituzionale che sorreggono la tecnica penalistica. Primo tra tutti viene in discussione il principio di offensività, atteso che, nella selezione dei fatti penalmente rilevanti, sia il legislatore che il giudice, devono punire le sole condotte idonee a ledere, ovvero porre in pericolo, i beni giuridici tutelati. Viene poi in rilievo il principio di legalità, di cui all'art. 25, comma 2, Cost., che si declina (per quanto più di interesse) nel principio di riserva di legge e di determinatezza, tassatività e precisione. Non da ultimo, occorre tenere in considerazione anche il principio di materialità del diritto penale, il cui contenuto massimo si sostanzia nel brocardo latino "cogitationis poenam nemo patitur", in virtù del quale nessun individuo può essere punito per aver concepito nella sola forma del pensiero la commissione di un fatto criminoso che non si sia tradotto in azione, ovvero non sia stato messo in atto.

Nel quadro giuridico comparato è emerso che la complessità della disciplina dettata e l'indeterminatezza delle fattispecie incriminatrici introdotte, concepite principalmente secondo le coordinate del modello istigatorio, hanno condotto, nella prassi, alla mancata applicazione di queste ultime a vantaggio delle fattispecie comuni. È quanto è accaduto in Germania e Spagna. Si può, quindi, rilevare che non sempre a uno sforzo legislativo sistematico o, finanche, definito di carattere "olistico", corrisponda l'efficacia della disciplina, la cui eccessiva forza espansiva compromette l'intellegibilità del sistema, al pari di quanto verificatosi in Francia.

Da ultimo, ai fini dell'accertamento della responsabilità degli autori, la giurisprudenza francese, tedesca e spagnola rimanda a un vaglio di proporzionalità, necessarietà e adeguatezza da eseguirsi nel caso specifico, in cui si prendano in considerazione sia i diritti fondamentali che vengono in rilievo, sia il rispetto dei principi che permeano la tecnica penalistica e la cui operatività deriva anche dall'effetto di irradiazione dei primi. Anche il cammino della Corte di Giustizia dell'Unione europea e della Corte di Strasburgo, nella faticosa opera di individuazione degli argini irrinunciabili della libertà di espressione, e la disamina offerta dal quadro normativo americano hanno condotto a esiti analoghi a quelli della Consulta italiana. Il diritto penale deve, necessariamente, continuare a essere utilizzato nel contrasto ai discorsi d'odio online come strumento principale, seppur integrato nell'ambito di una strategia complessa che passi anche attraverso la prevenzione.

Da questo punto di vista, la moderazione algoritmica realizzata dai PSI si presenta come uno strumento volto a realizzare un evidente vantaggio sociale, determinato dalla riduzione dell'inquinamento dell'ecosistema informazionale digitale. Come visto, ciò che deve essere necessariamente tenuto in debita considerazione è la fallibilità dei meccanismi di moderazione: un certo margine di errore è purtroppo ineludibile. Il rischio è rappresentato dal fatto che la moderazione attuata dalle piattaforme si riveli un silenziamento delle categorie già discriminate. Per evitare di tradire la *ratio* che ne giustifica l'utilizzo, è necessario pretendere una soglia di accettabilità dell'errore più esigente. Elaborare una strategia efficace di contrasto al fenomeno dell'odio significa necessariamente mettere in relazione le coordinate ermeneutiche del diritto e della tecnologia, alla luce del contesto sociale in cui ci si trova. Al riguardo, una soluzione valida oggi potrebbe non esserlo tra alcuni anni. Sarà compito primariamente del formante politico-legislativo e di quello giurisprudenziale individuare un equilibrio nell'ideazione e nell'interpretazione delle norme penali, al fine di preservare la flessibilità necessaria a fronteggiare temi come quello in esame, soggetto a mutamenti sempre più repentini.

In conclusione, nell'era delle democrazie digitali, emerge con chiara evidenza una c.d. "eterogenesi dei fini": se da un lato è indubbio che le piattaforme costituiscono una componente dell'infrastruttura democratica e favoriscono la libertà di manifestazione del pensiero, quest'ultima, nata come libertà da possibili ingerenze del potere statale, è assurta al rango di «diritto coessenziale al regime di libertà garantito dalla Costituzione»[276] e subisce i rischi connessi alle mutate modalità di diffusione delle idee tipiche della modernità tecnologica.

A fronte di ciò, seppure non si possa negare la "funzione didattico-terapeutica" svolta

[276] Corte cost., 23 marzo 1968, n. 11.

dalla censurabilità etica del web, occorre considerare che la Corte di Strasburgo, al pari della nostra Consulta, ha rilevato come il ricorso allo strumento pesante del diritto penale possa dirsi giustificato per la tutela di fondamentali valori democratici. Ciò in quanto i discorsi d'odio, in generale e ancor di più quelli che si dispiegano online, presentano un *quid pluris* rispetto alle altre condotte criminose, idoneo a compromettere i valori fondamentali dell'agire democratico plurale nell'era digitale.

La rete è divenuta il «centro di un grande motore di democrazia»[277] e, in tale contesto, l'*online hate speech* riveste uno *status* peculiare, in quanto la gravità e la pericolosità insite in esso per i valori fondamentali della convivenza civile costituiscono di per sé un giustificativo per il ricorso alla tecnica penalistica. Questa è l'unica in grado di difendere il fulcro immutabile di norme fondamentali garantiste su cui si regge il nostro sistema costituzionale, secondo lo schema del "diritto penale minimo", che persegue il fine di razionalizzare e minimizzare la violenza dell'intervento punitivo, vincolandolo a limiti rigidi imposti a tutela dei diritti della persona[278].

A ciò dovrebbe senz'altro aggiungersi l'adozione di un approccio chirurgico[279] nella progettazione delle fattispecie incriminatrici e nella loro interpretazione, finalizzato a garantire il rispetto dei canoni di legalità, offensività e materialità che permeano l'ordinamento penalistico.

Soltanto un diritto penale costituzionalmente orientato può garantire la repressione di condotte effettivamente e gravemente pregiudizievoli, non soltanto della dignità della vittima, bensì, anche della sopravvivenza della democrazia eterogenea e multiculturale cui essa appartiene. Del resto, «sostituire il diritto penale con qualcosa di meglio potrà avvenire solo quando avremo sostituito la nostra società con una società migliore»[280].

[277] L. Califano, *La libertà di manifestazione del pensiero,* cit., 6.

[278] L. Ferrajoli, *Cos' è il garantismo*, in *Criminalia*, 2014, 130.

[279] G. Ziccardi, *Il contrasto dell'odio online: possibili rimedi*, cit., 38.

[280] A. Baratta, *Criminologia critica e critica del diritto penale: introduzione alla sociologia giuridico-penale*, Milano, 2019, 1982.

Prime osservazioni sul rapporto tra libertà religiosa e intelligenza artificiale, a partire dall'AI Act*

Martina Palazzo

Abstract

La diffusione delle tecnologie che impiegano l'Intelligenza Artificiale (AI) desta l'interesse e talvolta la preoccupazione di numerosi studiosi, a causa della manifesta pervasività del fenomeno. Il diritto e, in particolare, il diritto ecclesiastico non fa eccezione, dovendosi porre i quesiti necessari per fornire protezione e garanzie agli individui, a livello nazionale e sovranazionale. Questo contributo si pone l'obiettivo di restituire un sintetico quadro di alcune possibili implicazioni che può determinare lo sviluppo dell'AI rispetto alla libertà di pensiero, coscienza e religione, con un focus specifico sulla normativa vigente e con alcune suggestioni per la prospettiva futura.

The spread of technologies employing Artificial Intelligence (AI) has become a concern for numerous scholars across various fields, due to the pervasive nature of the phenomenon. Law, and in particular law and religion, is no exception, as it must address the necessary questions to provide protection and guarantees to individuals at both national and supranational levels. This contribution aims to provide a concise overview of some possible implications that AI may have on freedom of thought, conscience, and religion, with specific attention on current legislation and some suggestions for future perspectives.

Sommario

Keywords

Intelligenza Artificiale – libertà religiosa – diritti fondamentali – rischi – possibilità

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

1. Introduzione: Intelligenza Artificiale e diritti fondamentali

L'Intelligenza Artificiale costituisce un fenomeno pervasivo e in costante crescita, in grado di condizionare e, per certi versi, rivoluzionare i paradigmi concettuali che descrivono e orientano le attività umane[1].

Gli strumenti algoritmici, in virtù delle loro rapidità, efficienza e potenzialità, sono impiegati oramai in un gran numero di ambiti, con importanti ricadute economiche e sociali che gli operatori del diritto e le istituzioni non possono (e non hanno interesse a) trascurare.

In particolare, uno dei dibattiti che si sono sviluppati negli ultimi anni riguarda le implicazioni dell'AI sui diritti fondamentali perché, se è vero che essa può produrre importanti benefici per l'umanità[2], è altrettanto vero che alcuni suoi impieghi potrebbero rappresentare una minaccia per la protezione degli individui[3].

---

[1] Con C. Casonato, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, in *DPCE online*, 1, 2022, 169 ss., «la AI è un ambito di ricerca scientifica e di applicazione tecnologica che presenta caratteristiche connotative uniche. Essa ha già inciso su molte delle nostre attività quotidiane, e ancor di più lo farà in futuro, ad un ritmo e con un impatto formidabili». Secondo P. Contucci, *Intelligenza artificiale tra rischi e opportunità*, in *il Mulino*, 4, 2019, 637 ss., spec. 640, si tratta di una vera e propria «rivoluzione con effetti dirompenti nell'intera società». Si è pervenuti così alla creazione della c.d. società *onlife*, come coniata da L. Floridi, *The Online Manifesto. Being Human in a Hyperconnected Era*, Cham, 2015.

[2] M. Fasan, *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in *DPCE online*, 1, 2022, 181 ss., spec. 184, illustra come «[i]n primo luogo, i sistemi artificiali possono contribuire a ridurre significativamente i tempi normalmente richiesti per l'adozione di determinate decisioni o per lo svolgimento di determinate funzioni, garantendo, quindi, benefici in termini di maggiore rapidità grazie alla potenza computazionale di analisi delle informazioni e alla capacità di individuare correlazioni rilevanti tra i dati esaminati. In secondo luogo, l'AI costituisce uno strumento vantaggioso nel migliorare l'efficacia delle soluzioni proposte, sia nella prospettiva di ridurne possibili margini di errore sia nell'ottica di personalizzarne i contenuti e renderli sempre più adatti agli interessi e ai desideri delle persone. Infine, i sistemi di AI possono contribuire in modo significativo dal punto di vista economico, diminuendo il costo dei processi predittivi e decisionali e incentivando, grazie alle dimostrate capacità di analisi e di correlazione dei dati, una corretta ed equilibrata allocazione delle risorse disponibili». Dello stesso avviso B. Custers-E. Fosch-Villaronga, *Humanizing Machines: Introduction and Overview*, in B. Custers-E. Fosch-Villaronga (a cura di), *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, The Hague, 2022, 3 ss., spec. 10, che affermano: «*AI is rapidly and radically changing the world around us. AI helps us to understand complex and vast amounts of data in many different areas (...). AI is often thought to be a great promise to make this world a better place. However, as the saying goes, the road to hell is paved with good intentions. Automating society, particularly when introducing highly sophisticated autonomous technologies, can result in disadvantages, undesirable side- effects, and unforeseen new applications. This may call for regulation, for instance, to offer sufficient protection to citizens and to reflect specific norms and values in the design of such new technologies*».

[3] S. Greenstein, *Preserving the rule of law in the era of artificial intelligence*, in *Artificial Intelligence and Law*, 20, 2022, 291 ss., spec. 307, sottolinea infatti che *«[t]echnology is often described as a 'double-edged sword' as its effects on society can be both beneficial but also risky. For example, technology may curtail freedom of expression but at the same time facilitate it. The inherent nature of AI is without doubt a threat to the rule of law and these must therefore be addressed. It is therefore necessary first to highlight some of the risks to the rule of law»*. Anche la Commissione europea, nella Relazione sull'applicazione della Carta dei diritti fondamentali dell'Unione europea, 10/12/2021, COM[2021] 819 final, p. 18, ha sottolineato che «[l]'uso delle tecnologie di intelligenza artificiale (IA) può avere importanti effetti positivi sulle nostre società. Può aumentare l'efficienza dei processi o stimolare l'innovazione e la ricerca. Può inoltre servire a promuovere una serie di diritti fondamentali, quali i diritti alla libertà di espressione e di informazione o all'assistenza sanitaria, e promuovere importanti questioni di interesse pubblico quali la sicurezza pubblica o la sanità pubblica. D'altro canto, quando l'IA è utilizzata senza garanzie e controlli di qualità adeguati per automatizzare

Tali considerazioni non sono sfuggite all'attenzione delle istituzioni che si sono proposte come primi regolatori del fenomeno, in primo luogo in ambito eurounitario.
Nel 2020, ad esempio, il Consiglio d'Europa aveva manifestato preoccupazione con la Risoluzione n. 2344, che significativamente accenna nel titolo a «*new rights or new threats to fundamental freedoms?*», con cui l'Assemblea Parlamentare aveva analizzato sommariamente i progressi registrati nell'ambito delle nuove tecnologie, fornendo una serie di principi etici e direttive volti a regolamentarli nel rispetto dell'individuo e della sua dignità[4].
Già nel 2018, peraltro, la Commissione europea aveva constatato la crescente importanza del contributo dell'Unione nello sviluppo e nella regolamentazione dell'AI, che avrebbe potuto porre l'UE in una posizione di parità nel dialogo (specie di natura economica) con gli altri competitor mondiali nel settore[5]. Nel farlo, la Commissione aveva tuttavia evidenziato l'esigenza che l'implementazione e la diffusione dei sistemi di AI dovesse necessariamente coniugarsi con il sistema valoriale dell'UE, garantendo ai cittadini europei i rispettivi diritti, in forma tradizionale o tecnologicamente connotata[6].
Si tratta di un tratto distintivo della strategia regolatoria eurounitaria in materia di AI, che mira a perseguire lo sviluppo economico e commerciale delle nuove tecnologie per un'affermazione sul mercato globale dell'Unione, senza per questo determinare sacrifici inaccettabili in termini di tutela dei cittadini e delle loro prerogative fondamentali[7].
Tale approccio costituisce un *unicum* nel panorama mondiale: Cina e Stati Uniti d'America, infatti, che si sono attivati ben prima delle istituzioni europee, hanno adottato politiche diverse da quella implementata dall'UE, in termini sia strutturali sia concettuali.
Da un lato, le prime manifestazioni d'interesse da parte della Cina nei confronti dell'Intelligenza Artificiale si sono espresse nel 2012, quando, durante il XVIII Congresso del Partito Comunista Cinese, si era rilevata l'importanza della diffusione delle tecnologie basate sull'AI nel Paese, per agevolarne lo sviluppo commerciale e sociale[8].

---

o sostenere i processi decisionali o per attività quali la sorveglianza, può anche violare i diritti delle persone. Tali violazioni possono verificarsi su larga scala, a seconda della diffusione dell'utilizzo di un sistema, e possono essere difficili da prevenire o rilevare quando il sistema di IA non è sufficientemente trasparente o le persone non sono a conoscenza del suo utilizzo».

[4] La Risoluzione è disponibile al sito *europeanrights.eu.*

[5] Cfr. Comunicazione «L'intelligenza artificiale per l'Europa», 25/04/2018, COM[2018] 237 final.

[6] A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. Pajno-F. Donati-A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Vol. 1: Diritti fondamentali, dati personali e regolazione*, Bologna, 2022, 127 ss., spec. 128, ha spiegato efficacemente che «[l]a ricerca di soluzioni che permettano di garantire che la progettazione e l'utilizzo delle applicazioni tecnologiche avvengano in conformità ai diritti fondamentali implica valutazioni complesse che richiedono, anzitutto, l'individuazione dei diritti e principi che, sia attualmente sia in modo potenziale, possono venire in rilievo. Se infatti la rilevanza di alcuni diritti fondamentali è del tutto evidente – come, in particolare, il rispetto della vita privata o il diritto alla salute – le implicazioni delle tecnologie basate sull'intelligenza artificiale mettono in gioco altri diritti e principi la cui individuazione può avvenire solo con riguardo alle concrete applicazioni, verificando in modo in modo empirico le conseguenze che queste comportano sia nella sfera privata che in quella pubblica».

[7] Il Libro Bianco «Un approccio europeo all'eccellenza e alla fiducia», 19/02/2020, COM[2020] 65 final, fornisce un quadro sintetico degli obiettivi che si è posta l'Unione europea nella gestione di questi temi, senza trascurare le implicazioni economiche né i potenziali effetti dell'AI sui diritti fondamentali.

[8] Ne fornisce una sintesi il rapporto elaborato dalla Commissione per l'esame dell'economia e della

In seguito, il governo cinese ha sviluppato il *Next Generation Artificial Intelligence Development Plan* (AIDP), che disponeva le prime misure concrete volte all'ottenimento della leadership globale cinese in tema di AI, coniugando azioni locali e centrali[9].
Nel 2021 è stato pubblicato altresì il Libro Bianco sulla standardizzazione dell'AI, che descriveva il contesto in cui si sarebbero dovuti innestare i lavori per la costruzione e lo sviluppo delle infrastrutture dedicate all'AI.
Gli interventi adottati dal governo cinese per lo sviluppo delle tecnologie algoritmiche, in generale e in estrema sintesi, sono volti al raggiungimento dell'efficienza tecnica e all'ottenimento del primato economico nel settore, con ricadute sui cittadini che si declinano anche in forme di controllo morale da parte del governo[10]. L'obiettivo, nonostante le (vaghe) affermazioni etiche di principio da parte delle istituzioni cinesi, pare essere quello di sfruttare le dinamiche commerciali legate all'AI per rafforzare il sistema politico centrale, assicurandogli stabilità e controllo[11].
Dall'altro lato, gli Stati Uniti si sono posti da subito in concorrenza con l'investitore cinese, implementando una disciplina regolatoria di tipo verticale, emanata perlopiù dall'esecutivo ma sprovvista di forza vincolante, adottando un approccio c.d. *hands-off*[12].
Il percorso statunitense, intrapreso dall'amministrazione Obama nel 2016 e proseguito sotto la presidenza Trump, era dedicato perlopiù allo sfruttamento economico e all'innovazione, con l'assegnazione di un ruolo primario agli investitori privati autoregolantisi[13].
Con l'amministrazione Biden il focus si era leggermente spostato, come testimoniato dall'accordo tra gli Stati Uniti e l'Unione europea per la formazione del Consiglio per lo Scambio e la Tecnologia, che dovrebbe coadiuvare la cooperazione intercontinentale in materia di AI[14] e costituire pertanto un punto d'incontro anche con riferimento alle priorità da perseguire in termini globali.
Tuttavia, per quel che riguarda il contesto statunitense strettamente inteso, l'attenzione

sicurezza degli Stati Uniti e della Cina, al sito *uscc.gov*.

[9] Cfr. *Next Generation Artificial Intelligence Development Plan Issued by State Council.*

[10] Così A. Monreale, *Rischi etico-legali dell'Intelligenza Artificiale*, in *DPCE online*, 3, 2020, 3391 ss., spec. 3393. In proposito, A. Venanzoni, *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2, 2019, 237 ss., spec. 280, ritiene che la Cina «in Hangzhou ha fatto trionfare lo stritolamento dei dati personali e della riservatezza». Il Parlamento europeo stesso, nella Risoluzione su una politica industriale europea globale in materia di robotica e intelligenza artificiale, 12/02/2019, P8_TA(2019)0081, al punto 13, ha espresso «profonda preoccupazione per l'utilizzo di applicazioni di intelligenza artificiali, ivi compreso il riconoscimento facciale e vocale, in programmi di "sorveglianza emotiva" […], talvolta combinati con sistemi di "credito sociale", come ad esempio già accade in Cina» e ne ha preso le distanze, sottolineando che «tali programmi contraddicono per loro natura i valori e le norme europei che tutelano i diritti e le libertà degli individui», nonché la politica eurounitaria in tema di AI.

[11] Così J. Zeng, *Artificial Intelligence with Chinese Characteristics. National Strategy, Security and Authoritarian Governance*, Singapore, 2022, 68.

[12] L'approccio viene specificato e spiegato nel memorandum del novembre 2020 *Guidance for Regulation of Artificial Intelligence Applications*, disponibile all'indirizzo *whitehouse.org*.

[13] Sulle criticità – specie legate all'enforcement – dei processi di autoregolazione da parte degli investitori privati, cfr. E. Chiti-B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della Regolazione dei mercati*, 1, 2020, 29 ss., spec. 43.

[14] Cfr. European Commission, *EU-US Trade and Technology Council*.

rivolta ai diritti fondamentali si è posta (e, per certi versi, continua a porsi) in subordine rispetto a quella nei confronti degli sviluppi economici e commerciali dell'AI, mancando delle disposizioni programmatiche e sistematiche per la protezione dei diritti umani[15].

Le esperienze cinese e statunitense, rivestendo il primato commerciale in tema di AI, costituiscono il contraltare più immediato per compiere una disamina comparata della strategia eurounitaria, che si propone anch'essa, seppur in maniera differente, di ottenere una leadership nel settore. Tuttavia, gli esempi comparatistici potrebbero essere i più vari e, tra i molti, numerosi sarebbero quelli che si accostano, per modalità e obiettivi, ai modelli extraeuropei finora delineati. È il caso, in particolare, del programma nazionale di Israele sull'AI, creato per assicurare al paese il primato scientifico che possiede in altri contesti tecnologicamente connotati[16]. L'obiettivo è perseguito, ancora una volta, con l'adozione di una struttura marcatamente *hands-off*, che non si manifesta con un disegno organico a livello nazionale, nella convinzione che questo potrebbe inibire la flessibilità delle procedure e quindi essere d'intralcio all'innovazione[17].

In definitiva, quindi, il contesto regolatorio più completo e più *human-rights oriented* pare essere oggi quello eurounitario. Invero, a partire dai più embrionali atti di *soft law* in materia di AI, proseguendo per il GDPR e il *Digital Services Act Package* e culminando nel recente AIA, l'attenzione rivolta da parte delle istituzioni europee ai diritti fondamentali pare essere massima, al punto che nel Regolamento dedicato all'AI l'approccio è basato sul rischio, con particolare attenzione al momento della programmazione degli algoritmi[18].

Se questo è il quadro sommario con il quale è necessario confrontarsi, ne emerge una chiara esigenza di ragionare in termini puntuali sulle implicazioni dell'AI rispetto a ogni posizione soggettiva garantita agli individui, in modo da intercettare le peculiarità e le criticità proprie di ciascuno rispetto alla diffusione dei sistemi algoritmici.

In questo discorso, la libertà di pensiero, coscienza e religione merita, al pari degli altri diritti, una considerazione autonoma, in grado di fornire alcuni spunti rispetto al suo atteggiarsi con l'emersione delle nuove tecnologie, a partire dalla c.d. *digital religion*.

---

[15] Ad esempio, la *Blueprint for an AI Bill of Rights* si è occupata di proporre una serie di principi etici, tuttavia non provvisti di forza vincolante.

[16] In passato, ad esempio, era già stato evidenziato come Israele, al pari della Cina, abbia sviluppato tra i primi al mondo capacità di *cyber warfare* formidabili: così A. Singh Gill, *Artificial Intelligence and International Security: The Long View*, in *Ethics & International Affairs*, 2, 2019, 169 ss., 172.

[17] Così G. Paltieli, *Visions of Innovation and Politics: Isreael's AI initiatives*, in *Discover Artificial Intelligence*, 2, 2022, par. 4: «*Officially, Israel still does not have a coherent national AI strategy, but only a "national program". This might seem like a semantic difference, but as I tried to show it bears a meaningful significance that needs to be explained. […] Adhering to the well-known tech maxim "it's not a bug, it's a feature", […] a lack of strategy allows agility, and agility can enable innovation*». Al programma nazionale israeliano è dedicato il sito internet *aiisrael.org.il*. Nel 2020 è stato altresì stilato, per il Comitato ad Hoc per l'AI costituito in seno al Consiglio d'Europa, il report *Harnessing Innovation: Israeli Perspectives on AI Ethics and Governance*, che forniva un quadro della normativa israeliana in tema di regolamentazione dell'AI.

[18] Il testo dell'AI Act, entrato ufficialmente in vigore il 1° agosto 2024, è reperibile sul sito della Gazzetta Ufficiale dell'UE; per la versione italiana, cfr. *https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689*.

2. Sviluppo tecnologico e *digital religion* (cenni)

La c.d. *digital religion* rappresenta l'oggetto di studio di una branca scientifica autonoma che, secondo parte della dottrina, si occupa di indagare le modalità di sviluppo delle pratiche religiose online e come queste si interfacciano con gli elementi del contesto religioso offline[19].

Altra dottrina invece ritiene invece che il campo d'elezione di questa disciplina non sia da ricercarsi nell'interazione tra religione online e offline, bensì in un terzo spazio, rappresentato unicamente dalle manifestazioni religiose digitali, senza alcun riferimento alle estrinsecazioni religiose per così dire analogiche[20].

Inoltre, c'è chi ha studiato il fenomeno trasponendo online l'art. 9 della Convenzione europea dei diritti dell'uomo[21], che tutela la libertà di pensiero, coscienza e religione, assumendo come punti fermi le quattro estrinsecazioni espressamente richiamate da questa disposizione (culto, insegnamento, pratiche e osservanza dei riti) per la disamina della libertà di religione digitale[22].

Si tratta dunque di una scienza ancora *in fieri*, che si nutre dell'esperienza empirica e che può essere osservata da più angoli visuali[23].

Tale considerazione è stata confermata e per certi versi potenziata durante e dopo la pandemia da Covid-19, quando la delicata gestione della crisi sanitaria e la diffusa preoccupazione per la garanzia del godimento dei diritti fondamentali, tra cui la libertà

---

[19] H.A. Campbell, Z. Sheldon, *Community*, in H.A. Campbell-R. Tsuria (a cura di), *Digital Religion. Understanding Religious Practice in Digital Media*, Abingdon-New York, 2013, 57 ss., descrivono la *digital religion* come «framework for articulating the evolution of religious practices online which are linked to online and offline contexts simultaneously». In particolare, i relativi studi si occupano di esaminare «*the technological and cultural space that is evoked when we talk about how online and offline religious spheres have become blended and integrated*». Per una panoramica dell'evoluzione degli studi in materia di *digital religion*, cfr. C. Helland, *Digital Religion*, in D. Yamane (a cura di), *Handbook of Religion and Society. Handbooks of Sociology and Social Research*, Cham, 2016, 177 ss.; L. Berzano, *La religione nell'era digitale*, in *Historia religionum: an international Journal*, 12, 2020, 165 ss.; H.A. Campbell-G. Evolvi, *Contextualizing current digital religion research on emerging technologies*, in *Human Behavior and Emerging Technologies*, 2, 2020, 5 ss.; N. Pannofino, *La digitalizzazione del sacro. Nuovi culti e nuovi media*, in *Quaderni di diritto e politica ecclesiastica*, 1, 2022, 233 ss.

[20] Si tratta della teoria elaborata da N. Echchaibi, S.M. Hoover (a cura di), *The Third Spaces of Digital Religion*, Londra, 2023.

[21] La norma sancisce, al par. 1, che «[o]gni persona ha diritto alla libertà di pensiero, di coscienza e di religione; tale diritto include la libertà di cambiare religione o credo, così come la libertà di manifestare la propria religione o il proprio credo individualmente o collettivamente, in pubblico o in privato, mediante il culto, l'insegnamento, le pratiche e l'osservanza dei riti». Al par. 2, la norma stabilisce che «[l]a libertà di manifestare la propria religione o il proprio credo non può essere oggetto di restrizioni diverse da quelle che sono stabilite dalla legge e che costituiscono misure necessarie, in una società democratica, alla pubblica sicurezza, alla protezione dell'ordine, della salute o della morale pubblica, o alla protezione dei diritti e della libertà altrui». Sul tema, cfr., *ex multis*, M. Toscano, *Il fattore religioso nella convenzione europea dei diritti dell'uomo. Itinerari giurisprudenziali*, Pisa, 2018.

[22] In questo senso si è orientato C. Ashraf, *Exploring the impacts of artificial intelligence on freedom of religion or belief online*, in *The International Journal of Human Rights*, 26-5, 2022, 757-791.

[23] H.A. Campbell-R. Tsuria (a cura di), *Digital Religion. Understanding Religious Practice in Digital Media*, Abingdon-New York, 2021, 2, rilevano come «*[a]lmost a decade after the launch of the first edition of this book, the field of digital religion has changed in various ways. During this decade, this field of study became much more established. […] And, during the last 10 years, the internet itself has continued to shift, forcing the field of study to change with it. Social media have become much more prevalent, and new arenas of digital culture have emerged, such as virtual reality, artificial intelligence, the internet of things, and big data, to name a few*».

religiosa, hanno sostituito le altre tematiche al centro del dibattito ecclesiasticistico[24].
Invero, in un momento in cui molte delle libertà costituzionalmente garantite sono state limitate, anche la libertà religiosa ha dovuto far fronte alle esigenze determinate dalla fase emergenziale, ricorrendo ampiamente a internet e agli altri strumenti tecnologici, dando luogo a una trasposizione digitale di quanto si era fino a quel momento svolto in presenza[25].
L'esempio più macroscopico di questo adattamento è stato rappresentato dall'impiego dello streaming online per svolgere le liturgie, specie in alcuni momenti particolarmente rilevanti per le comunità di fedeli, come la Pasqua e il Natale per i cristiani cattolici e ortodossi, il Ramadan per i musulmani o la Commemorazione della Morte di Gesù Cristo per i Testimoni di Geova[26].
Tuttavia, l'importanza delle tecnologie in generale e di internet in particolare non rappresenta il portato della sola pandemia: la partecipazione delle religioni sul web era consistente ben prima della diffusione del virus[27].
Ad esempio, già nel 1997 papa Giovanni Paolo II aveva scelto di introdurre la Chiesa cattolica al mondo online, attribuendole di fatto la qualifica di internauta istituziona-

[24] *Ivi*, 6: «[e]specially during the COVID-19 pandemic, when millions of people could not participate in religious gatherings offline, the use of digital technologies to engage religious populations was evident». M.L. Movesian, *Law, Religion, and the COVID-19 Crisis*, in *Journal of Law and Religion*, 37-1, 2022, 9 ss., osserva come *«[i]n the United States at the start of 2020, lawyers and scholars were preoccupied with other issues, such as whether local governments could exclude religious schools from public scholarship programs, and whether religious believers could claim exemptions from public accomodations law that prohibit discrimination based on sexual orientation and gender identity […]. Those debates have not ended. But a central issue on the law-and-religion agenda, one that has drawn academic, judicial, and popular attention, has turned out to be something completely different: whether, and to what extent, government can legally restrict collective worship during a public health emergency*». Sul tema delle configurazioni assunte dalla libertà religiosa e dalla rispettiva tutela durante la pandemia da COVID-19, cfr. quantomeno M. Toscano, *Emergenza sanitaria e libertà di religione*, Torino, 2024; J. Martínez-Torrón, *State, Religion and COVID-19: can Religious Freedom be Guaranteed in Exceptional Circumstances?*, in *Stato, Chiese e pluralismo confessionale*, 16/I, 2022, 7 ss; G. Macrì, *Brevi note in tema di libertà di culto in tempo di pandemia*, in *Il diritto Ecclesiastico*, 1/2, 2020, 49 ss.; R. Santoro, *La libertà di religione nel contesto pandemico*, in *Diritto e religioni*, 2, 2020, 157 ss.; N. Colaianni, *La libertà di culto ai tempi del coronavirus*, in *Stato, Chiese e pluralismo confessionale,* 7, 2020, 25 ss.; A. Licastro, *Annotazioni sugli standard di tutela della libertà di culto nella seconda fase di gestione della pandemia (spunti per una comparazione tra Italia, Francia e Stati Uniti d'America)*, in *Consulta online*, 3, 2020, 758 ss.

[25] Con P. Perri, *La tutela dei dati personali nei social networks e nelle app religiose*, in *JusOnline*, 3, 2020, 82 ss., partic. 82, «[n]essuna attività è stata risparmiata: lavoro, istruzione, associazionismo, volontariato, eventi culturali e qualsiasi altra iniziativa che, nella norma, viene svolta anche con persone estranee al proprio nucleo familiare, è stata trasferita in Internet, grazie ai diversi strumenti d'interazione offerti da piattaforme già esistenti quali i *social media* […] e le *app* per i telefoni portatili». In questo senso, P. Consorti, *La libertà religiosa travolta dall'emergenza*, in *Forum di Quaderni Costituzionali*, 2, 2020, 369 ss., spec. 376, ha rilevato che il regime riservato alla libertà religiosa in fase emergenziale non è stato discriminatorio, dal momento che «[l]a ragione della sospensione delle manifestazioni pubbliche è direttamente funzionale all'eliminazione di possibili evidenti cause di contagio. Allo stesso modo, le disposizioni successive [all'istituzione della c.d. "zona rossa"] non hanno mai discriminato la materia religiosa riservandole un trattamento diverso da quello attribuito alle altre libertà».

[26] P. Palumbo, *Digital religious celebrations during and after the Covid-19. Limits and opportunities for regulation*, in *Stato, Chiese e Pluralismo Confessionale*, 17, 2021, 77 ss., spec. 82.

[27] C. Helland, *Online religion as lived religion. Methological issues in the study of religious participation on the internet*, in *Online – Heidelberg Journal of Religions on the Internet*, 1.1, 2005, 1 ss., partic. 1, osservava a suo tempo che *«[o]ne of the greatest difficulties in studying religion on the Internet is keeping pace with its rapid development and changes. This has been a significant issue when developing theoretical frameworks for examining religious participation on the World Wide Web. Religion has always had a significant online presence*».

lizzato, con la creazione del sito ufficiale della Santa Sede. Nello stesso anno, la Federazione delle Organizzazioni Islamiche in Europa aveva lanciato il sito istituzionale dell'European Council for Fatwa And Research.

Sono queste solo alcune delle molteplici declinazioni materiali della *webreligion* che, pur avendo subito una fortissima accelerazione nell'ultimo decennio, si è sviluppata subito dopo la nascita e la diffusione della rete, implicando da subito alcune criticità legate alla navigazione «in un oceano di siti dedicati al doppio lemma *religione/spiritualità*»[28].

Tale quadro è complicato dalla diffusione delle tecnologie algoritmiche, che hanno una massiva presenza non solo su internet come tradizionalmente concepito, ma che si trovano altresì a produrre importanti ripercussioni sulla vita analogica degli individui, che sempre più s'intreccia ed è condizionata dalle dinamiche che animano il funzionamento dell'AI.

Le medesime preoccupazioni sono state (e sono tuttora) oggetto di riflessione anche da parte delle istituzioni religiose, che si sono dotate di strumenti atti a contenere e a inquadrare i nuovi fenomeni algoritmici, con l'obiettivo di partecipare attivamente al dibattito e rendersi attori in prima linea per lo sviluppo etico dell'AI[29]. Il riferimento è, in particolare, alla *Rome Call for AI Ethics*, firmata nel febbraio 2020 a esito di «un convegno dedicato alla ricerca di un algoritmo buono»[30]. Il documento, che ha l'ambizione di impegnare tutti i soggetti coinvolti a «*guarantee an outlook in which AI is developed with a focus not on technology, but rather for the good of humanity and the environment, of our common and shared home and of its inhabitants*»[31], non si pone come obiettivo la risoluzione analitica delle criticità legata all'impiego delle tecnologie di AI, mirando invece a costituire un invito all'*accountability* rivolto a tutti i soggetti coinvolti[32].

Tali standard etici, che rinnovano quanto già proposto dalle istituzioni europee nei documenti di *soft law* sul tema, hanno prodotto una risonanza a livello internazionale: secondo l'*AI Index Report*, ossia la pubblicazione annuale dell'Institute for Human-Centered AI (HAI) in seno alla Stanford University, l'impegno del Vaticano è da

[28] E. Pace-G. Giordan, *La religione come comunicazione nell'era digitale*, in *Humanitas*, 5-6, 2010, 761 ss., spec. 762.

[29] Con A.S.J. Spadaro-P. Twomey, *Intelligenza artificiale e Giustizia Sociale. Una sfida per la Chiesa*, in *La Civiltà Cattolica*, 1, 2020, 121 ss., spec. 131, «[l]'evoluzione dell'IA contribuirà in grande misura a plasmare il XXI secolo. La Chiesa è chiamata ad ascoltare, a riflettere e a impegnarsi proponendo una cornice etica e spirituale alla comunità dell'IA, e in questo modo a servire la comunità universale. Seguendo la tradizione della *Rerum novarum*, si può dire che qui c'è una chiamata alla giustizia sociale. C'è l'esigenza di un discernimento. La voce della Chiesa è necessaria nei dibattiti politici in corso, destinati a definire e ad attuare i principi etici per l'IA».

[30] M. Ventura, *Il concordato sull'algor-etica del Papa con Microsoft e IBM*, 8 marzo 2020, Corriere della Sera.

[31] Rome Call for AI Ethics, 3. Nello stesso senso, P. Benanti, *The urgency of an algorethics*, in *Discover Artificial Intelligence,* 11, 2023, 1 ss., spec. 7, denuncia che *«[o]ur human condition convinces us that technology is a gift. But our sapiential knowledge […] tells us that our existence is always marked by possibility: for good or evil. To choose good and avoid evil, we need ethics. To do this today, with the help of artificial intelligence, we urgently need algorethics*».

[32] P. Annichino, *Tra algor-etica e regolazione. Brevi note sul contributo dei gruppi religiosi al dibattito sull'intelligenza artificiale nel contesto europeo*, in *Quaderni di diritto e politica ecclesiastica*, 2, 2020, 341 ss., spec. 347, specifica che la Call «non si prefigge di affrontare nel dettaglio tutte le problematiche etiche relative all'adozione delle tecnologie dell'intelligenza artificiale, ma, come ha sottolineato Monsigor Vincenzo Palla, costituisce un "appello a riconoscere e poi ad assumere la responsabilità che proviene dal moltiplicarsi delle opzioni rese possibili dalle nuove tecnologie digitali"».

annoverare tra i cinque *topic* che hanno catalizzato la maggiore attenzione nel 2020 con riguardo all'uso etico dell'IA[33].
L'impegno assunto dalle istituzioni religiose, come quello assunto dagli altri attori istituzionali, risponde pertanto all'esigenza di tutelare la posizione del fedele (e non) che si ritrovi a navigare o utilizzare gli strumenti tecnologici a sua disposizione per fare acquisti, ottenere informazioni, viaggiare o comunicare con le altre persone: in poche parole, che utilizzi (scientemente o meno) le tecnologie algoritmiche.

3. Profilazione e pubblicità comportamentale

Le attività degli internauti sui social network, su altre piattaforme analoghe e nel corso della navigazione *tout court* determinano la rivelazione – spesso inconsapevole – di dati personali relativi alle proprie abitudini, alle proprie idee e alle proprie preferenze.
Tale condivisione non risparmia alcun fattore personale, tantomeno quello religioso o a-religioso.
Sulla base dello sfruttamento di questo genere di informazioni si è sviluppata la pubblicità comportamentale online o *online behavioural advertising* (OBA), vale a dire «una forma di propaganda o pubblicità – cioè di comunicazione d'idee, credenze, fedi, ideologie, culture, e, più in generale, informazioni finalizzate alla persuasione dei destinatari, non solo a scopo commerciale – che si fonda sull'analisi dei comportamenti di ogni singolo utente del *web* cui è rivolta, al fine di adattarsi dinamicamente ai suoi gusti, idee politiche o religiose»[34].
Lo scopo di questo tipo di strategie pubblicitarie consiste nell'identificazione, tramite l'impiego dei *cookie*, ossia i file creati autonomamente dalla rete quando l'utente navighi sui browser di internet, le caratteristiche tipiche dello *user*, in modo da utilizzarle per selezionare i contenuti in grado di soddisfare più pienamente il suo gusto, le sue preferenze o le sue esigenze particolari[35].

[33] Il report, reperibile al link *https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf*, 12, annovera «the Vatican's AI ethics plan» insieme al White Paper rilasciato dalla Commissione europea, il licenziamento da parte di Google di Timnit Gebru (co-leader del team per l'IA etica), il comitato per l'IA etica formato dalle Nazioni Unite e l'uscita di IBM dal business del riconoscimento facciale. L'iniziativa ha avuto seguito con l'AI Ethics: An Abrahamic commitment to the Rome Call, allestito per includere nella Call gli esponenti delle altre religioni abramitiche, in modo da coinvolgerle nella missione etica e permettere ai firmatari originari di rinnovare la propria dedizione.

[34] D. Morelli, Perché non possiamo non dirci tracciati: *analisi ecclesiasticistica della pubblicità comportamentale on-line*, in *Stato, Chiese e Pluralismo Confessionale*, 37, 2012, 1 ss., spec. 1. Sul tema della pubblicità comportamentale cfr., quantomeno, E.C. Pallone, *La profilazione degli individui connessi a Internet: privacy online e valore economico dei dati personali*, in *Ciberspazio e diritto: rivista internazionale di informatica giuridica*, 16-2, 2015, 295 ss.; G. D'Ippolito, *Profilazione e pubblicità targettizzata on line. Real-Time Bidding and behavioural advertising*, Napoli, 2021; A. Adinolfi-A. Simoncini (a cura di), *Protezione dei dati personali e nuove tecnologie: ricerca interdisciplinare sulle tecniche di profilazione e sulle loro conseguenze giuridiche*, Napoli, 2022; A. Di Cerbo, *L'inquadramento giuridico dei dati personali ceduti per la fruizione dei servizi digitali*, in *European Journal of Privacy Law & Technologies*, 2, 2022, 293 ss.

[35] L'enciclopedia Treccani online fornisce la seguente definizione di *cookie*: «In informatica, il file di servizio che viene inviato da un sito Internet all'utente che si collega con esso, allo scopo di registrarne l'accesso e di lasciare sullo schermo un'icona che renda immediato il collegamento in una successiva circostanza. Talvolta con il termine c. si indica anche l'icona stessa». Sul tema cfr., quantomeno, G.

Si tratta di quella che viene definita profilazione dal legislatore eurounitario che, all'art. 4 n. 4 del General Data Protection Regulation (GDPR), la descrive quale «qualsiasi forma di trattamento automatizzato dei dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica»[36].

Grazie a questa tecnica, è possibile ottenere un quadro più o meno completo delle caratteristiche personali e dei dati dell'utente di internet, essenziali per l'alimentazione e il funzionamento dei sistemi di AI, che a sua volta li tratta e li impiega, talvolta automaticamente, per permettere ai meccanismi algoritmici di dispiegare i propri effetti[37].

Da un lato, lo strumento della profilazione offre grandissime potenzialità per l'attuazione dei diritti *ex* art. 19 Cost., poiché permette alla propaganda religiosa di assumere una rilevanza e di avere una risonanza senza precedenti, altrimenti difficili da replicare[38].

Tuttavia, queste attività conducono alla produzione di elaborati sistemi di archiviazione, trattenimento e impiego dei dati sensibili (quali sono quelli riguardanti la fede) che gli utenti, per scarsa dimestichezza o per informative talvolta non trasparenti, non comprendono o non percepiscono di alimentare[39].

Marino, *Internet e tutela dei dati personali: il consenso ai* cookie, in *Jus Civile*, 2, 2020, 398 ss.; L.N. Jayakumari, *Cookies 'n' Consent: An empirical study on the factors influencing of website users' attitude towards cookie consent in the EU*, in *DBS Business Review*, 4, 2021, 1 ss.; M.R. Perugini, *Cookies e consenso: le nuove prospettive*, in *European Journal of Privacy Law & Technologies*, 1, 2021, 1 ss.; F. Zorzi Giustiniani, *Il panopticon digitale: i cookies tra diritto e pratica nell'Unione europea*, in *Freedom, Security & Justice: European Legal Studies*, 2, 2022, 241 ss.

[36] Il Regolamento è disponibile al sito *eur-lex.europa.eu*.

[37] A. Ceserani, *Profilazione religiosa e sicurezza: alcune riflessioni su un quadro normativo in divenire*, in *Il diritto Ecclesiastico*, 4, 2023, 867 ss., spec. 867-868, ha rilevato che «[n]el mondo digitale col termine 'profilare' s'intende, come noto, l'operazione, generalmente automatizzata, di raccolta, trattamento ed elaborazione di informazioni riguardanti soggetti, perlopiù persone fisiche, al fine da renderne possibile la classificazione a più fini. Il risultato è la costruzione – più o meno completa, ma anche più o meno attendibile – di una identità personale sulla cui base chi ne dispone può anche assumere una decisione, con modalità che possono essere altrettanto automatizzate».

[38] *Ibid.*: «il tracciamento comportamentale religioso *online* dell'utente può essere utilizzato da soggetti privati per fini di natura commerciale, poiché il tracciamento favorisce la fornitura di beni e servizi religiosi personalizzati o l'individuazione di nuovi potenziali clienti; a volte, la profilazione religiosa serve solamente per valutare l'affidabilità del consumatore di beni religiosamente neutri. Tale profilazione è inoltre usata da Chiese e organizzazioni confessionali per fornire servizi di natura religiosa o per svolgere attività di proselitismo e di propaganda». G. Pavesi, *I* social media *come strumento di propaganda religiosa*, in *Il diritto Ecclesiastico*, 4, 2023, 891 ss., partic. 892, ha osservato come «l'immediata accessibilità della rete e, più nello specifico, delle piattaforme *social* consenta di raggiungere, in tempo reale, un pubblico ampio e dai confini indeterminati, in nessun modo paragonabile al novero dei destinatari delle tradizionali attività di propaganda, quali il volantinaggio o la predicazione 'porta a porta' nonché, più di recente, la trasmissione tramite canali radiotelevisivi. A ciò si aggiunga che, grazie alle sofisticate tecniche di profilazione dei *social media*, la scelta del 'bersaglio' diviene particolarmente accurata, con la possibilità di cucire sartorialmente su di esso il messaggio, in modo da renderlo altamente persuasivo ed efficace». Sul punto cfr. anche V. Pacillo, *Ciberspazio e fenomeno religioso: profili giuridici*, in *Ciberspazio e diritto*, 1, 2022, 17 ss.

[39] Secondo G. Pavesi, *I* social media, cit., 893, «i moderni mezzi di comunicazione e condivisione hanno portato con sé alcune insidie, suscettibili di mettere a dura prova la tenuta di diritti e libertà fondamentali».

Nel caso specifico dei social network, ci si trova spesso a fornire, nell'ambito del proprio profilo, informazioni «relative alla propria appartenenza confessionale, ma spesso gli utenti non si rendono conto di quale possa essere l'ambito di diffusione di questa informazione e di quanto possa essere facile da estrarre, utilizzando sistemi di *data mining* volti a profilare i soggetti sulla base di specifici indicatori»[40].

Non avvedendosi di tali meccanismi, il fedele (e non) online si ritrova bombardato dalla pubblicità comportamentale sensibile, che si serve dei dati afferenti al fattore religioso raccolti nella navigazione per proporgli messaggi pubblicitari elaborati *ad hoc* per fare leva sulle sue inclinazioni personali[41]. In tal modo, tra le altre cose, si produce una violazione del diritto a non manifestare le proprie convinzioni (a)religiose[42].

Inoltre, la personalizzazione derivante dalla profilazione può nuocere potenzialmente al pluralismo (anche) confessionale, dal momento che, in base ai sistemi legati ai meccanismi di *like, retweet* o simili, gli strumenti di pubblicità comportamentale saranno portati a proporre una sempre maggiore quantità di contenuti della medesima tipologia, creando così delle vere e proprie *religious filter bubbles*[43].

La profilazione finisce così per determinare una fissità di *content display*: gli algoritmi di AI alla base degli strumenti di navigazione in senso lato acquisiscono i dati riguardanti

---

[40] P. Perri, *La tutela dei dati personali*, cit., 90. In proposito, G. Pavesi, *I* social media, cit., 893, ha rilevato che «l'ampio margine di discrezionalità a lungo concesso dagli ordinamenti alle piattaforme digitali ha indotto queste ultime ad atteggiarsi "a presidio del sistema dei valori alla base degli ordinamenti liberal-democratici" arrogandosi "il potere di determinare unilateralmente, in base a regole autoprodotte, quali tipi di comportamenti, informazioni e contenuti possono essere espressi attraverso i servizi da esse offerti e quali invece debbano essere inibiti o censurati"». Per uno studio sulla neutralità della rete e dei social network, cfr. A. Negri, Social network *e fattore religioso: verso nuove forme di neutralità*, in *Il diritto Ecclesiastico*, 4, 2023, 883 ss.

[41] Con G. Mobilio, *La profilazione algoritmica e le nuove insidie alla libertà di religione*, in *Il diritto Ecclesiastico*, 1-2, 2023, 147 ss., spec. 148, la profilazione permette di «personalizzare i servizi offerti, di intercettare le preferenze delle persone , di soddisfare meglio i bisogni sulla base delle esigenze; allo stesso tempo, questo tipo di tecniche espone le persone e i loro dati agli interessi lucrativi delle imprese o a forme di controllo molto penetranti da parte delle autorità pubbliche».

[42] V. Pignedoli, *Privacy e libertà religiosa*, Milano, 2001, 77 ss.

[43] La locuzione *filter bubble* è stata coniata da Eli Pariser, Autore di *The Filter Bubble: What The Internet Is Hiding From You*, Londra, 2012. M. Fasan, *Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 101 ss., spec. 109, specifica che le *filter bubbles* consentono di far sì che «i contenuti visualizzati dall'utente tend[a]no ad essere sempre più in linea con gli interessi e le opinioni dello stesso. Gli algoritmi intelligenti, infatti, al fine di individuare le notizie di maggior interesse per l'utente, filtrano solo ed esclusivamente le informazioni e le opinioni che siano coerenti con la visione della realtà dei singoli individui. La conseguenza principale di questo fenomeno è che le persone, durante la loro navigazione in rete, hanno accesso e ricevono contenuti che rappresentano soltanto un'eco delle opinioni e dei gusti che già in precedenza hanno manifestato, cosicché sono portati a rafforzarsi ancor di più nei loro convincimenti personali». G. Mobilio, *La profilazione algoritmica*, cit., 153-154, ha osservato come «grazie alla profilazione algoritmica, è possibile creare sistemi di c.d. pubblicità comportamentale on-line che, a partire dalle pagine visitate su internet, propongono – o al contrario escludono da – l'acquisto di beni o servizi di ispirazione religiosa; oppure sistemi che, a partire dall'analisi del linguaggio impiegato sui *social media*, segnalano siti di informazione di tendenza sul piano religioso, filosofico o politico, dando origine a quel fenomeno delle 'bolle filtro' per cui l'utente riceve solamente informazioni conformi alle proprie opinioni o pregiudizio».
Sul tema cfr. anche W.H. Dutton-E. Dubois-G. Blank, *Social Shaping of the Politics of the Internet Search and Networking: Moving Beyond Filter Bubbles, Echo Chambers, and Fake News*, in *Quello Center Workin Paper*, 2944191, 2017, 1 ss., spec. 3; G. Pitruzzella, *La libertà di informazione nell'era di Internet*, in questa *Rivista*, 1, 2018, 19 ss.

l'utente (appunto, profilandone le caratteristiche anche religiosamente connotate) e apprendono a sponsorizzare e mostrare contenuti sempre più simili a quelli con i quali ha già interagito in precedenza[44].

Quel che ne consegue è la produzione delle c.d. *echo chamber*, ossia ambienti telematici delimitati dalle attività svolte dall'utente e destinati a riproporsi indefinitamente, appunto, riproducendo il meccanismo dell'eco[45].

Alla luce del quadro sinteticamente descritto, si comprendono le criticità e gli sviluppi legati alla pubblicità comportamentale online e alla profilazione del fedele (e non) online, i cui dati (il cui trattamento è protetto dal GDPR) vengono acquisiti e utilizzati per scopi che spesso esulano dal suo diretto controllo[46].

4. *Bias* e discriminazioni algoritmiche religiosamente connotate

L'impiego delle tecnologie basate sull'AI implica spesso il rischio dell'insorgenza di quelle che sono state definite vere e proprie discriminazioni algoritmiche[47]. Tali discri-

[44] C. Ashraf, *Exploring the impacts of artificial intelligence on freedom of religion*, cit., 771, osserva che *«[w]ith content display, AI systems learn to serve content which generates user interactions by examining data of what users have interacted with previously. The result is that the AI systems typically serve more of the same content, with the emphasis being that content display focuses on what can be seen online»*.

[45] Sul punto cfr., ad esempio, Y. Benkler-R. Faris-H. Roberts-N, Bourassa, *Understanding Media and Information Quality in Age of Artificial Intelligence, Automation, Algorithms and Machin Learning*, in *cyber.harvard.edu/story/2018-07/understanding-media-and-information-quality-age-artificial-intelligence-automation*, 2018; A. Nicita, *Libertà di espressione e pluralism 2.0: I nuovi dilemmi*, in questa *Rivista*, 1, 2019, 314 ss.; M. Cinelli-G. De Francisci Morales-A. Galeazzi-W. Quattrociocchi-M. Stranini, *The echo chamber effect on social media*, in *Proceedings of the National Academy of Sciences*, 118:9, 2021, 1 ss.; L. Terren-R. Borge, *Echo Chambers on Social Media: A Systematic Review of the Literature*, in *Review of Communication Research*, 9, 2021, 9 ss.;

[46] Con D. Durisotto, *La libertà religiosa individuale. Contenuti e problematiche*, in R. Benigni, *Diritto e religione in Italia. Principi e temi*, Roma, 2021, 57 ss., spec. 71, «[i] dati raccolti (big data) raggiungono a livello mondiale una quantità imponente e in crescita esponenziale. Il trattamento dei dati personali, infatti, non coinvolge solo il singolo sito web o piattaforma social che si sta visitando, ma costituisce il frutto di un'azione combinata delle attività dell'utente durante la sua navigazione su uno o più dispositivi, come computer, smartphone o smart tv. Un reale problema può sorgere quando tali piattaforme o altre applicazioni con finalità religiose (ad esempio una raccolta di preghiere), forniscono in modo illegittimo i dati religiosi, concessi consapevolmente dall'utente, a terze società, per finalità che l'utente non conosce». Sul punto cfr. anche A. Fuccillo, *Diritto, religioni, culture. Il fattore religioso nell'esperienza giuridica*, Torino, 2019, 318 ss.

[47] A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 63 ss., spec. 84, denunciava la necessità dell'enunciazione nella normativa eurounitaria di un principio «che potremmo chiamare *di non discriminazione algoritmica*», riferibile in generale all'attività degli algoritmi predittivi e in particolare al caso della profilazione. Nel frattempo l'Unione europea è intervenuta con l'approvazione dell'AIA, che affronta diffusamente il tema delle discriminazioni, con particolare riferimento ai sistemi di AI ad alto rischio. In termini generali, il considerando n. 28 stabilisce che «[l]'IA presenta, accanto a molti utilizzi benefici, la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per pratiche di manipolazione, sfruttamento e controllo sociale. Tali pratiche sono particolarmente dannose e *abusive* e dovrebbero essere vietate poiché sono contrarie ai valori dell'Unione relativi al rispetto della dignità umana, alla libertà, all'uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori». V. Molaschi, *Algoritmi e discriminazione*, in *Fundamental Rights*, 2, 2022, 19 ss., spec. 28-29, propone

minazioni interverrebbero a causa dei c.d. *bias* o condizionamenti di matrice eminentemente umana, trasposti tuttavia nella programmazione informatica. Considerato il funzionamento sostanzialmente automatico delle tecnologie algoritmiche, meccanismi di questo tipo potrebbero condurre a discriminazioni sistematiche su larga scala[48].

Una concreta ripercussione di questi meccanismi può intervenire rispetto alla c.d. *content moderation*, uno strumento elaborato tramite l'AI che gestisce grandi masse di dati e che, tramite la relativa analisi, identifica correlazioni tra gli stessi e ne determina le tendenze e i risultati. A partire da questi, vengono filtrati o rimossi quelli non conformi agli standard predeterminati per il servizio offerto[49].

Una tra le problematiche legate alla *content moderation* consiste nell'eventualità (non remota) che i parametri impostati dai programmatori della piattaforma o del servizio online siano di per sé discriminatori e determinino quindi la rimozione forzata di un contenuto. Nel caso questi ultimi siano religiosamente connotati, l'AI, che dispone automaticamente se possano o meno essere visualizzati, influenza di fatto le modalità in cui la religione o le credenze affini vengono manifestate (o non manifestate) online, prima ancora che il contenuto stesso venga visualizzato dagli individui, rendendone così impossibile la consultazione[50].

In questo senso Ahmed Shaheed, che ha detenuto dal 2016 al 2022 la carica di *Special Rapporteur on Freedom of Religion or Belief* in sede all'ONU, ha evidenziato come la *AI content moderation* possa culminare, in un contesto di intolleranza religiosa, in una «*over-policing of certain faith coomunities and further inhibit communicative action*»[51].

---

l'equazione per la quale dire discriminazione algoritmica è uguale a dire *bias*, specificando che «[t]ali "pregiudizi" caratterizzano sistemi informatici che discriminano sistematicamente e ingiustamente certi individui o gruppi di individui in favore di altri, negando opportunità o beni ovvero attribuendo un risultato indesiderato sulla base di motivazioni irragionevoli o inappropriate». Per uno studio comparato cfr. altresì E. Falletti, *Discriminazione algoritmica. Una prospettiva comparata*, Torino, 2022.

[48] Emblematico, tra tutti, è stato il noto caso statunitense Loomis, nel quale il software informatico COMPAS, che valutava il rischio di recidiva e la pericolosità sociale degli individui in base a dati statistici, precedenti giudiziari e questionari, aveva utilizzato come ulteriore parametro di giudizio il genere, il contesto sociale di provenienza e il colore della pelle. Inoltre, il codice sorgente alla base del software utilizzato non era stato reso noto all'imputato. Cfr. *Criminal Law — Sentencing Guidelines — Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing – State v. Loomis, 881 N.W.2d 749 (Wis. 2016)*, in *Harvard Law Review*, 130-5, 2017, 1530 ss. F. Donati, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020, 415 ss., spec. 423 ss., riporta alcune riflessioni svolte dalla giustizia amministrativa in tema di procedure decisionali automatizzate, in particolare riprendendo Cons. Stato, Sez. VI, 8 aprile 2019, n. 2270.

[49] Sul tema cfr. N. Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, in *Big Data & Society*, 7-2, 2020; M.E. Bucalo, *La libertà di espressione nell'era dei social network fra content moderation e necessità di una regolazione flessibile*, in *Diritto pubblico comparato ed europeo,* 1, 2023, 143 ss.; U. Ruffolo, *Piattaforme e* content moderation *nella dialettica tra libertà di espressione ed autonomia privata*, in *European Journal of Privacy Law & Technology*, 1, 2023, 9 ss.

[50] Con C. Ashraf, *Exploring the impacts of artificial intelligence on freedom of religion*, cit., 773, «*AI can influence how religion or belief manifests online by moderating content related to religion or belief before it is even seen by individuals online, eliminating entire conversations, pages, videos, events, and other content from social media. The potential harm of this approach is significant as it can deprive individuals and groups of the ability to exercise the right to FoRB [freedom of religion or belief] 'alone or in community with others' by completely eliminating the ability to do so*».

[51] A. Shaheed, *Freedom of Religion or Belief: Report of the Special Rapporteur on Freedom of Religion or Belief*, A/HRC/40/58, 5 marzo 2019, 15. In particolare, viene spiegato che «*online tools designed to combat expression that constitutes incitement are not guaranteed to be free from human bias, and their use might reinforce societal prejudices against minorities, exposing them to further stigmatization, discrimination and marginalization […]. Individuals and*

Alcuni studi hanno stimato che questo strumento alimentato dall'Intelligenza Artificiale ha il 150% di probabilità in più di indicare come offensivi i *tweet* scritti da afroamericani[52], nonché di discriminare altri gruppi etnici e geografici minoritari e i migranti. Evidentemente i gruppi religiosi, tantopiù quelli minoritari, sono esposti al rischio di discriminazioni algoritmiche relative al *content display*[53].

Un'ulteriore possibile discriminazione determinata dai sistemi di AI può essere generata dalle nuove tecnologie relative al riconoscimento facciale[54].

Invero, la FRA ha esaminato le implicazioni negative dell'impiego di questi sistemi di identificazione ai danni delle minoranze etniche e religiose, sottolineando l'importanza dell'applicazione dei criteri di proporzionalità e di necessità per limitarne «l'applicazione a casi particolari, come la lotta al terrorismo»[55].

Proprio con l'esimente del contrasto al terrorismo, la Cina ha adottato nella regione dello Xinjiang, popolata dalle minoranze degli uiguri e dei kazaki, sofisticate tecniche di sorveglianza, «rivelando in maniera plastica come l'installazione di telecamere e la diffusione di app di controllo sociale si presti a declinazioni liberticide nei rapporti tra

whole communities may also be targeted through the manipulation of online filters, and the use of some tools, such as facial recognition technology, risks undermining the activities of civil society actors that peacefully pursue the exercise of fundamental human rights».

[52] M. Sap-D. Card-S. Gabriel-Y. Choi-N.A. Smith, *The Risk of Racial Bias in Hate Speech Detection*, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Firenze, 2019, 1668 ss.; T. Davidson-D. Bhattacharya-I. Weber, *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, in *Proceedings of the Third Workshop on Abusive Language Online*, 2019.

[53] FRA, *Bias in Algorithms – Artificial Intelligence and Discrimination*, Lussemburgo, 2022. Il report spiega che «*automated hate speech detection is unreliable. Harmless phrases such as 'I am Jewish' or 'I am Muslim' may get flagged as offensive. And yet offensive content may easily slip through*», 3; e che «*in English, the use of terms alluding to 'Muslim', 'gay' or 'Jew' often lead to predictions of generally non-offensive text phrases as being offensive. In the German-language algorithms developed for this report, the terms 'Muslim', 'foreigner' and 'Roma' most often lead to predictions of text as being offensive despite being non-offensive. In the Italian-language algorithms, the terms 'Muslims', 'Africans', 'Jews', 'foreigners', 'Roma' and 'Nigerians' trigger overly strong predictions in relation to offensiveness. Such bias clearly points to language differences in predictions of 'offensiveness' for different groups by ethnic origin, which means that people who use such phrases are treated differently. Such biased flagging and blocking practices can, for example, lead to differences in access to communication services based on ethnicity. For example, a Jewish person may use the term 'Jew' more often in the online content they post, which may be more readily flagged as offensive and be removed*», 11. Il tema è stato altresì esplorato da R. Xenidis, *When Computers Say No: Towards a Legal Response to Algorithmic Discrimination in Europe*, in B. Brożek-P. Pałka-O. Kanevskaia (a cura di), *Research Handbook on Law and Technology*, Cheltenham, 2024, 222 ss.

[54] FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, Lussemburgo, 2019, 27, riporta che « [*p*]*henotypical characteristics – i.e. the expression of genes in an observable way, such as hair or skin colour – might influence the outcome of biometric matching in facial recognition systems: reflection of light affects the quality of facial images of very fair-skinned persons, and not enough light affects the quality for very dark-skinned persons. When comparing their facial images against a database or watchlist, such people are, therefore, exposed to a higher likelihood of being wrongly matched as false positives. This may result in certain groups of persons being wrongly stopped more frequently due to their colour of the skin*». Rispetto alle discriminazioni legate al colore della pelle, l'MIT ha promosso il progetto *gender shades*, i cui risultati sono visionabili al sito internet *gendershades.org*. A. Pin, *AI, the Public Space, and the Right to Be Ignored*, in J. Temperman-A. Quintavalla (a cura di), *Artificial Intelligence and Human Rights*, Oxford, 2023, esamina invece l'impatto del riconoscimento facciale sul diritto alla riservatezza degli individui e all'anonimato nello spazio pubblico. Rispetto alla possibilità della sorveglianza pubblica degli individui e dell'impiego dei loro dati biometrici a questo scopo è stata promossa *Reclaim Your Face*, una campagna di attivisti per la protezione della privacy: cfr. *reclaimyourface.eu*.

[55] I. Valenzi, *Libertà religiosa e intelligenza artificiale: prime considerazioni*, in *Quaderni di diritto e politica ecclesiastica*, 2, 2020, 353 ss., spec. 362.

autorità e cittadini. Le nuove tecnologie, comprese quelle di riconoscimento facciale, sono utilizzate nell'ambito di una politica di "rieducazione" promossa dal partito comunista e che comprende anche forme di detenzione di massa»[56].

Sulla base di questa e altre esperienze[57], è risultato evidente come gli algoritmi alla base dei sistemi di riconoscimento facciale siano spesso fallibili, impiegando quali parametri, ad esempio, i connotati etnico-razziali o la caratterizzazione sessuale[58].

La discriminazione su base religiosa è coinvolta a pieno titolo nei potenziali rischi che questa tipologia di tecnica determina, al punto che anche il Consiglio d'Europa ha richiesto l'imposizione di un divieto rispetto alle tecniche di riconoscimento facciale «for the sole purpose of determining a person's skin colour, religious or othere belief, sex, racial or ethnic origin, age, health or social status»[59].

Si pensi ad esempio al caso frequente delle donne islamiche che indossano il velo, o ancora alle suore cattoliche, anch'esse provviste di un velo copricapo: in questi casi un sistema di riconoscimento facciale non impostato per riconoscere questo tipo di accessorio (o, peggio, impostato per bloccare i soggetti che portino un velo con un intento deliberatamente discriminatorio) determinerebbe una duplice penalizzazione: l'una legata al genere, l'altra legata al credo[60].

---

[56] M. Colacurci, *Riconoscimento facciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini*, in *Sistema penale*, 9, 2022, 1 ss., spec. 11. I. Valenzi, *Libertà religiosa e intelligenza artificiale*, cit., 362, ritiene che si tratti di «una delle manifestazioni più efferate di utilizzo della tecnologia predittiva in violazione dei diritti fondamentali. […]. Con numerose Risoluzioni a partire dal 4 ottobre 2018 il Parlamento Europeo denuncia e condanna la condizione di internamento in campi di rieducazione di cittadini per motivi di appartenenza religiosa, condizione definita come la più grande detenzione di massa di una minoranza etnica mai operata».

[57] Il riferimento è agli USA, dove la Federal Federal Trade Commission, ad esempio, ha adottato nel 2020 alcune linee guida per gli operatori economici che impiegano l'IA. A fronte della potenziale lesività discriminatoria degli algoritmi, l'organo ha richiesto una preliminare analisi dei dati, dei potenziali bias al loro interno e di una loro gestione etica. In generale, il tema dell'utilizzo dei dati biometrici per il riconoscimento facciale negli USA è molto dibattuto: cfr. A. Chen, *Why San Francisco's ban on face recognition is only the start of a long fight*, in *MIT Technology Review*, 16 maggio 2019; L. Barrett, *Ban Facial Recognition Technologies For Children – And for Everyone Else*, in *Boston University Journal of Science & Technology Law*, 26-2, 2020, 223 ss.; N. Statt, *Massachusetts on the verge of becoming first state to ban police use of facial recognition*, in *The Verge*, 2 dicembre 2020.

[58] Sul tema cfr. I. Raji, J. Boulamwini, *Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. Conference on Artificial Intelligence, Ethics, and Society*, 2019, in *media.mit.edu*; G. Mobilio, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Milano, 2021; F. Paolucci, *Riconoscimento facciale e diritti fondamentali: è la sorveglianza un giusto prezzo da pagare?*, in questa *Rivista*, 1, 2021, 204 ss.

[59] Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data, *Convention 108. Guidelines on Facial Recognition*, T-PD(2020)03rev4, 28 gennaio 2021, spec. 5. Sul documento, tuttavia, F. Paolucci, *Riconoscimento facciale e diritti fondamentali*, cit., 213, denunciava che «il nodo centrale della problematica sono sempre gli stessi due elementi [...]. In primo luogo, il documento in esame è uno strumento che si propone di dare dignità agli individui guidando i governi e i privati, ma non imponendo agli stessi alcunché. Inoltre, non è salutare distinguere tra usi buoni ed usi cattivi di riconoscimento facciale finché le discriminazioni puntualizzate esistono».

[60] Sulla doppia discriminazione religiosa e di genere, cfr. FRA, *Bias in Algorithms*, cit., 12: «*Algorithms can also exhibit bias in relation to the gender categories of certain terms. The gender categories of terms were investigated for the German- and Italian-language data, as these languages use gendered nouns. The analysis shows that available language models (pre-trained AI algorithms based on a large amount of text) can lead to gender bias. This bias can lead to differential predictions, for example by considering the feminine version of a term more offensive than its masculine counterpart, or vice versa. For example, the feminine version of 'Muslim' in Italian ('Musulmana') is rated by the models*

La discriminazione potrebbe essere altresì perpetrata, con esclusivo riferimento alla confessione religiosa d'appartenenza, ai danni di uomini che indossino il turbante sikh o ancora di religiosi sciiti.
In definitiva, per scongiurare i rischi illustrati l'implementazione dei sistemi algoritmici sommariamente esaminata deve necessariamente essere supportata da una programmazione *ethical-by-design*, «valorizzando il contesto sociale nella sua dimensione plurale. Si tratta cioè di promuovere da un lato la partecipazione consapevole dei gruppi culturalmente e religiosamente connotati nell'immissione di dati di buona qualità, rendendoli attenti alle potenzialità della propria presenza sulla rete, e, dall'altro, procedere con la costruzione di algoritmi allenati al riconoscimento e alla valorizzazione delle differenze»[61].

5. Violazioni della riservatezza del fedele (e non)

La riservatezza dell'utente nelle interazioni su internet è uno degli elementi che la normativa e la giurisprudenza eurounitarie tengono debitamente in considerazione nella prospettiva del c.d. mercato unico digitale (MUD)[62].
La diffusione delle tecnologie che sfruttano l'Intelligenza Artificiale (e le tecniche algoritmiche in generale) richiede oggi la necessaria garanzia del diritto alla riservatezza, inteso come quello a non essere ridotti a «oggetto dal quale vengono costantemente estratte, con le tecniche più diverse, tutte le possibili informazioni non solo per le tradizionali, anche se continuamente dilatate, forme di controllo, ma sempre più intensamente per costruire profili e identità, per stabilire nessi e relazioni di cui ci si serve soprattutto per finalità economiche, per ritagliare dalla persona quel che interessa il mercato»[63].
Il diritto alla riservatezza dei dati personali idonei a rivelare le proprie convinzioni afferenti alla religione o alla coscienza può essere senz'altro ricompreso nell'ambito di protezione dell'art. 19 Cost., che tutela anche la libertà di religione negativa, nella forma dell'astensione dall'esprimere le proprie idee in materia religiosa[64].

more negatively than its masculine counterpart ('Musulmano'). This also reflects intersectional hatred, as the rating is based on gender in combination with ethnic origin or religion».

[61] I. Valenzi, *Libertà religiosa e intelligenza artificiale*, cit., 361.

[62] Con F. Ferri, *Il bilanciamento dei diritti fondamentali nel mercato unico digitale*, Torino, 2022, 170, «la giurisprudenza UE sui diritti fondamentali di privacy digitale incide sull'affermazione del MUD, specialmente perché l'obsolescenza che a lungo ha contrassegnato la disciplina UE in materia di dati personali e il progresso tecnologico sempre più diffuso hanno portato la Corte a pronunciarsi su questioni che il diritto derivato governava a fatica». Per una disamina più approfondita della società europea digitale cfr. quantomeno S. Calzolaio-A. Iannuzzi-E. Longo-M. Orofino-F. Pizzetti, *La regolazione europea della società digitale*, Torino, 2024.

[63] G. Ziccardi, *Sorveglianza elettronica, data mining e trattamento indiscriminato delle informazioni dei cittadini tra esigenza di sicurezza e diritti di libertà*, in *Ragion pratica*, 1, 2018, 29 ss., spec. 39.

[64] In proposito D. Morelli, Perché non possiamo non dirci tracciati: *analisi ecclesiasticistica*, cit., 11-12, osserva come «[d]a un punto di vista ecclesiasticistico, sembra corretto ritenere che il diritto alla riservatezza in ambito religioso trovi fondamento non soltanto – come il diritto alla riservatezza tout court, specie quando riferito al c.d. "nucleo duro" della privacy – negli artt. 2 e 3 Cost. (nelle parti in cui essi tutelano rispettivamente i diritti fondamentali dell'uomo e la dignità sociale dei cittadini), ma

Nella medesima prospettiva, la Corte EDU ha avuto modo di censurare pratiche o richieste che possano direttamente o indirettamente condurre un individuo a rivelare forzatamente i propri convincimenti (a)religiosi[65].

Il documento *Privacy and Freedom of Expression in the Age of Artificial Intelligence*, redatto nel 2018 dalle organizzazioni per i diritti umani Article 19 e Privacy International, ha esaminato e illustrato una serie di potenziali rischi prodotti dall'AI rispetto alla riservatezza in materia religiosa. Tra gli altri, sono annoverati la raccolta non consensuale dei dati nei prodotti di consumo, la profilazione degli individui sulla sola base di dati riguardanti il ceto sociale di appartenenza, inferenze generate automaticamente e riguardanti l'identità a partire da dati non ritenuti strettamente sensibili.

A causa dell'opacità che spesso caratterizza gli strumenti di AI è complesso determinare in ogni frangente della navigazione o dell'utilizzo degli strumenti tecnologici l'estensione delle invasioni nella privacy degli utenti[66].

Risulta evidente però la capacità degli algoritmi di dedurre automaticamente i convincimenti religiosi di un individuo a partire dalle sue interazioni online (*like*, *retweet*, tempo di visualizzazione di una determinata pagina web), anche senza una sua azione positiva o un'esplicita affermazione sul tema. A partire da tali inferenze l'AI è in grado di adattare le proprie attività e di orientare quelle dello *user*, che si è ritrovato, senza avvedersene, a fornire in prima persona le coordinate per questo orientamento tecnologico[67].

Nello scenario problematico più estremo, in un clima di intolleranza religiosa, i dati disvelati potrebbero altresì condurre all'identificazione di minoranze religiose, apostati, blasfemi, atei o altre categorie la cui religione o i cui convincimenti personali rappresentino un target per la sorveglianza di massa, l'arresto o la tortura[68].

---

anche nell'art. 19 Cost., il quale, infatti, non può non ritenersi anche il fondamento del diritto di non essere obbligati a rendere manifeste le proprie convinzioni religiose». Il tema, indiscusso da tempo, ha lungamente occupato la dottrina ecclesiasticistica in passato e può sintetizzarsi con le parole di P. Fedele, *La libertà religiosa*, Milano, 1963, 16, che spiegava che «la libertà religiosa non consiste soltanto nella libertà di non credere ad una determinata religione, di non professare una determinata fede, ma consiste altresì nella facoltà spettante all'individuo di credere a quello che più gli piace o di non credere, se più gli piace, a nulla: ciò vuol dire che la libertà religiosa deve essere intesa non soltanto in senso positivo, ma anche in senso negativo». Cfr. altresì, *ex multis*, F. Ruffini, *La libertà religiosa come diritto pubblico subiettivo*, Bologna, 1992; G. Catalano, *Il diritto di libertà religiosa*, Bari, 2007.

[65] Cfr., ad esempio, *Buscarini & others v. San Marino*, ric. 24645/95 (1999); *Alexandridis v. Greece*, ric. 19516/06 (2008); *Grzelak v. Poland*, ric. 7710/02 (2010); *Dimitras and others and others n. 2 v. Greece,* ricc. 42837/06, 3237/07, 3269/07, 35793/07, 6099/08 and 34207/08, 6365/09 (2012).

[66] Per approfondimenti sul tema dell'opacità o effetto *black box* cfr. F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge (MA), 2016; D. Pedreschi et al., *Meaningful Explanations of Black Box AI Decision Systems*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 1, 2019, 9780 ss.; T. Wischmeyer, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer, T. Rademarcher (a cura di), *Regulating Artificial Intelligence*, Cham, 2020, 75 ss.; G. Fioriglio, *La società algoritmica fra opacità e spiegabilità: profili informatico-giuridici*, in *Ars interpretandi*, 1, 2021, 53 ss.

[67] Con Falletti E., *Discriminazione algoritmica*, cit., 58, «la persona che interagisce sulla Rete, sia attraverso Internet o un social network, può appartenere a una certa categoria che raggruppa soggetti aventi in comune determinate caratteristiche focalizzate su certe caratteristiche identitarie, che possono delineare sia ciò che si è sia ciò che si pensa di essere». Nello stesso senso cfr. L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, 167 ss.

[68] C. Ashraf, *Exploring the impacts of artificial intelligence on freedom of religion*, cit., 775, fa riferimento al fenomeno di Clearview AI, che permette alle agenzie governative di individuare le persone tramite

Si capisce quindi che «riguardo alla libertà religiosa, il diritto di riservatezza garantisce contro ogni potenziale discriminazione ad opera delle autorità pubbliche o di determinati soggetti privati nei confronti degli orientamenti religiosi dei singoli individui»[69]. Alla luce di quanto illustrato risulta evidente la rilevanza della garanzia della privacy. A primo acchito, dal momento che questo diritto è caratterizzato da un contenuto negativo, consistendo in un'astensione da parte dei terzi – di qualsiasi natura – dall'invadere la sfera altrui, potrebbe sembrare semplice fornire protezione a questa pretesa soggettiva online dove, per definizione, ciascuno naviga da solo. A ben vedere, tuttavia, a fronte del quadro delineato finora, la trasposizione innanzitutto digitale e in secondo luogo algoritmica di questo diritto, che si pone a presidio, ai sensi del GDPR e degli artt. 7 e 8 della CDFUE, dell'identità personale (ivi compresa quella religiosa) degli individui, pone criticità per certi versi più sottili di quelle osservate sul piano meramente analogico.

Ancora una volta si prospetta la necessità di un complesso bilanciamento: da un lato, il funzionamento automatizzato – con i relativi *pro* e *contra* – dei processi algoritmici alla base dell'AI, il cui sviluppo è interesse di numerosi soggetti e la cui efficienza è al servizio dell'utente stesso; dall'altro lato, «il diritto alla *privacy*, alla riservatezza dei dati e alla loro non divulgazione pubblica»[70].

6. Acquisizione e trattamento dei dati sensibili del fedele (e non)

I sistemi di Intelligenza Artificiale svolgono le loro attività, alimentando i meccanismi di *machine learning* che permettono loro di svilupparsi, sulla base dei dati che riescono a raccogliere e trattenere in seguito al passaggio dell'utente sulla rete: più nutrita è la quantità di dati raccolti, migliori sono le prestazioni dell'AI e le risposte che fornisce[71]. Dato il massiccio utilizzo di tali dati, in ambito eurounitario si è affermato il diritto alla

un'analisi incrociata dei loro social network e dei loro post, anche solo uno solo, in cui appaiano chiese, moschee, sinagoghe o simili.

[69] D. Durisotto, *Diritti degli individui e diritti delle organizzazioni religiose nel Regolamento (UE) 2016/679. I "corpus completi di norme" e le "autorità di controllo indipendenti*, in *federalismi.it*, 27, 2020, 38 ss., spec. 39.

[70] L. Pedullà, *Accesso a internet, libertà religiosa informatica e buon costume*, in *Stato, Chiese e Pluralismo Confessionale*, 35, 2012, 1 ss., spec. 4. L'Autore illustra il «quesito "tragico" tra quale, in caso di conflitto, debba prevalere: il diritto di informare, d'informarsi e di essere informati - quale diritto di "cercare, ricevere, diffondere con qualunque mezzo di espressione" le notizie e le idee - o il diritto alla privacy, alla riservatezza dei dati e alla loro non divulgazione pubblica?», notando altresì come anche su internet «i diritti fondamentali poss[a]no entrare tra loro in conflitto e in tal caso la loro risoluzione non p[ossa] che essere poggiata sull'applicazione del principio di ragionevolezza».

[71] C.M. Reale-M. Tomasi, *Libertà d'espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale*, in *DPCE online*, 1, 2022, 325 ss., spec. 169 spiegano che «alla base della proposta e dell'organizzazione delle informazioni vi sono algoritmi elaborati da sistemi di Machine Learning che "personalizzano" l'informazione sulla base delle impronte che le singole persone lasciano sul web (*big data*)». G.F. Italiano, *Intelligenza Artificiale: passato, presente, futuro*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 207 ss., spec. 217, ribadisce al pari che «[l]e tecnologie alla base di tutto questo [dei sistemi di AI che usiamo quotidianamente] sono (...) algoritmi di *machine learning*, capaci di apprendere velocemente vari compiti analizzando le miriadi di dati che provengono anche dalle nostre continue interazioni con il dispositivo».

protezione degli stessi, che presuppone un'azione positiva da parte degli ordinamenti per il controllo sul trattamento e la salvaguardia dei dati[72].

Nell'ordinamento italiano l'appartenenza confessionale e le convenzioni personali godono del principio di tendenziale irrilevanza giuridica, vale a dire sono insuscettibili di determinare vantaggi o svantaggi per i singoli[73].

Tuttavia, nel momento in cui la dimensione di riferimento cessi di essere quella analogica per trasformarsi in quella algoritmica, tenere fede a tale principio diviene più complesso, considerata l'essenzialità dei dati nell'economia di funzionamento dei sistemi di AI che lavorano per fornire un'offerta personalizzata al massimo livello[74].

Spesso l'utente non ha la percezione dei dati sensibili che cede navigando in rete: il credente (o non) fornisce pertanto a soggetti perlopiù privati, gestori delle piattaforme, specifiche informazioni personali, che possono essere raccolte e messe in circolo per fornire mappature spesso incomplete di una realtà sociale sfaccettata, condizionando la fruizione dei servizi che offrono in base all'appartenenza religiosa dell'individuo.

In ambito nazionale il d.lgs. 196/2003 (Testo Unico per il trattamento dei dati personali) annovera, tra i diritti meritevoli di tutela quanto alla privacy, la libertà religiosa, che coinvolge e produce dati sensibili. L'art. 4 T.U. definisce infatti come «dati sensibili» quelli «idonei a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni od organizzazioni a carattere religioso, filosofico, politico o sindacale, nonché i dati personali idonei a rivelare lo stato di salute e la vita sessuale».

La normativa statale è stata modificata in seguito alla promulgazione del GDPR, il cui «oggetto regolato e la finalità perseguite [...] non possono non interessare da vicino anche le attività di trattamento dei dati aventi natura religiosa, talmente rilevanti da ricevere la qualifica di 'dati sensibili'»[75].

Il considerando n. 4 del Regolamento precisa che esso «rispetta tutti i diritti fondamentali e osserva le libertà e i principi riconosciuti dalla Carta, sanciti dai trattati, in particolare [...] la libertà di pensiero, di coscienza e di religione [...] nonché la diversità culturale, religiosa e linguistica».

L'art. 9 del GDPR vieta inoltre il trattamento di «dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche». Il par. 2,

---

72 Cfr. S. Rodotà, *Il mondo nella rete. Quali diritti, quali vincoli,* Roma-Bari, 2014.

73 C. Cardia, *Principi di diritto ecclesiastico: tradizione europea legislazione italiana*, Torino, 2005, 162. D. Durisotto, *Diritti degli individui e diritti delle organizzazioni religiose nel Regolamento (UE) 2016/679,* cit., fornisce alcune esemplificazioni concrete della portata del principio in parola, sottolineando altresì come in altri ordinamenti europei esso non sia altrettanto radicato e conduca quindi a una certa ambiguità nell'atteggiamento in materia da parte della Corte EDU.

74 In generale, sul tema dei *big data*, cfr. quantomeno G. D'Acquisto-M. Naldi, *Big data e privacy by design*, Torino, 2017.

75 M. Ganarin, *Salvaguardia dei dati sensibili di natura religiosa e autonomia confessionale. Spunti per un'interpretazione secundum Costitutionem del regolamento europeo n. 2016/679*, in *Stato, Chiese e Pluralismo Confessionale*, 11, 2018, 1 ss., spec. 2. Per una disamina della normativa italiana precedente e successiva all'entrata in vigore del Regolamento n. 679/2016 cfr. M. Mazzoni, *Le Autorizzazioni Generali al trattamento dei dati sensibili da parte delle confessioni religiose. Osservazioni alla luce delle recenti riforme in materia di privacy*, in *Stato, Chiese e Pluralismo Confessionale*, 7, 2020, 66 ss.; V. Marano, *Protezione dei dati personali, libertà religiosa e autonomia delle chiese*, in V. Cuffaro-R. D'Orazio-V. Ricciuto (a cura di), *I dati personali nel diritto europeo*, Torino, 2019, 579 ss.

lett. a) della stessa norma, tuttavia, esclude l'operatività del divieto laddove il soggetto abbia prestato il proprio consenso esplicito al trattamento[76].

Con queste modalità, il Regolamento perseguiva l'affermazione del principio della *privacy-by-design,* con l'obiettivo di responsabilizzare i titolari del trattamento, conferendogli alcuni obblighi per una gestione corretta del rischio[77].

Il meccanismo del consenso, tuttavia, non si sta dimostrando efficace quanto si sarebbe auspicato, specie a fronte del funzionamento dell'Intelligenza Artificiale che, con i suoi procedimenti automatizzati, trasforma anche il consenso stesso in una sorta di automatismo che le conoscenze e la consapevolezza dell'utente medio non permettono di intercettare[78].

L'art. 9 par. 2, lett. d) del GDPR sancisce inoltre che il consenso dello *user* non è necessario laddove il trattamento sia eseguito «da una fondazione, associazione o altro organismo senza scopo di lucro che persegua finalità politiche, filosofiche, religiose o sindacali, a condizione che il trattamento riguardi unicamente i membri, gli ex membri o le persone che hanno regolari contatti con la fondazione, l'associazione o l'organismo a motivo delle sue finalità e che i dati personali non siano comunicati all'esterno senza il consenso dell'interessato». La disposizione non ha rappresentato una novità rispetto al precedente apparato legislativo: l'art. 8 della Direttiva 95/46/CE e l'art. 26 co. 3, lett. a) del T.U. Privacy prevedevano già questa eccezione.

Il Regolamento include però anche gli ex membri «fra i soggetti dei quali è consentito il trattamento dei dati»[79]: in questo senso, il GDPR pare riservare un margine di autonomia alle Chiese e alle altre organizzazioni religiose nell'impiego dei dati, posto che queste devono operare nel rispetto delle norme nazionali, ai sensi dell'art. 17 TFUE.

La soluzione è d'altronde costituzionalmente imposta anche dal nostro ordinamento, in cui la «*libertas agendi* delle confessioni esige che esse abbiano il diritto di approntare unilateralmente soluzioni normative, dinanzi alle quali peraltro le autorità statali rinunciano a verificarne in via preventiva il contenuto»[80].

---

[76] Ai sensi dell'art. 4 par. 1, punto 11 del GDPR il consenso consiste in «qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso, mediante dichiarazione o azione positiva inequivocabile, che i dati personali che lo riguardano siano oggetto di trattamento». Ai sensi del considerando 32 «il consenso dovrebbe essere prestato mediante un atto positivo inequivocabile con il quale l'interessato manifesta l'intenzione libera, specifica, informata e inequivocabile di accettare il trattamento dei dati personali che lo riguardano, ad esempio mediante dichiarazione scritta, anche attraverso mezzi elettronici, o orale».

[77] Sul tema cfr. S. Calzolaio, Privacy by design. *Principi, dinamiche, ambizioni del nuovo Reg. Ue 2016/679*, in *federalismi.it*, 24, 2017, 1 ss.

[78] In questo senso G. Mobilio, *L'intelligenza artificiale e le regole giuridiche alla prova: il caso paradigmatico del GDPR*, in *federalismi.it*, 16, 2020, 266 ss.

[79] V. Marano, *Protezione dei dati personali*, cit., 584.

[80] M. Ganarin, *Salvaguardia dei dati sensibili di natura religiosa*, cit., 14. L'Autore specifica altresì che «la connotazione laica delle istituzioni pubbliche preclude la possibilità di valutazioni nel merito di atti confessionali. Per converso, la specificazione del dettato costituzionale nella disciplina sulla tutela dei dati personali responsabilizza le confessioni in ordine alla regolamentazione di una materia che coinvolge, seppure secondo un angolo prospettico divergente, gli interessi 'vitali' dell'ordine confessionale e quelli indisponibili dell'ordine statale». Nello stesso senso cfr. F.D. Busnelli-E. Navarretta, *Battesimo e nuova identità atea: la legge n. 675/1996 si confronta con la libertà religiosa*, in *Quaderni di diritto e politica ecclesiastica*, 3, 2000, 855 ss.

Le criticità legate ai dati non si esauriscono peraltro con riguardo a quelli sensibili e, quindi, quelli strettamente attinenti ai convincimenti religiosi e affini: al contrario, questi vengono combinati dai sistemi algoritmici con quelli non sensibili per ricostruire un profilo individuale dell'utente, ledendo così «sia la sua privacy (a causa dell'invasività totale nella sfera privata di questa), sia la sua dignità (rendendola "nuda" e "trasparente" di fronte a chi entra in possesso dei suoi dati rielaborati»[81].
Di qui all'elaborazione *ex novo*, tramite i meccanismi di *machine learning*, di dati che non sono stati elaborati da operatori umani ma che si autoproducono a partire da quanto già descritto dai dati di partenza, il passo è breve. Sul punto, pare necessario interrogarsi quanto ai modi in cui i dataset stessi, una volta acquisiti, vengano impiegati successivamente dalle strutture dell'Intelligenza Artificiale[82].
La riflessione che si impone, in termini generalissimi, riguarda così le misure attuabili per fronteggiare un fenomeno i cui connotati sono sempre più imponenti, sfuggenti e articolati. Le valutazioni dovrebbero essere parametrate agli specifici mezzi con cui i dati vengono raccolti e impiegati e, in particolare, quanto «all'ampiezza dei dati trattati, alla natura dei dati che si andranno a chiedere all'utente, all'obbligatorietà o meno di conferire quei dati per usufruire del servizio e al periodo di conservazione»[83].
Il risultato auspicato sarebbe la garanzia offerta all'utente di navigare in sicurezza manifestando se – e solo se – lo desidera i suoi convincimenti di coscienza, senza per questo essere coinvolto in un sistema generale di profilazione e di impiego dei suoi dati per i più disparati scopi commerciali (inclusi quelli di progettazione dell'AI stessa).

7. Quadro normativo: il contesto italiano

Fino all'approvazione dell'AI Act, in Italia non esisteva una fonte legislativa né di rango sub-primario idonea a disciplinare, in alcuna sua forma, l'Intelligenza Artificiale[84].

[81] E. Falletti, *Discriminazione algoritmica*, cit., 155.

[82] C. Ashraf, *Exploring the impacts of artificial intelligence on freedom of religion*, cit., 779, si chiede ad esempio *«[w]hat kind of data is being used? How was this data collected? What religious or belief groups were targeted or excluded in data collection? What groups are identified in the training data? Which groups were excluded and for what reason? How do the inclusions and exclusions reflect the real population which will be impacted by the AI system? How might various datasets impact these groups in practice, observance, worship, and teaching online? What errors were encountered during testing? How did these errors impact the various identified groups as well as their ability to practice, observe, worship, and teach online? What inferences can be drawn from these errors? What kind of online content will this AI impact the most? Who or what might it inadvertently impact? How are these distinctions manifest in theory, testing, and implementation?».*

[83] P. Perri, *La tutela dei dati personali*, cit., 93.

[84] G.F. Italiano-S. Civitarese Matteucci-A. Perrucci, *L'intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno-F. Donati-A. Perrucci (a cura di), *Diritti fondamentali, principi democratici e rule of law*, cit., 43 ss., denunciavano «che il nostro Paese non è all'avanguardia nell'applicazione di sistemi di IA nel mondo della produzione e, più in generale, nell'economia e nella società. Come è stato notato da diversi esperti della materia, ad esempio dall'Osservatorio sull'intelligenza artificiale della Bocconi, in Italia ci sarebbe il potenziale per svolgere un ruolo di primo piano nel campo dell'IA: vantiamo posizioni di rilievo nella preparazione dei talenti e nella ricerca, che tuttavia risulta molto frammentata, diverse nicchie di specializzazione, una apprezzabile capacità brevettuale, limitatamente ad alcune applicazioni, una significativa crescita degli investimenti da parte delle imprese. Ciò che serve è una strategia adeguata, un approccio di sistema, che coinvolga pubblico

L'inerzia del nostro legislatore si è a lungo spiegata con l'arretramento dei diritti positivi degli Stati membri auspicato da una parte della dottrina a fronte dell'iniziativa assunta dalle istituzioni europee con la proposta di Regolamento dell'aprile 2021[85].

Nel frattempo peraltro non sono mancati documenti in materia di AI, provenienti da più direzioni, atti a fornirne un inquadramento giuridico coerente con la traiettoria descritta dalle istituzioni UE.

Il Consorzio interuniversitario nazionale per l'informatica (CINI) aveva elaborato già nel 2010 il documento *AI for Future Italy* con l'obiettivo, coerente con la strategia in seguito impostata dall'UE, della creazione di un'AI affidabile e sostenibile, che mirasse al benessere umano a livello individuale e sociale, tramite sistemi che incorporassero i valori etici europei e garantissero il rispetto dei diritti umani e dei valori democratici[86].

Il Libro Bianco sull'IA al servizio del cittadino ha offerto una panoramica complessiva dei risultati raggiunti nel campo dell'AI e di quanto era verosimile attendersi negli anni a venire. Il documento proponeva inoltre alcune sfide poste dall'AI, mantenendo in ogni caso ferma la necessità che «in ogni contesto l'IA [sia] al servizio delle persone»[87].

Nel Programma Nazionale per la Ricerca (PNR) 2021-2027, curato dal MUR, si rinviene una sezione dedicata all'Intelligenza Artificiale, nella quale si ribadisce che l'AI «è una priorità assoluta per tutti i Paesi e per l'Europa *in primis*», definendola altresì «una irripetibile opportunità per il rilancio del nostro Paese dell'industria digitale»[88].

Il Programma Strategico Intelligenza Artificiale 2022-2024 ha fornito inoltre un resoconto sintetico del contesto in cui l'Italia si trovava nel 2021, annoverando tra i suoi cinque principi guida quello per il quale «l'IA italiana è un'IA europea»[89].

Non stupisce quindi che il governo italiano abbia atteso l'approvazione della bozza finale dell'AI Act europeo per sviluppare un disegno di legge in materia di Intelligenza Artificiale, approvato lo scorso aprile[90].

---

e privato».

85 C. Casonato, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, cit., 170, rilevava «la strutturale incapacità del livello statale di affrontare tematiche di per sé transanazionali», quale è l'AI. Nello stesso senso cfr. F. Alicino, *Diritto e religioni alla prova dell'intelligenza duale della globalizzazione*, in *Quaderni di diritto e politica ecclesiastica*, 1, 2021, 143 ss., spec. 146. L'Ufficio Rapporti con l'Unione europea della Camera dei deputati ha pubblicato a novembre 2021 il dossier n. 57 , relativo all'AIA, nel quale la proposta di Regolamento veniva analiticamente esaminata e presentata nei suoi punti. Lo studio ha illustrato che al momento della sua pubblicazione, sulla base dei dati forniti dal sito IPEX, l'esame della proposta era stato concluso da Austria, Croazia, Repubblica Ceca, Francia, Spagna e Italia. In particolare, in Italia era stato finalizzato dal Senato a luglio 2021, accertate la correttezza della base giuridica e il rispetto dei principi di sussidiarietà e proporzionalità.

86 Questi gli obiettivi espressi dal documento a p. 0010, disponibile al sito *magazine.fbk.eu.*

87 Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, disponibile in *ia.italia.it,* 2018, 41. Le sfide proposte dal documento riguardavano in particolare qualità e neutralità dei dati, responsabilità, trasparenza e apertura, tutela della sfera privata.

88 Programma nazionale per la ricerca (PNR) 2021-2027, 2020, 94, disponibile al sito *miur.gov.it*

89 Programma Strategico Intelligenza Artificiale 2022-2024, 2021, 14, disponibile al sito *assets.innovazione.gov.it.* Gli altri quattro principi guida sono: 2) l'Italia sarà un polo globale di ricerca e innovazione dell'IA; 3) L'intelligenza artificiale italiana sarà antropocentrica, affidabile e sostenibile; 4) Le aziende italiane diventeranno leader nella ricerca, nello sviluppo e nell'innovazione di IA; 5) Le pubbliche amministrazioni italiane governeranno l'IA e governeranno con l'IA.

90 Il disegno di legge è disponibile al sito *senato.it.*

Il disegno di legge è costituito da 25 articoli, i primi dei quali (segnatamente dal 3 al 5) sanciscono i principi da adottare in materia, tra i quali emergono la correttezza, l'attendibilità, la sicurezza, la qualità, la trasparenza e l'appropriatezza. Per quel che riguarda il fattore religioso, esso viene citato un'unica volta, nell'ambito dell'art. 10, che dispone quanto all'uso dell'intelligenza artificiale in materia di lavoro. In particolare, il co. 3 sancisce che l'AI impiegata nel rapporto di lavoro «garantisce l'osservanza dei diritti inviolabili del lavoratore senza discriminazioni in funzione del sesso, dell'età, delle origini etniche, del credo religioso, dell'orientamento sessuale […] in conformità con il diritto dell'Unione europea».
Da questo punto di vista, pertanto, il ddl pare rimandare integralmente al Regolamento UE.

8. (segue): il legame con il contesto europeo

Nel solco della politica eurounitaria adottata in tema di AI come sommariamente delineata *supra* (par. 1), l'AI Act si preoccupa precipuamente di garantire che le tecnologie che impiegano gli algoritmi sul territorio UE non sfruttino le vulnerabilità individuali o dei gruppi, anche minoritari[91]. Tra le vulnerabilità personali considerate dall'AIA è infatti annoverata anche l'appartenenza religiosa, specie di minoranza[92].
In secondo luogo, il fattore religioso viene considerato dall'AIA nell'ambito della categorizzazione biometrica: il Regolamento protegge i dati (e ne vieta l'utilizzo) dai quali si potrebbero dedurre informazioni strettamente personali, come, appunto, l'opinione e l'appartenenza religiosa[93].
Infine, l'AI Act fa riferimento al fattore religioso all'art. 5, che vieta l'utilizzo dei meccanismi basati sulla manipolazione degli individui, ancora una volta, tramite i dati biometrici[94].
Il Regolamento UE, in definitiva, pone l'accento sui momenti di implementazione e

[91] Cfr. AI Act, , considerando n. 29: «*AI systems may also otherwise exploit vulnerabilities of a person or a specific group of persons due to their age, disability within the meaning of Directive (EU) 2019/882, or a specific social or economic situation that is likely to make those persons more vulnerable to exploitation such as persons living in extreme poverty, ethnic or religious minorities. Such AI systems can be placed on the market, put into service or used with the objective to or the effect of materially distorting the behaviour of a person and in a manner that causes or is reasonably likely to cause significant harm to that or another person or groups of persons, including harms that may be accumulated over time and should therefore be prohibited*».

[92] *Ibid.*

[93] Ivi, considerando n. 30: «*[b]iometric categorisation systems that are based on individuals' biometric data, such as an individual person's face or fingerprint, to deduce or infer an individuals' political opinions, trade union membership, religious or philosophical beliefs, race, sex life or sexual orientation should be prohibited*». L'utilizzo di tali dati è condizionato al rispetto delle norme nazionali o eurounitarie in materia: segnatamente, il GDPR si occupa del tema all'art. 9 par. 1, che vieta di trattare «i dati biometrici intesi a identificare in modo univoco una persona fisica».

[94] Ivi, art. 5 par. 1 lett. (g): «[*t]he following artificial intelligence practices shall be prohibited: […] the placing on the market, the putting into service for this specific purpose, or the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement*».

sviluppo dei sistemi di AI, coerentemente con l'approccio basato sul rischio che lo ispira. Nell'ambito di tale atteggiamento, peraltro, l'attenzione nei confronti dei diritti fondamentali è per così dire indiscriminata, fornendone una garanzia (di principio) generalizzata. Il fattore religioso, in tale contesto, non pare garantito in maniera specifica e puntuale, posto che esso viene spesso associato alla protezione delle minoranze, senza un'attenzione alla credenza religiosa *tout court*.

Ne consegue dunque che il rimando svolto dalla (potenziale) normativa italiana all'AIA non pare, allo stato, soddisfare l'esigenza di protezione della libertà di pensiero, coscienza e religione, posto che anche in questo testo il fattore religioso è ricompreso in una categoria assai eterogenea di elementi, tutti genericamente riconducibili ai diritti fondamentali.

9. Alcuni spunti per il futuro

Il panorama che pare oggi svilupparsi a fronte della diffusione dell'Intelligenza Artificiale pone l'individuo in una continua interazione con gli altri e con strumenti inediti, rendendolo sempre più sensibile (e vulnerabile) a condizionamenti e suggestioni.

In questo quadro, la libertà di pensiero, coscienza e religione transita dalla sfera eminentemente interiore per ritrovarsi partecipe di un dialogo incessante tra la persona e la dimensione pubblica, analogica e tecnologica[95].

Alla luce di quanto sinteticamente illustrato, la libertà religiosa si sta arricchendo di caratteristiche, implicazioni e strutture fino a un decennio fa impensabili, che contribuiscono a ripensare globalmente le maniere in cui individui e gruppi si ritrovano a vivere, praticare e propagandare il culto.

Da un lato, la rivoluzione tecnologica in atto, di cui è protagonista l'AI, non costituisce il primo dirompente mutamento della *forma mentis* individuale: basti pensare, in questo senso, al cellulare e, soprattutto, agli *smartphone*, che hanno stravolto completamente il modo di vivere delle persone. Dall'altro lato, tuttavia, stiamo assistendo all'emersione e allo sviluppo rapidissimo di un fenomeno altamente pervasivo, le cui modalità di insediamento sono spesso impercettibili ma non per questo meno rilevanti, che, come una sorta di rumore bianco, agisce costantemente e diffusamente sull'agire umano, senza esentare, evidentemente, la sfera religiosamente connotata[96].

La rilevanza acquisita dalle tecnologie algoritmiche ha di fatto determinato l'insorgenza

[95] Con L. Pedullà, *Accesso a internet, libertà religiosa informatica e buon costume*, cit., 6, la libertà religiosa «è (oggi più di ieri) frutto di un continuo confronto dialettico tra l'individuo e la dimensione pubblica, tant'è che anche se non si volesse ammettere l'influenza totalizzante di *internet* sui principi che reggono la libertà religiosa, difficilmente potrebbe negarsi la sua notevole influenza sulla *formazione* della coscienza religiosa».

[96] K.A. Bingaman, *Religious and Spiritual Experience in the Digital Age: Unprecedented Evolutionary Forces. New Directions in Pastoral Theology Conference (Honoring Lewis Rambo)*, in *Pastoral Psychology*, 69, 2020, 291 ss., 293, in proposito afferma che «*we could approach this [the digital revolution] more as a "nothing new under the sun" scenario; we have been here many times before, have lived through many other evolutionary twists and turns. But not like this one, which is now simultaneously and conterminously driven by biological and technological evolution, the organic and the inorganic. In a word, it is very much an unprecedented evolutionary transition, as Shoshana Zuboff makes clear in her recent and important book, Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*».

di una mutata dimensione, nella quale risulta «sempre più sbiadito il confine tra l'*online* e l'*offline*»[97]. Le novità tecnologiche sono qui in grado di creare nuove regole e fornire nuovi connotati a concetti tradizionali, incentivando le persone a proiettare, più o meno consapevolmente, se stesse e le proprie convinzioni, anche religiose, quanto più possibile all'esterno di sé, attribuendo al foro esteriore una centralità che, probabilmente, prima dell'avvento di internet e dei mezzi connessi, sarebbe stata difficilmente concepibile.

In tale situazione, caratterizzata da frammentazione e moltiplicazione di istanze e posizioni, l'esperienza religiosamente connotata pare adeguarsi ai nuovi canoni tecnologicamente imposti e mutare quindi i propri tratti tradizionali.

Le sollecitazioni offerte dalla dimensione *always on* e *data-driven*, interconnessa, fondata sullo scambio e sul dialogo, forniscono quindi l'occasione di riflettere e interrogarsi sulle modalità più efficienti per fornire garanzia alla libertà religiosa dell'individuo.

Se queste sono le premesse, quanto finora registrato dal punto di vista normativo non pare cogliere appieno la profondità, la specificità e la complessità dell'operazione da svolgere.

Da un lato, infatti, il contesto nazionale si preoccupa di armonizzare le proprie regole a quelle dettate dall'Unione europea e, con riferimento ai diritti fondamentali *tout court*, vi rinvia integralmente.

Dall'altro lato, le istituzioni europee hanno svolto, con la stesura dell'AI Act, una delicata operazione di bilanciamento, tentando di coniugare le implicazioni economiche e commerciali determinate dall'AI e le istanze di protezione dei diritti umani. Nel farlo, il Regolamento ha posto l'UE in una posizione di assoluta primazia normativa in termini globali, rappresentando la prima proposta regolatoria organica in materia di AI.

Tuttavia, per quel che riguarda le posizioni giuridiche soggettive fondamentali, l'AIA dispone una tutela concentrata al momento preliminare della programmazione degli algoritmi e, in un certo senso, generalizzata, rivolta a una categoria assai eterogenea di diritti. Se è vero che questa scelta ha il pregio di permettere un rapido adeguamento agli sviluppi futuri dell'Intelligenza Artificiale, è pure vero che i diritti umani singolarmente considerati si ritrovano carenti di copertura normativa precisa e specifica, in grado di fornire adeguata tutela alle caratteristiche singolari di ognuno.

Le valutazioni concrete sul punto, evidentemente, si svolgeranno in seguito all'entrata in vigore del Regolamento e della potenziale legislazione elaborata in seno all'ordinamento italiano. Allo stato attuale, tuttavia, a fronte delle potenziali criticità determinate dall'AI e della prima lettura delle fonti normative in materia, pare che le riflessioni in materia siano tutt'altro che concluse.

[97] R. Santoro-F. Gravino, *Internet, culture e religioni. Spunti di riflessione per un web interculturale*, in *Stato, Chiese e Pluralismo Confessionale*, 20, 2020, 99 ss., spec. 99.

Neurolaw, Neurorights and Neuroprivacy: Theoretical and Constitutional Issues

Francesco Cirillo

Abstract

Advancements in neurosciences and neurotechnologies have prompted the proposal of new neurorights to address the unique protection needs arising from risks of technological interference. This work examines the critical and doctrinal questions sorrounding neurorights, starting with the interaction between cognitive sciences and the conceptual categories of legal culture. The paper explores neurotechnological practices and their potential risks to individual rights, focusing on current legal frameworks in biolaw, criminal procedure, and data protection. The analysis reviews key neurorights proposals, such as cognitive liberty, mental privacy, and psychological continuity, and discusses the theoretical and practical challenges in affirming these protections within the broader legal and ethical context.

I progressi nelle neuroscienze e nelle neurotecnologie hanno condotto alla proposta di nuovi neurodiritti per rispondere alle particolari esigenze di protezione derivanti dai rischi delle interferenze tecnologiche. Questo lavoro esamina le questioni critiche e dogmatiche dei neurodiritti, muovendo dall'interazione tra le scienze cognitive e le categorie concettuali della cultura giuridica. Esplora le tecniche neurotecnologiche e i loro potenziali rischi per i diritti delle persone, concentrandosi sulle questioni giuridiche attuali nel biodiritto, nella giustizia penale e nella protezione dei dati. L'analisi si sofferma sulle principali proposte in tema di neurodiritti, come la libertà cognitiva, la *privacy* mentale e la continuità psicologica, e discute le sfide teoriche e pratiche nell'affermare queste protezioni all'interno del più ampio contesto giuridico ed etico.

Table of contents

Keywords

Neurolaw – neuorights – neuroprivacy – neural data protection – neurotechnology

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

1. Neurorights and Neuroprivacy

The progressive invention of increasingly effective neurotechnology for extracting information related to cognitive activity and manipulating cognitive processes has led some scholars to highlight the necessity of new neurorights[1].

From this perspective, the issue of neuro-rights emerges as a complex and intellectually stimulating manifestation of the debate on "new" rights[2]. A new technology appears on the scene, presenting unprecedented risks to individuals; thus, a debate arises on the assertion of new rights. In this debate, typically opposing positions clash: some argue for the necessity of recognizing autonomous rights, others attempt to relate the expectations of protection to already recognized rights or deny that such a necessity exists, with various possible intermediary positions[3]. This pattern, for example, characterized the emergence of the right to privacy[4]: With the advent of new media, American scholars highlighted the relevance of a new right[5], which perhaps was not entirely new as it emphasized new dimensions of protection for already recognized rights, and thus, through a complex process, led to the recognition of new claims, gradually made effective by court case-law[6]. Nonetheless, the same has happened for data protection, which appeared as an instrumental aspect of privacy, or for the new dimensions of the freedom of emotional or sexual relationships.

However, the issue of neurorights is not merely a variation of the problem of new rights emerging with societal and technological changes. Indeed, neurorights necessitate reflection on general legal theory, the metaprinciples that inspire it, and the possible doctrinal solutions that should accompany their affirmation[7]. Emblematically, the assertion of the right to free will, cognitive liberty, or mental privacy raises far more

[1] R. Yuste - S. Goering *et al.*, *Four Ethical Priorities for Neurotechnologies and AI*, in *Nature*, 551, 2017, 159 ss.; M. Ienca - R. Andorno, *Towards New Human Rights in the Age of Neuroscience and Neurotechnology*, in *Life Sciences, Society and Policy*, 13, 2017, 1 ss.; O. Pollicino, *Costituzionalismo, privacy e neurodiritti*, in *Rivista di diritto dei media,* 2, 2021, 9 ss.; R. de Asís Roig, *Derechos y tecnologías*, Madrid, 2022, 123 ss.

[2] E.g. F. Modugno, *I «nuovi diritti» nella giurisprudenza costituzionale*, Torino, 1995, 1 ss.

[3] *Ex multis*, P. Alston, *Making Space for New Human Rights: The Case of the Right to Development*, in *Harvard Human Rights Yearbook*, 1, 1988, 3 ss.

[4] See M. Luciani, *Il diritto al rispetto della vita privata: le sfide digitali, una prospettiva di diritto comparato*, Studi del Servizio Ricerca del Parlamento europeo, Bruxelles, 2018, 1 ss.; A. Cerri, *Riservatezza (Diritto alla), Diritto costituzionale*, in *Enciclopedia giuridica*, vol. XXVII, Rome, 1991. A broad picture of the first decades of the right in W.L. Prosser, *Privacy*, in *California Law Review*, 3, 1960.

[5] Emblematically, S. Warren - L. Brandeis, *The Right to Privacy*, in *Harvard Law Review*, 5, 1890, 193 ss., who present their "new" right as in reality an ancient and always recognized law, which just shows a "new" aspect in front of new media.

[6] A short map of this route: A. Lukács, *What is Privacy? The History and Definition of Privacy*, in G. Keresztes (ed.), *Tavaszi Szél*, Budapest, 256-265.

[7] S. Fuselli, *Neurodiritto. Per una introduzione*, in Id. (ed.), *Neurodiritto. Prospettive epistemologiche, antropologiche e giuridiche*, Milano, 2016, 7 ss.; A. D'Aloia, *Law Challenged. Reasoning About Neuroscience and Law*, in A. D'Aloia - M.C. Errigo, *Neuroscience and Law. Complicated Crossings and New Perspectives*, Cham, 2020; or V. D'Antonio - G. Sica, *Neurodiritti* e mental privacy*: alla ricerca di un* framework *normativo,* in V. D'Antonio (ed.), *Diritti digitali*, Scafati, 2022, 293 ss.

complex issues than those involved in extending the content of a right or asserting a new claim. Free will, cognitive activity, mind, psyche, consciousness, or memory are challenging to treat precisely as the object or content of an individual right because the autonomy of will and the psychological dimension belong to the extra-juridical or meta-juridical foundation of the legal order, and of fundamental rights[8]. They are the (non-juridical) premise of a theoretical construction that derives the recognition of certain fundamental rights from a specific anthropological vision, according to various and converging views, the individual has rights to freedom precisely because they possess free will or autonomy of will. Whereas it is much more challenging to assert rights to freedom so that such autonomy is preserved or promoted. In this shift, the autonomy of will moves from the meta-juridical level to the general theory of law, eventually finding its place within doctrinal discourse (finding its place within the dogmatic level of legal doctrinal discourse).

For these reasons, the issue of neurorights has a significant philosophical-legal dimension concerning both the non-juridical premises of their affirmation and the connections with cognitive sciences, as well as the theoretical-dogmatic aspects of their "construction". Furthermore, as potential new (fundamental) rights, their emergence also poses a constitutional problem, as they intertwine with already recognized rights (on the national, European, and international level). Lastly, the issue can—and should—be examined from specific disciplinary perspectives. Neurorights engage with a range of issues well-known to criminal law and criminal procedure law (for instance, neuroscientific evidence in trials); they can be viewed from the perspective of biolaw, as neurotechnology is a practice accessing the biological realm, similar to surgery or pharmacology; and finally, especially concerning neural data protection, they fit within the broader debate on regulating the digital environment (consider, for example, the current reflection on the manipulative potential of AI[9]). This interdisciplinary nature of neurorights underscores the breadth and depth of the topic, engaging scholars from various fields[10].

Bringing together these distinct planes is neither an easy operation nor a fully achievable objective. Nonetheless, the current state of the debate and its ability to permeate different disciplinary sectors make an extensive and nuanced view increasingly necessary, aiming to outline a possible common framework within which to inscribe future research directions. Moreover, as illustrated in the following paragraphs, each level allows for illustrating or characterizing specific classes of rights proposed in the literature. The philosophical-legal reflection around the paradigm of cognitive sciences

[8] A picture of the philosophical debate in D.A.J. Richards, *Rights and Autonomy*, in *Ethics*, 1, 1981, 3 ss. See also J. Kiper, *Do Human Rights Have Religious Foundations?*, in *Religion & Human Rights*, 2, 2012, 109 ss. The classic thesis of the religious foundation of metaprinciples can be traced back to E.W. Böckenförde, *The Fundamental Right of Freedom of Conscience* (1970), now in M. Künkler - T. Stein (eds.), *Religion, Law, and Democracy: Selected Writings*, Oxford, 2020, 168 ss., or to C. Schmitt, *Political theology: Four Chapters on The Concept of Sovereignty* (1922), Chicago, 2005, particularly through his argument that modern political concepts are secularized theological concepts (free will and autonomy, in this case).

[9] . *AI Act*, art. 5; on this topic, R. Uuk, *Manipulation and the AI Act*, Brussels, 2022, 2-5; or M. Ienca, *On Artificial Intelligence and Manipulation*, in *Topoi*, 3, 2023, 833 ss.

[10] *Inter alia*, L. Palazzani, *Dalla bio-etica alla tecno-etica: nuove sfide al diritto*, Torino, 2017, 333 ss.; S. Amato, *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Torino, 2020, 113 ss.

(§ 2) enables the discussion of the criticalities of the right to free will, the so-called cognitive liberty (such as the autonomy of cognitive activity as a doctrinal object). The realm of criminal justice and procedure (§ 3) provides a framework for neurorights as positions implicit in the right to a fair trial and evidence or the free expression of thought (and its free omission): namely, those neurorights formulated as aspects of privacy *versus* authority. The biolegal field allows for projecting the issue of neurorights into the context of bodily rights, thus addressing the so-called *habeas mentem*[11], psychic integrity, psychological continuity, and other related concepts. Finally, the area of data protection offers an apparently specific yet potentially expansive viewpoint: technologies for processing personal data (such as personal neural data)—that is, any neurotechnology that utilizes data processing—fall within the application scope of data regulation; hence, the issues related to a possible autonomous category of psychic or neural data, and those related to the conditions and limits of their processing. Given these considerations, it will be possible to outline the main issues on the philosophical-legal, general theoretical, doctrinal, and positive legal levels, and to demonstrate that the concept of neuroprivacy, although ambiguous like the concept of privacy itself, is perhaps more suitable for addressing the issues discussed in the literature without hastening the affirmation of new rights fraught with conceptual ambiguities.

2. Neurolaw as Cognitive Science?

From an initial perspective, like neuroethics[12], neurolaw aspires to be an interdisciplinary field of study where legal disciplines meet cognitive sciences or, where possible, neuroscience in the strict sense[13]. In other words, alongside a law of neuroscience (that is, a law regulating neuroscience and neurotechnologies), the past few decades have seen the scientific ambition to construct a neuroscience of law, which is the study of legally relevant issues through the lenses of cognitive sciences[14]. We will not dwell here on the ambiguities of referring to the neural level (of the prefix '-neuro'), nor on its appropriateness, but we can immediately observe that the attempt to reshape legal categories based on the "knowledge" of psychic activity is certainly not new.

[11] A principle that is not brand new: F.H. Sanford, *Creative Health and the Principle of Habeas Mentem*, in *American Journal of Public Health and the Nations Health*, 2, 1956, 139-148. See also A. Barbera, *Un moderno "Habeas Corpus"*, in *Costituzione Economia Globalizzazione. Liber amicorum in onore di Carlo Amirante*, Napoli, 57 ss.; or A. Baldassarre, *Diritti della persona e valori costituzionali*, Turin, 1997, 372 ss. For some authors it is about that old general right/concept of moral freedom (which does not exist according to A. Pace, *Problematica delle libertà costituzionali*, Padova, 1992).

[12] For the first concept see A. Roskies, *Neuroethics*, in *Stanford Encyclopedia of Philosophy*, 2021; for the second one, recently G.D. Caruso, *Neurolaw*. Cambridge, 2024; or S.M. Wolf, *Neurolaw: The Big Question*, in *American Journal of* Bioethics, 8, 2008, 21–22. The term "neurolaw" is over thirty years old: J.S. Taylor - J.A. Harp - T. Elliott, *Neuropsychologists and Neurolawyers*, in *Neuropsychology*, 5, 1991, 293 ss.

[13] J.A. Chandler, *Neurolaw and Neuroethics*, in *Cambridge Quarterly of Healthcare Ethics*, 4, 2018, 590-598.

[14] A meta-analysis of the «extraordinary growth in the amount of legal scholarship, legal practice, and public policy at the intersection of law and neuroscience» in F.X. Shen, *The Law and Neuroscience Bibliography: Navigating the Emerging Field of Neurolaw*, in *International Journal of Legal Information*, 3, 2010, 352 ss.

In the Italian context, at the end of the 19th century a broad debate arose between two Schools of criminal law, the Classical and the Positive, the former leaning towards preserving the traditional concepts of criminal law (primarily, the freedom-responsibility binomial, from which the retributive function of punishment derives), and the latter advocating for criminal justice far from the category of guilt and closer to that of dangerousness[15]. The Italian Positive School had obvious strong connections with the broader positivist ambitions[16] of the late 19th century, and its fortune did not differ from that of positivism *tout-court*: an ambitious and all-encompassing research program, yet often relying on simplistic worldviews and excessively rigid deterministic models, frequently destined to end in paradoxically unscientific results, as well as often veering towards racist or reactionary tendencies. An exemplary case was precisely the now outdated results of the Positive School, which, in its attempt to offer an early criminology and a scientific theory of delinquency, ended up producing a rather confused "system" imbued with prejudices and racism[17].

These early premature attempts to introduce a form of (pseudo)scientific determinism into the law—described by Foucault and authors inspired by him as the medicalisation of criminal law[18]—encountered a progressive failure but left peculiar legacies in legal systems. This initial phase can be linked to technologies that aim to "read" the minds of defendants or reduce their tendency towards reticence, such as lie detectors or truth serums, which, transplanted from European positivism into other cultures, are still used in various countries today. In the next paragraph, we will focus on this aspect, analysing the relationships between neurolaw and criminal justice. In any case, the premature ambition of the late 19th century found new fortune precisely in the wake of developments in neuroscience over the last few decades. Thus, the increasing ability to read cognitive activity and manipulate mental processes has led, on one hand, to the belief that humans are much less free than they assume to be and, on the other hand, to highlight the risk of manipulative intervention by these technologies. Emblematically, we could consider the literature that evaluated the impact of Libet's experiments in the law context[19], but the discussion did not only involve the free will.

---

[15] M.A. Musmanno, *The Italian Positive School of Criminology*, in *American Bar Association Journal*, 7, 1925, 427-430, where a singular interview with Enrico Ferri, one of the founders of the Italian Positive School, is reported: «The fundamental principle embraced in our school goes back to Galileo Galilei and Leonardo da Vinci, a principle which was followed, disciplined and systematized by Francis Bacon in his "Novum Organum", namely, the inductive method of reasoning which consists in observing facts, particularizing, classifying and reproducing them in experiments where possible, and then drawing from them the general conclusions or the legal norm».

[16] It is not referred to the so-called legal positivism or any positive legal theory («the thesis that the existence and content of law depends on social facts and not on its merits», according to L. Green, *Legal Positivism*, in *Stanford Encyclopedia of Philosophy*, 2003), but it is referring to the philosophical ambition of reducing the social sciences to deterministic paradigms (even in the absence of evidence of the validity of the explanatory models), like in the Comtian thought.

[17] The case of Lombroso, *inter alia*, S. Montaldo, *Lombroso: The Myth, the History*, in *Crime, Histoire & Sociétés*, 2, 2018, 31 ss. See also F. Rotondo, *Un dibattito per l'egemonia. La perizia medico legale nel processo penale italiano di fine Ottocento*, in *Rechtsgeschichte*, 12, 2008, 139 ss.

[18] Especially in M. Foucault, *Les anormaux: cours au Collège de France (1974-1975)*, Paris, 1999. See also M. Mitjavila, P. Mathes, *Labyrinths of Crime Medicalization*, in *Saúde e Sociedade*, 25, 2016, 847 ss.

[19] S. Pockett, *The Concept of Free Will: Philosophy, Neuroscience and the Law*, in *Behavioral Sciences & the Law*,

Consider, for example, the impact of neuroscience on our (still vague) understanding of memory, the claims to extract information or to manipulate its processes[20].
Indeed, in this field, the renewed confidence that science offers reductionistic explanations of cognitive activity leads some authors to propose a complete revision of legal categories, from imputability to autonomy, from culpability to intentionality[21]. This hypothesis generates all kinds of criticisms, as the autonomy of law is asserted, the reductionistic paradigm of the hard sciences is opposed (especially in some contexts of the humanities), and the process of "naturalization" of law is criticised, as it is seen as bending to logics that are foreign to it[22]. From this perspective, the debate surrounding the concept of neurolaw is a specific manifestation of the broader debate between the "two cultures"[23], with proponents on one side advocating for the absolute uniqueness of the human being (and its dignity), and on the other, a tendency to encompass the human being within a naturalistic paradigm.
Nonetheless, it is likely that this debate will not cross the thresholds of academia: Even if one were to embrace a reductionist and materialistic view of cognitive activity, it would be unclear with which categories to reconstitute a law devoid of the presupposition of the autonomy or free will. Furthermore, for those who expect neuroscience to "prove" the non-existence of free will, it is likely that this result will never be achieved, partly because there is no clear concept of free will, and partly because, as in any field, it is impossible to prove the non-existence of something[24]. Additionally, despite many advances in the field of neuroscience, very few results seem robust enough to be exported into the theory and practice of organising human societies, as

2, 2007, 281 ss.

[20] On the memory detection see again S.M. Wolf, *Neurolaw: The Big Question*, in *The American Journal of Bioethics*, 1, 2008, 21-22; P. Catley, *The Future of Neurolaw*, in *European Journal of Current Legal Issues*, 2, 2016; D.V. Meegan, *Neuroimaging Techniques for Memory Detection: Scientific, Ethical, and Legal Issues*, in *The American Journal of Bioethics*, 8, 2008, 9 ss.; or A. Farano, *Neuroscienze e diritto: un primo bilancio*, in S. Salardi – M. Saporiti (eds.), *Le tecnologie 'morali' emergenti e le sfide etico-giuridiche delle nuove soggettività*, Torino, 2020, 42 ss.

[21] F. Corso – A. Lavazza, *Neuroetica e neurodiritto: fine dell'imputabilità?*, in M.F. Pacitto (ed.), *Neuroetica. Convegni cassinati. Scuola di Alta Formazione in Neuroetica e Filosofia delle Neuroscienze*, Roma, 2020, 153 ss. See also M.C. Errigo, *Neuroscienze, tecnologia e diritti: problemi nuovi e ipotesi di tutela*, in *Dirittifondamentali.it*, 3, 2020, 244 ss.

[22] C. Sarra, *Questioni pregiudiziali*, in S. Fuselli (ed.), *Neurodiritto*, cit., 78 ss.; A. Pirozzoli, *La libertà di coscienza e le neuroscienze cognitive*, in *Consulta OnLine, Liber amicorum per Pasquale Costanzo*, 2020, 6; N. Irti, *L'uso giuridico della natura*, Roma-Bari, 2013, 33. Some authors speak about neuroessentialism or ruthless reductionism. «Neuroessentialism is the position that, for all intents and purposes, we are our brains» [B. P. Reiner, *The Rise of Neuroessentialism*, in J. Illes - B. Sahakian (ed.), *The Oxford Handbook of Neuroethics*, Oxford, 2011, 1]. About ruthless reductionism, J. Bickle, *Philosophy and Neuroscience: A Ruthlessly Reductive Account*, Alphen aan den Rijn, 2003.

[23] The formula of C.P. Percy, *The Two Cultures*, London, 1959: «Literary intellectuals at one pole—at the other scientists, and as the most representative, the physical scientists. Between the two a gulf of mutual incomprehension—sometimes (particularly among the young) hostility and dislike, but most of all lack of understanding». Some remarks in G. Lumia, *Il diritto tra le due culture*, Milano, 1971. On this relationship see the conclusions of P. Sommaggio, *Neurocivilizzazione o libertà cognitiva?*, in S. Fuselli (ed.), *Neurodiritto*, cit., 150 ss.

[24] A typical logical fallacy (I. Copi – C. Cohen, *Introduction to Logic*, Harlow, 2014, 132–133). On the free will debate: M. De Caro - A. Lavazza, *La libertà nell'era della scienza*, in M. De Caro, A. Lavazza - G. Sartori (eds.), *Siamo davvero liberi? Le neuroscienze e il mistero del libero arbitrio*, Torino, 2019, VII ss.

the case of law[25].

For all these reasons, although further research is desirable, neurolaw, understood as the neuroscience of law, or as the cognitive sciences of the legal phenomenon, represents a program more than a reality, an ambition more than a body of knowledge.

The other aspect of neurolaw, the law of neuroscience, involves no lesser critical issues. The general assumption driving authors, particularly in the bioethical and neuroscientific fields, but today also in legal contexts, is that neurotechnology poses a risk to humans due to their interference with cognitive activity (whether extracting information or manipulative or inductive interference).

The vast majority of neurotechnology applications are currently confined to the fields of therapy and scientific experimentation, so the issue seems to reduce to the development of good practices[26] for clinical experimentation on humans (consider the protocols approved by the Food and Drug Administration for Neuralink's brain-computer implants[27]). In contrast, the main concerns are directed at a potential widespread use of these technologies in various sectors of society[28] both by private stakeholders and in the relationships between citizens and authorities, envisioning forms of profiling, control, and manipulation that pose risks to individual freedom and dignity.

Certainly, some authors deploy a general fear of technologies toward neurotechnologies, a fear that seems prevalent in public debate from artificial intelligence to biotechnology[29]. But even steering clear of pessimistic views and irrational fears, one must acknowledge that concerns about potential cognitive manipulation are justified. It is primarily based on these considerations that proposals for new neurorights emerged within the academic community to address these new risks. Some of these proposals were favourably received by legislators and governments, as seen in the case of Chile[30], the Spanish government's digital rights charter[31], the soft law documentation of many

[25] I am not referring to individual applications in the context of criminal proceedings, where neuroscience could well progressively support or replace psychiatric expertise (for the current value of neuroimaging in insanity assessments see for instance G. Meynen, *Neuroscience-Based Psychiatric Assessments of Criminal Responsibility: Beyond Self-Report?*, in *Cambridge Quarterly of Healthcare Ethics*, 3, 2020, 446 ss.). *Contra*, a "dystopic" example: J.M.R. Delgado, *Physical Control of the Mind. Toward a Psychocivilized Society*, New York-London, 1971.

[26] See also R. Yuste, *Advocating for Neurodata Privacy and Neurotechnology Regulation*, in *Nature Protocols*, 18, 2023, 2869 ss.

[27] A. J. Jawad, *Engineering Ethics of Neuralink Brain Computer Interfaces Devices*, in *Perspective*, 4, 2021; D. Hurley, *Ethical Questions Swirl Around Neuralink's Computer-Brain Implants*, in *Neurology Today*, 10, 2024, 1 ss.

[28] *The Neurorights Foundation: Market Analysis: Neurotechnology*, at *neurorightsfoundation.org*, 2023; see also the report of J. Genser - S. Damianos - R. Yuste (eds.), *Safeguarding Brain Data: Assessing the Privacy Practices of Consumer Neurotechnology Companies*, at *neurorightsfoundation.org*, April 2023.

[29] A lucid criticism is in V. Zeno-Zencovich, *Artificial Intelligence, Natural Stupidity and Other Legal Idiocies*, in *MediaLaws*, 1, 2024.

[30] The new art. 19 of the proposed Constitution: «*La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella*». See. P. López Silva - R. Madrid, *Acerca de la protección constitucional de los neuroderechos: la innovación chilena*, in *Prudentia Iuris*, 94, 2022, 39 ss.

[31] *Carta de Derechos Digitales* adopted by Spanish Governament, 14 July 2021, § XXVI.

international and European institutions[32], and most recently, the European legislator's consideration of the manipulative risks of AI[33].

However, this aspect of neurolaw, now as the law of neurotechnology, also presents philosophical, theoretical, doctrinal, and constitutional problems. Notably, the proposed catalogs of neurorights differ in identifying which rights they encompass[34].

While it is possible, as we will try to do, to identify homogeneous classes of rights referable to individual proposals, these rights appear, on the one hand, already affirmed in many legal traditions, and on the other hand, grounded in a rather obscure theoretical framework[35].

The class of rights that presents the most critical issues from a philosophical and general-theoretical perspective includes the right to free will, cognitive liberty, the right to cognitive autonomy. In the case of these rights, the autonomy of will, variously named, which often features in legal philosophy as a metaprinciple akin to dignity in the foundation of human rights[36]—an assumed value from which the existence of rights derives—becomes the content or object of the right. As if an individual, just as they can claim the right to move or express themselves freely when impeded, could similarly claim to be free in their will when manipulated by others. This presents a glaring recursion, like the image of Baron Munchausen lifting himself out of the swamp by his hair[37]. Besides the theoretical paradox inherent in such a right, what is more surprising is that the proposal embraces the reductionist paradigm of neuroscience, admitting that cognitive activity is entirely predictable and manipulable; yet it repudiates this paradigm by asserting free will. In other words, if free will truly existed as an indeterminate principle immune to external factors, then it would not be subject to manipulation by neurotechnology.

The glaring recursion and the contradiction inherent in admitting free will while denying its presence suggest that the issue is probably misconceived. If one assumes the paradigm of neuroscience, that is, assuming that cognitive activity is nothing different

---

[32] UNESCO, *Report of the International Bioethics Committee of UNESCO (IBC) on ethical issues of neurotechnology*, SHS/BIO/IBC-28/2021/3, § 190; OCSE, *Recommendation on Responsible Innovation in Neurotechnology*, 11 December 2019,

[33] Resolution of EP on «*Artificial Intelligence in a Digital Age*» (2020/2266(INI), 3 May 2022, § 247; and *funditus* in the *AI Act* (*e.g.* art. 5).

[34] For example, according to Ienca and Andorno, four neurorights can be identified: cognitive liberty, mental privacy, mental integrity, and psychological continuity. The Neurorights Initiative at Columbia University proposes five neuro-rights: mental privacy, personal identity, free will, fair access to mental enhancement, and protection from algorithmic bias. Based on this proposal and the amended art. 19 of the Chilean Constitution, the bill on the protection of neurorights and mental integrity, and the development of research and neurotechnologies, reflects the Neurorights Initiative's proposal: prohibiting neurotechnological interference that harms the psychological and mental continuity of a person, personal identity, autonomy of will, and the ability to make decisions freely, and protecting the mental substrate of personal identity (art. 4); elevating neural data to a special category of health data (art. 6), subjecting their dissemination and transmission to organ transplant regulations (sic, art. 7); and promoting fair access to neurotechnologies (art. 10).

[35] A critical analysis in J.C. Bublitz, *Novel Neurorights: From Nonsense to Substance*, in *Neuroethics*, 17, 2022, 12.

[36] For instance, N. Bobbio, *Libertà*, in *Enciclopedia del Novecento*, Rome, 1978, § 4;

[37] A quite common *topos*: yet in R. von Jhering, *Der Zweck im Recht*, II ed., Leipzig, 1884, 3-4.

from other natural phenomena, then free will does not exist, and claiming it would make no more sense than claiming a theological or abstract object as a human right. However, this paradigm is not compatible with the current dogmatics of fundamental rights and principles such as culpability or responsibility.
Conversely, by adopting a mentalist, psychological paradigm, or one that nonetheless affirms the unique indeterminacy of human choices—a common assumption in all Western legal traditions—the foundational metaprinciples of legal categories are justified, but the manipulative potential of technologies cannot be fully embraced.
These incompatibilities between paradigms can also be observed in the other perspectives discussed in the next three paragraphs (criminal proceedings, biolaw, and data protection), but they are certainly more evident when one seeks to affirm a right to free will or the right to control over one's mental states.

3. Neurolaw and Criminal Justice

The field of law that first encountered "knowledge" (scientific or not) aimed at the psychic dimension was, as previously mentioned, criminal law and criminal proceedings. Following the initial attempts to establish a new theory and practice oriented toward psychology (as in the case of the Italian Positive School), criminal proceedings have since maintained varying degrees of intersection with psychology, psychiatry, and, more recently, neuroscience[38]. In criminal law, it is alongside the asserted relevance of the criminal (f)act that the importance of psychological elements emerges: the imputability of the offender, guilt and intent, the genuineness of testimony; all aspects that refer to awareness, intentionality, agency, memory, and so on.
However, the intersections between psychology and criminal law have not always been successful, to the extent that legislators often limited the use of psychiatric evidence in criminal proceedings to avoid determining solutions or culpability based on ambiguous psychological hypotheses. A notable example is the Italian Code of Criminal Procedure of 1930, which, despite being drafted in an authoritarian context, took care to exclude psychiatric evidence for assessing culpability (art. 314, para. 2)[39].
Such limits persisted in legal systems, and, in the Italian case, they even found justification on constitutional grounds that were not present at their inception. Indeed, the current prohibitions on using lie detectors or truth tests are justified by the Constitutional Court as being detrimental to an individual's moral freedom (or moral autonomy)[40], a principle that, though philosophically or religiously framed, does not seem far from the more secular concept of cognitive liberty discussed earlier[41].

[38] O. Di Giovine, *Ripensare il diritto penale attraverso le (neuro-)scienze?*, Torino, 2019, 17 ss.

[39] Similarly, now art. 220 Italian Code of Criminal Procedure.

[40] Constitutional Court, decisions nos. 124/1970, 179/1973, 229/1998.

[41] G. Vassalli, *Il diritto alla libertà morale (Contributo alla teoria dei diritti della personalità)*, in *Studi in memoria di Filippo Vassalli*, vol. II, Torino, 1960, 1670-1701; but see also A. Barbera, *I principi costituzionali della libertà personale*, Milano, 1967; A. Bonomi, *Le neuroscienze in rapporto alla libertà morale: aspetti di diritto costituzionale*, in *Forum di Quaderni Costituzionali*, 2018, 1 ss.; or A. Santosuosso - B. Bottalico, *Neuroscienze e genetica comportamentale nel processo penale italiano. Casi e prospettive*, in *Rassegna italiana di criminologia*, 6, 2013, 72 ss.

Nonetheless, with the invention of new neuroscientific diagnostic tools, the debate on the admissibility of neuroscientific evidence in criminal proceedings found new momentum. Several Italian legal cases, following the same pattern and analogous to those in other jurisdictions, illustrate the current situation[42]. During the trial, the defendant claims their right to use any means of evidence to prove their innocence, proposing to undergo a neuroscientific test. The most common case involves tests (I.A.T. or autobiographical Implicit Association Test, and T.A.R.A., or Timed Antagonistic Response Alethiometer) to detect mnemonic traces of violent episodes or, based on the absence of such traces, to provide elements supporting the defendant's innocence. Thus, on one side, there is the defendant's right to evidence, and on the other, their "moral freedom", which would prevent the use of neurotechnological devices in criminal proceedings. Currently, the interpretation by the Supreme Court and the Constitutional Court favours the inviolability of moral freedom over the right to evidence. However, the unreliability of the diagnostic tests plays a certain role in the background of these cases.

Indeed, the courts frequently refer to three elements: the explicit prohibition by the legislator; the constitutional legitimacy of the prohibition, ostensibly to protect the constitutional value of moral freedom; and lastly, questioning the relevance of the first two arguments, the scientific unreliability of the diagnostic tests. However, imagine that a new diagnostic test could reliably detect traces of a criminal event in an individual's memory or equally reliably ensure the absence of such traces. Could the inviolability of cognitive freedom justify prohibiting a defendant from using such enlightening evidence? Moreover, what would happen if the test produced an opposite result? Could a judge disclose traces of a criminal event in memory as a basis for a verdict?

Similar issues arise in different contexts: assessing the offender's imputability, where psychiatric evidence is admitted and can utilize neuroscientific tools; evaluating subjective elements, intent, and guilt; assessing testimony, and so forth. All actors of criminal procedure are involved in the "neurohype"[43], not least the judge, whose decision could be subject to psychological or neuroscientific evaluation[44].

In any case, the introduction of neuroscientific tools in criminal proceedings is strongly limited by two factors: the explicit presence of prohibitions on using psychiatric or neuroscientific evidence at certain trial stages and the substantial unreliability of the

[42] A. Farano, *Neuroscienze e diritto*, in S. Salardi, M. Saporiti (eds.), *Le tecnologie 'morali' emergenti e le sfide etico-giuridiche delle nuove soggettività*, Torino, 2020, 48 ss.; G. Gullotta - M. Caponi Beltramo, *Neurodiritti: tra tutela e responsabilità*, in *Sistema penale*, 1 October 2021, 7 ss.; G. Gennari, *Oscillazioni neuro...scientifiche: test a-IAT e macchina della verità*, in *Sistema penale*, 10 December 2020; O. Di Giovine, *Prove di dialogo tra neuroscienze e diritto penale*, in *Giornale italiano di psicologia*, 4, 2016, 336; G. di Chiara, *Il canto delle sirene. Processo penale e modernità scientifico-tecnologica: prova dichiarativa e diagnostica della verità*, in *Criminalia*, 2007.

[43] A. Wexler, *Separating Neuroethics from Neurohype*, in *Nature Biotechnology*, 9 August 2019, 988 ss.

[44] Recently A. Santosuosso - M. Giustiniani, *Vulnerable Defendants: Redefining Decision-Making through the Lenses of Neuroscience, Law and Artificial Intelligence*, in H. Wishart, CM Berryessa (eds.), *Neurolaw in the Courtroom. Comparative Perspectives on Vulnerable Defendants*, Abingdon-NewYork, 2024, 37 ss.; O.D. Jones - J.D. Schall - F.X. Shen - M.B. Hoffman - A.D. Wagner, *Brain Science for Lawyers, Judges, and Policymakers*, Oxford, 2024; M.A. Thomaidou - C.M. Berryessa, *Bio-Behavioral Scientific Evidence Alters Judges' Sentencing Decision-Making: A Quantitative Analysis*, in *International Journal of Law and Psychiatry*, 95, 2024.

scientific results of the tests.
However, both factors may only have a historical character and may prove precarious. Legislative limits can be reviewed or overcome due to technological improvements. However, even their constitutional legitimacy, the constitutional legitimacy of the prohibitions, could face a similar fate. This can happen, especially considering the growing relevance of scientific evidence in the judicial review of laws[45]. Particularly in cases involving a collision of fundamental rights (to evidence and moral freedom), the judgment on the reasonableness of the legislative prohibition can well be grounded in scientific evaluations. Consider, notably, the pandemic context, where the proportionality of legislative prohibitions often relied on scientific parameters related to the necessity and suitability of imposed measures[46]. Based on these premises, it is one thing to affirm the constitutional illegitimacy of a technologically unreliable device; it is another to persist in deeming the prohibition of scientifically founded evidence legitimate.
Future debate may be framed within this theoretical context, particularly due to the development of technologies far more reliable than those now clumsily attempting to enter criminal proceedings.

4. Neurolaw as Biolaw

Suppose the philosophical perspective and criminal law reveal a web of complex relationships between law and neuroscience, in which the fundamental questions may be unresolvable. Then, the perspective of biolaw is more solid and perhaps more oriented towards positive law. Neurotechnological devices are predominantly and most effectively used in research and medical therapy, and it is in these sectors that a regulatory dimension can be observed. The law of research, experimentation, and medical therapy is a very specific area. However, the field may be the laboratory for observing issues that will interest other sectors of society sooner or later.
If we adopt the perspective of the "body", perhaps even with ruthless reductionism, the interference of neurotechnology on the human being appears as a practice of accessing the body itself: these are interventions with therapeutic purposes, possibly also for experimentation and research, which as such are subject to a complex framework of rules on clinical experimentation. At this level, all proposals for those neurorights

[45] A. Ruggeri, *Diritti fondamentali e scienza: un rapporto complesso*, in *Consulta OnLine*, I, 2022, 251 ss.; G. Fontana, *Tecnoscienza e diritto al tempo della pandemia (considerazioni critiche sulla riserva di scienza)*, in *Osservatorio sulle fonti*, 1, 2022, 808; F. Pacini, *Ai confini della normatività.* Hard law *e* soft law *in "tempi difficili"*, in *Gruppodipisa.it*, 18 June 2022, 7 ss.; S. Penasa, *Il dato scientifico nella giurisprudenza della Corte costituzionale: la ragionevolezza scientifica come sintesi tra dimensione scientifica e dimensione assiologica*, in *Politica del diritto*, 2, 2015, 295 ss.; C. Casonato, *La scienza come parametro interposto di costituzionalità*, in *Rivista AIC*, 2, 2016, 5 ss.; S. Zorzetto, *Dal "sogno cartesiano" alla "razionalità limitata": usi e abusi della scienza nella politica legislativa*, in Ead. - F. Ferraro (eds.), *La motivazione delle leggi*, Torino, 2019, 167 ss.; P. Veronesi, *La Corte costituzionale e la scienza: alcune tendenze e punti fermi*, in *BioLaw Journal*, 2, 2024, 125 ss.; L. Busatta, *Tra scienza e norma: il fattore scientifico come oggetto, strumento e soggetto della regolazione*, in *Costituzionalismo.it*, 1, 2021, 143 ss.

[46] Constitutional Court., decision no. 14/2023; see also F. Girelli - F. Cirillo, *Immuni e green pass. Prospettive di bilanciamento nella pandemia*, in *Consulta Online*, 1, 2022, 254 ss.

that already seem protected by the legal systems of Western traditions should be considered, such as psychophysical integrity or the right to access therapies, albeit in new forms of human enhancement.

Reconstructing the regulatory framework is quite challenging because various historical matrices built it. Consider the affirmation of bioethical principles, gradually transferred into the legal realm, the emergence of the principle of informed consent in North American jurisprudence and then in European law, European or national laws, and international norms on clinical experimentation. We can summarize this regulatory framework in a single principle escorted by a regime of exceptions: interventions on the human body are generally prohibited unless the law and, where applicable, the individual's consent legitimizes exceptional interventions under highly limited conditions and purposes.

Viewed from another perspective, we are faced with one of the possible variations of the concept of privacy, now understood as a prohibition of interference in an individual's private life, whose most exclusive domain coincides with their body (consider the reasons that led the case-law to derive the right to abortion from the right to privacy). Regarding practices involving interventions on a person's body, such as neuropharmacology, neurosurgery, or even structural or functional neural imaging diagnostic techniques, their (legitimate) use outside a healthcare context would be difficult to envisage for factual reasons even before legal ones. This is especially true when considering that even scientific research on the human body – the so-called clinical research – must always take place within the healthcare sector, if only due to the limited availability and high costs of the necessary equipment[47].

Regarding the regulation of healthcare treatments, reference must be made to a broad framework of international, European, and domestic sources, which can only be briefly summarized. This framework includes fundamental constitutional and international principles (above all, informed consent), healthcare organization law, norms on the liability (civil or criminal) of healthcare professionals[48], technical sources (guidelines and international standards), and professional ethics. For example, at the international level, the 1964 Declaration of Helsinki[49] and the 1997 Oviedo Convention affirm a broad set of principles, from the right to privacy to the right to receive information collected about one's health, thus leading to the affirmation of the right to the protection of health-related information. In the European Union context, regarding only the sector of pharmaceutical experimentation, the need for harmonized regulations across different states led, for example, in April 2014, to the approval of Regulation EU/536/2014 on clinical trials on medicinal products for human use[50]. Similarly, in

[47] See A. Iannuzzi (ed.), *La ricerca scientifica fra possibilità e limiti*, Napoli, 2015; G. Marsico, *La sperimentazione clinica: profili bioetici*, in L. Lenti - E. Palermo Fabris - P. Zatti (eds.), *Trattato di biodiritto. I diritti in medicina*, Milano, 2011, 625 ss.

[48] E. Catelani, P. Milazzo, *La tutela della salute nella nuova legge sulla responsabilità medica. Profili di diritto costituzionale e pubblico*, in *Istituzioni del Federalismo*, 2, 2017, 305 ss.

[49] About the Declaration, U. Schmidt - A. Frewer (eds.), *History and Theory of Human Experimentation. The Declaration of Helsinki and Modern Medical Ethics*, Stuttgart, 2007.

[50] M. Fasan, C.M. Reale, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, un'occasione mancata?*, in *BioLaw Journal*, 4, 2022, 272 ss.; see also C. Casonato, *I farmaci, fra speculazioni e logiche*

the Italian context, in line with the principles mentioned above, Clinical Trial Ethics Committees[51] were established in the 1990s, progressively expanding their role to serve as an «organizational model balancing scientific freedom and the protection of individuals in biomedical research»[52]. Therapeutic activities, broadly understood to include diagnostic and clinical research activities, are overseen by the Italian National Guidelines System[53], consistent with a broader horizon of internationally derived guidelines and standards[54].

The wide range of methods and techniques mentioned in connection with neuroscience and neurotechnologies would require distinct evaluation on a case-by-case basis. Thus, a detailed examination of such a broad range of legal and technical norms is precluded here.

For the purposes of this inquiry, however, it is relevant to focus on the emergence of the principle of informed consent, which unites the entire referenced normative framework within the broader context of strengthening patient protections linked to the affirmation of the right to privacy, particularly in the North American context.[55]. Nevertheless, it is worth emphasizing the centrality of the concept of privacy, which, despite its semantic ambiguities, demonstrates sufficient flexibility to address various issues[56].

This general framework of guarantees and protections for the body already seems entirely suitable to accommodate new forms of protection against neurotechnological interference, especially because, unlike the initial levels observed, the perspective of the body adopts a paradigm wholly compatible with the reductionism that characterizes the neuroscientific approach. We are not facing a clash of paradigms again, but rather two compatible and consistent viewpoints.

costituzionali, in *Rivista AIC*, 4, 2017, *passim*.

[51] The Committees «are independent bodies responsible for ensuring the protection of the rights, safety, and well-being of individuals involved in experimentation and for providing public assurance of such protection. Where not already assigned to specific bodies, ethics committees may also perform consultative functions concerning ethical issues related to scientific and healthcare activities, with the aim of protecting and promoting the values of the individual» (Italian Minister of Health Decree, February 8, 2013, art. 1), established in the context of clinical trials, but whose consultative functions in the healthcare sector have progressively expanded. In relation to the topic of experimentation, see F. Giunta, *Lo statuto giuridico della sperimentazione clinica e il ruolo dei comitati etici*, in *Diritto pubblico*, 2, 2002, 631 ss., on the birth of the *Committees*, 634 ss.

[52] W. Gasparri, *Libertà di scienza, ricerca biomedica e comitati etici. L'organizzazione amministrativa della sperimentazione clinica dei farmaci*, in *Diritto pubblico*, 2, 2012, 586.

[53] The law 24/2017 on professional liability attributed fundamental importance to this, giving the National Institute of Health, through the National Center for Clinical Excellence, Quality, and Safety of Care, the role of methodological guarantor and national governance of the process of producing the guidelines themselves. On this topic, see C.M. Masieri, *Linee guida e responsabilità civile del medico. Dall'esperienza americana alla legge Gelli-Bianco*, Milano, 2019, 23 ss.

[54] For example, consider the system implemented thanks to the European Network for Health Technology Assessment (see *eunethta.eu*).

[55] See C. Casonato, *Il Principio di autodeterminazione. Una modellistica per inizio e fine vita*, in *Osservatorio AIC*, 1, 2022, 54 ss.; G. Razzano, *Principi costituzionali ed ambito di applicazione del consenso informato*, in *Trattato di diritto e bioetica*, Napoli, 2017, 11 ss.

[56] Regarding the preference for a broader, ambiguous, and provisional concept of neuroprivacy, see the conclusions (§ 7).

5. Neurolaw as Data Protection Law

Considering once again a different point of view, neurotechnologies almost all perform digital data processing related to individuals: this is certainly the case for diagnostic devices, but also for statistical research in neuropharmacology, robotics in neurosurgery, or brain-computer implants.

This circumstance brings the activities in question within the scope of data regulation, specifically the European Regulation EU/679/16 (GDPR), which primarily safeguards the right to data protection (enshrined in art. 8 of the Charter of Nice at the EU level and protected in the context of the Council of Europe with reference to art. 8 of the ECHR and Convention 108/1981) but also the complex balance between this right and other fundamental rights[57]. The GDPR requires, in brief, that data be processed lawfully, fairly, and transparently, collected for specified legitimate purposes and limited to them, stored for a limited time, and secured appropriately (principles set out in art. 5). It assigns tasks, responsibilities, and obligations based on the types of processing, organizational context, technologies used, purposes, types of data processed, etc.

Among the most relevant (and problematic) principles of the GDPR, art. 25 addresses data protection by design and by default, which mandates that data controllers, considering a complex set of variables[58], implement appropriate technical and organizational measures to enforce the necessary principles and safeguards to meet the requirements of this Regulation and protect the rights of data subjects. This implies that operators intending to implement data processing technology must conduct a thorough analysis of the impact on fundamental rights and design the technology and organization needed for processing activities to achieve the best balance between the involved rights and interests.

The regulation, therefore, not only predetermines balancing operations between the right to data protection and other rights or interests but also complements the role of member states (legislators, constitutional courts, ordinary courts, supervisory authorities) with that of the norm's recipients, who are involved in defining the regulatory horizon it addresses, following a model that is both self-regulatory and techno-regulatory[59].

---

[57] See C. Colapietro, *Il diritto alla protezione dei dati personali in un sistema delle fonti multilivello*, Napoli, 2018, 21 ss.; or O. Pollicino, *Judges, Privacy and Data Protection from a Multilevel Protection Perspective*, in *Federalismi.it*, 4, 2022.

[58] Art. 25, para 1.: «Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects».

[59] By self-regulation, we mean a form of regulation not unilaterally imposed by the state or another institutional actor, adopted in light of the observation of the ineffectiveness or suboptimal effectiveness of authoritative public intervention. (G. Napolitano, *Autoregolamentazione*, in *Dizionario di Economia e Finanza*, Roma, 2012). In self-regulation, the stakeholders are called to cooperate in defining the

The innovative GDPR approach is not without critical aspects, which nonetheless involves all data processing operators in defining the protection not only of the right to data protection but also of every fundamental right (including potential neuro-rights) whose consideration is determined (and, indeed, primarily) by the design of a data processing technology. This would occur even more so if we accept the perspective that algorithmic profiling produces persuasive or inductive effects that threaten individuals' cognitive autonomy: in this case, it is highlighted that the freedom of a person to self-determination would be nonetheless an inviolable constitutional right, enforceable not only against public authorities but also against private entities[60].

This raises several critical and interconnected issues.

The first is primarily the actual sustainability of fundamental rights in private relations: rights originally conceived as protections against the authority's power (and thus also as duties of state non-intervention) risk, in their horizontal projection between citizens, transforming into an obligation of state intervention in every aspect of contractual autonomy.

Furthermore, significant doubts might arise regarding the current regulation of digital data. On one hand, its appropriateness is contested due to the lag in the European digital market, and on the other extreme, because faced with an enormous number of ever-evolving rules, the effectiveness of this intricate regulatory framework appears rather weak[61].

Moving beyond these first two observations, the literature highlights the prohibition on processing special categories of data, such as those revealing ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, as well as genetic data, biometric data, and data concerning the health or sex life of the data subject (art. 9 GDPR). This prohibition, given the provision of ten classes of exceptions (among which the mere consent of the data subject stands out), rather results in a strengthened protection regime for these special categories of data[62].

Regarding neural data, some would ensure their full classification as sensitive or spe-

relevant norms. In the field of data protection and the digital context, the self-regulatory approach is increasingly favored, although it is often implemented in models where public intervention coexists with regulatory negotiation between institutional actors, such as supervisory authorities, and the stakeholders themselves. [S. Sileoni, *I codici di condotta e le funzioni di certificazione*, in V. Cuffaro - R. D'Orazio - V. Ricciuto (ed.), *I dati personali nel diritto europeo*, Torino, 2019].

By techno-regulation, we mean the ability of technical standards and design choices in the digital environment to contribute to, if not replace, the traditional normative dimension. In other words, in the design of software, digital platforms like social networks, or apps, the developer defines, based on standards and choices that are initially technical in nature, what the user can and cannot do. This creates an overlap between the writing of the computer code and the political-legal decision on the freedoms or limitations that will characterize the user experience.

60 About the concept of *Drittwirkung,* A. Lamberti, *L'ambiente digitale: una sfida per il diritto costituzionale*, in *Federalismi.it*, 4, 2022, 448. Similarly, the possibility of deriving neuroethics and the consequent necessary framework of protections against neurotechnological risks within the notion of personhood is already in S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2013, 371 ss.

61 Again V. Zeno-Zencovich, *Artificial Intelligence, Natural Stupidity*, cit.*, passim.*

62 A. Thiene*, La regola e l'eccezione. Il ruolo del consenso in relazione al trattamento dei dati sanitari alla luce dell'art. 9 GDP*R, in A. Thiene - S. Corso (eds.), *La protezione dei dati sanitari. Privacy e innovazione tecnologica tra salute pubblica e diritto alla riservatezza*, Napoli, 2023, ss.

cial data categories[63]: it is true that neural data revealing health conditions are indeed health data outright, just as neural data related to sexual activity are sexual data in the proper sense. Similarly, we can include neural data in special categories even in the case of neuro-sensitive devices capable of recognizing an individual's approval or disapproval of political news and, therefore, revealing (or allowing the prediction of) a user's political opinions (or, likewise, in the case where the data allows the prediction of the user's health conditions)[64]. Nonetheless, if the processing is not aimed at collecting or predicting information related to special categories, the protection of neural data remains subject to the ordinary safeguards provided by the GDPR framework. This limitation could imply, however, that entire sectors of neural data processing are subject to relatively weak protections: consider, for example, the exploitation of neural data as indicators of consumer engagement and attention related to products or content[65], and more generally, the uses of neurotechnologies in the various sectors of the "attention economy" (the exploitation of neural data to determine audience preferences and improve content personalization in the media and entertainment industries)[66]. To be clear, the framework of positive norms, as interpreted by supervisory authorities and case law, does not allow psychic data related to consumer preferences and engagement to be subject to enhanced protection, nor would it be possible – and perhaps not even desirable – to completely preclude platforms from using "vaguely psychic" data. A reversal of this approach, although desirable in the opinion of some, could require positive intervention by the Legislator (European or national)[67]. In this context, the unique Chilean bill stands out, which, in art. 7, would subject the processing of neural data to the provisions regarding organ transplantation (*sic!*) [68].

Such a problem leads some authors to dismiss any issue regarding the protection of neural data, for the marginal cases where they are not already classified as special categories, by proposing the inclusion of a specific class of neurodata in art. 9 of the GDPR[69]. Such an intervention would likely be useful in extending the enhanced pro-

[63] "Neural data" undoubtedly fall into the category of sensitive data and must be treated according to the regulations of the new GDPR on personal data protection, according to R. Trezza, *La tutela della persona umana nell'era dell'intelligenza artificiale: rilievi critici*, in *Federalismi.it*, 2022, 300). See also P. Perlingieri, *Note sul "potenziamento cognitivo"*, in *Tecnologie e diritto*, 1, 2021, 214 ss.

[64] N. Minielly - V. Hrincu - J. Illes, *A View on Incidental Findings and Adverse Events Associated with Neurowearables in the Consumer Marketplace*, in I. Bárd - E. Hildt (eds.), *Developments in Neuroethics and Bioethics. Ethical Dimensions of Commercial and DIY Neurotechnologies*, Cambridge, 2020, 267 ss.

[65] J. Zhang - J. Ho Yun - E.-J. Lee, *Brain Buzz for Facebook? Neural Indicators of SNS Content Engagement*, in *Journal of Business Research*, 130, 2020.

[66] T. Terranova, *Attention, Economy and the Brain*, in *Culture Machine*, 13, 2012, 1 ss.

[67] D. Hallinan - P. Schütz - M. Friedewald - P. de Hert, *Neurodata and Neuroprivacy: Data Protection Outdated?*, in *Surveillance & Society*, 1, 2014, 55 ss.

[68] Art. 7: «*La recopilación, almacenamiento, tratamiento y difusión de los datos neuronales y la actividad neuronal de las personas se ajustará a las disposiciones contenidas en la ley n°19.451 sobre trasplante y donación de órganos, en cuanto le sea aplicable, y las disposiciones del código sanitario respectivas*». A nonsense according to C. Bublitz, *Novel Neurorights: From Nonsense to Substance*, cit., 7.

[69] «Insofar as some forms of neurodata are not covered but should be so, one may insert "neurodata" to art. 9, next to other types of data such as genetic data. No need for further reforms» (C. Bublitz, *ibid.*, 7).

tection regime to the grey area of neural data not classified as sensitive, but it would certainly not be entirely conclusive. First, because the threat it aims to address appears rather «invisible»[70]; second, because the regulation of the digital environment contends with its a-territorial nature (the lack of territorial boundaries)[71]; third, because classifying neural data as special categories does not circumvent the critical issues that still arise in this sector[72]; and lastly, because consent (even if free, specific, informed, explicit, given through an unequivocal positive act, etc.) would still tend to legitimize risky processing.

The perspective of data protection seems the most suitable for addressing the issue of technological interference in cognitive activity, especially because it extends its scope of interest to non-interventional technologies, which could hardly be understood from the perspective of bodily intervention and biolaw. Nonetheless, the debate has mainly focused on defining an autonomous category of neural data. This result could add a set of personal data that is difficult to define to the already crowded set of special categories. Even this result, in any case, would not allow overcoming the problems and limits found in the field of protecting data belonging to special categories, especially with reference to uncontrolled uses in non-healthcare contexts.

6. Theoretical Criticisms and Dogmatic Questions

The proposal by Yuste that substantiates the Neurorights Initiative identifies five neuro-rights[73]: mental privacy, personal identity, free will, fair access to mental enhancement, and protection from algorithmic bias. According to the proposal by Ienca and Andorno, four neuro-rights could be identified: cognitive liberty, mental privacy, mental integrity, and psychological continuity[74].

Analyzing these proposals, one can distinguish between several main classes of rights to be examined: the first class concerns integrity (psychic, psychological, of the mind, etc.); the second concerns privacy (of the mind, brain, neurons, etc.); the third concerns liberty (of the mind, psychic, cognitive, free will, autonomy of choice, etc.); the fourth concerns identity and continuity (psychic, psychological, etc.); and the fifth concerns access to enhancement (neural, cognitive, psychic, etc.).

The issue of integrity (whether psychic, psychological, or mental) is relatively straightforward, primarily because art. 3 of the Charter of Nice explicitly affirms that every-

[70] P. De Pasquale, *Verso una Carta dei diritti digitali (fondamentali) dell'Unione europea?*, in *Osservatorio europeo*, March 2022, 14 ss.

[71] About the «a-territorial nature of the Internet», G. De Minico, *Towards an Internet Bill of Rights*, in *Federalismi.it*, 5, 2016, 14 ss. See also F. Pizzetti, *Il sistema cinese di tutela e sicurezza dei dati e il quadro europeo nello scenario della competizione mondiale*, in *Federalismi.it*, 4, 2022.

[72] E. Catelani, *Nuove tecnologie e tutela del diritto della salute: potenzialità e limiti dell'uso della Blockchain*, in *Federalismi.it*, 4 2022, 214, ss.; or A. Thiene - S. Corso (eds.), *La protezione dei dati sanitari.* Privacy *e innovazione tecnologica tra salute pubblica e diritto alla riservatezza*, Napoli, 2023.

[73] See *neurorightsfoundation.org/mission*.

[74] M. Ienca - R. Andorno, *Towards New Human Rights in the Age of Neuroscience and Neurotechnology*, cit., *passim*.

one has the right to physical and mental integrity. Additionally, constitutional jurisprudence recognizes the "physical or mental integrity of individuals" as fundamental to the very existence of the legal system[75]. This right, which is relevant in discussions on potential neuro-rights, is not new. It is already acknowledged as psychic integrity and extensively protected by various safeguards. For instance, compensatory protection for psychic damage is treated as biological damage[76]; moreover, the constitutional prohibition of torture, enshrined in art. 613-bis of the Criminal Code, includes provisions against psychic torture, penalizing those who cause acute physical suffering or verifiable psychic trauma through severe violence, threats, or cruelty.[77]The specific nature of this integrity, whether termed mental, psychic, or psychological, does not necessitate a debate between dualism and biological reductionism. Violations of psychic integrity are already regarded as breaches of a unified, indistinct entity, which can be evaluated using different criteria, such as biological damage verified through psychiatric consultation.

Regarding privacy, the Neurorights Foundation asserts that «any NeuroData obtained from measuring neural activity should be kept private. If stored, there should be a right to have it deleted at the subject's request»; or, according to the other authors, «it should guarantee the systemic protection of brain information»[78]. Overall, the primary risk highlighted in the literature is the extraction of information and the unregulated processing of data flows related to individuals. The main concern, therefore, is linked to privacy, encompassing information related to the nervous system, cognitive processes, emotions, and more generally, the mind and thoughts of the person. The proposals primarily focus on data protection, for which it is already possible to outline a framework of claims and powers based on positive law[79]. In this case, it does not seem appropriate to discuss a new neuroright, although it is likely that the current regulation does not fully guarantee the protection of private life (arts. 8 ECHR and 7 CFREU) and the instrumental right to data protection (art. 8 CFREU) against neurotechnologies. The most problematic aspect is the possible definition of an autonomous category of (neuro)data to be subjected to an enhanced protection regime.

---

[75] Constitutional Court, decisions nos. 290/2001, 236/2020.

[76] Italian Supreme Court, civil división, III, 11 June 2009, no. 13547.

[77] Art. 13, para. 3, of the Italian Constitution states that «all physical and moral violence against persons subject to restrictions of freedom shall be punished», and art. 27, par. 3, provides that «punishments cannot consist of treatments contrary to the sense of humanity». The issue here would be more about effectiveness than recognition: C. Scialla, *L'inafferrabile reato di tortura nello spazio della detenzione*, in *BioLaw Journal*, 4, 2022, 113 ss.

[78] Respectively on the NeuroRights Foundation website and in M. Ienca, *On Neurorights*, cit., 7.

[79] As outlined in arts. 15-22 of the GDPR, these rights enable a broad range of claims and powers. These include the right to obtain information about which data is being processed (right to information); the right to request and receive data in an intelligible form (right of access); the right to obtain the update or correction of submitted data (right to rectification); the right to have data deleted (right to erasure); the right to oppose data processing (right to object); the right to revoke consent for data processing (right to withdraw consent); the right to oppose automated processing and not be subject to decisions based solely on automated processing, including profiling (right to object to automated processing); the right to obtain the blocking or limitation of data processed unlawfully and those no longer necessary for the processing purposes (right to restriction of processing); and the right to transfer data to another controller (right to data portability).

Certainly, here as elsewhere, the emphasis on the neural level has the drawback of overlooking the broader framework of data processing types, not all of which are solely focused on neural activity in the strict sense.
Nonetheless, the hypothesis of creating distinct categories of data for various reference levels (mental, neural, psychic, psychological, cognitive, etc.) is unconvincing. The implicit naive reductionism in the notion of neural data (neuronal or neurodata) seems preferable to a slippery multiplication of categories, which would draw any type of mental information related to the person into an enhanced protection regime (any taste, memory, or evaluation «regarding an identified or identifiable natural person», per art. 4 GDPR). It should be noted that the risk to fundamental rights is not so much determined by the qualification of the data, but by the type of processing and the context (as per art. 25 GDPR). Therefore, regulation should focus on neurotechnologies rather than on neurodata themselves.
Proponents of cognitive liberty emphasize the necessity of safeguarding an individual's freedom to control their mental states[80]. This conception of freedom appears to resonate more closely with Eastern spiritual practices, which are protected under art. 19 of the Constitution. The notion of cognitive liberty, however, is somewhat ambiguous, as it presupposes that the individual exists independently of their cognitive processes, mental states, or consciousness, and can thus assert control over them. More pragmatically, the advocacy for cognitive liberty seems to aim at preventing manipulative interference by covert or unwanted external agents. While the focus of mental privacy is on protecting "outgoing" information, cognitive liberty seeks to prevent the intrusion of information or stimuli that could subtly influence cognitive functions, regardless of whether such interference results in perceptible or permanent harm to psychic integrity. This concern is clearly addressed by the right to private life (arts. 8 ECHR and 7 CFREU) and is further supported by the constitutional protections of bodily integrity (arts. 2, 13, and 32 of the Constitution). It is, therefore, difficult to envision a legal framework that upholds the inviolability of personal freedom, domicile, and correspondence, yet remains indifferent to covert neuromodulation practices.
Cognitive liberty thus emerges as a demand for non-interference through neuromodulation technologies, potentially accompanied by informational rights regarding the risks associated with these technologies, as well as secondary protective measures. Consequently, the establishment of an autonomous right to free will or control over mental states appears superfluous. More beneficial would be enhanced regulatory measures and the widespread implementation of informational obligations concerning the inductive impact of these practices.
Turning to identity and the consequent psychological (or psychic) continuity, the proposals here are based on two necessary assumptions: that individuals possess an identity (i.e., a specific and stable essence) and that this identity is maintained over time through significant continuity of psychological aspects. The concept of psychological continuity is primarily developed through philosophical rather than psychological re-

[80] *I.e.*, «the positive liberty of having the possibility of acting in such a way as to take control of one's mental life» e «freedom of thought as the normative foundation of a person's autonomous control over her mind» (M. Ienca, *On Neurorights*, cit., 6-7); «Individuals should have ultimate control over their own decision making» for the Neurorights Initiave.

flection on the theme of identity[81]. Consider neurostimulation techniques that induce changes in musical preferences, as in the case of a sixty-year-old individual with obsessive-compulsive disorder treated with deep brain stimulation, who consequently developed an unexpected passion for Johnny Cash as a side effect[82]. It is thus assumed that neurotechnologies can modify—potentially deliberately and not only in relation to musical tastes—certain distinctive traits of personal identity[83].

There is a recognized need to protect personal identity and the continuity of psychological life from external alterations[84] and to prohibit technologies from disrupting one's sense of self or blurring the boundary between self-awareness and external technological inputs[85]. While it is debatable whether such an essentialist concept of identity can constitute a fundamental right, the preservation of personal identity against unwanted external forces is a complex issue[86]. It could be debated whether such a structural and essentialist concept of identity qualifies as a value and, if so, whether it is sufficiently shared to justify the establishment of an autonomous fundamental right. Similarly, one might question, in abstract terms, whether personal identity can ever be entirely preserved over time from the influence of unwanted external forces.

Nevertheless, the right to personal identity is already well-established in European law (ECtHR jurisprudence on art. 8) and in national contexts. Initially recognized as the right to a correct social projection, it has evolved to include the right to a name, control over personal information, and one's biological truth, ultimately becoming the «right to be oneself»[87] even allowing for significant changes in one's individual history (e.g., the right to be forgotten and the right to alter sexual characteristics). External interference in this domain is already broadly prohibited by numerous legal provisions.

---

[81] P. van Inwagen, *Materialism and the Psychological-Continuity Account of Personal Identity Source*, in *Philosophical Perspectives*, 11, 1997, 305 ss.

[82] M. Ienca, *Neurodiritti: storia di un concetto e scenari futuri*, in *Privacy e neurodiritti. La persona al tempo delle neuroscienze*, cit., 50; M. Mantione - M. Figee - D. Denys, *A Case of Musical Preference for Johnny Cash Following Deep Brain Stimulation of the Nucleus Accumbens*, in *Frontiers in Behavioral Neuroscience*, 8, 2014, 1 ss. Research «may suggest an association between DBS and changed musical preference». The improbability of such a late change in musical tastes is particularly noteworthy, especially considering the singularity of the individual having no pronounced musical preferences prior to the clinical treatment.

[83] On the subject of the relationship between neurotechnologies and music, a number of studies could be cited to the contrary, demonstrating the use of musical stimulation integrated with neuromodulation techniques: «[l]istening to modulated vocalizations/music is potentially an efficient strategy for neuromodulation of the autonomic nervous system» (N. Rajabalee *et al.*, *Neuromodulation Using Computer-Altered Music to Treat a Ten-Year-Old Child Unresponsive to Standard Interventions for Functional Neurological Disorder*, in *Harvard Review of Psychiatry*, 5, 2022, 311).

[84] M. Ienca - R. Andorno, *A New Category of Human Rights: Neurorights*, in *Research in Progress Blog*, 26 April 2017. Art. 4 of the Chilean proposal also addresses psychological continuity: «*si puede dañar la continuidad psicológica y psíquica de la persona*».

[85] According to NeuroRights Foundations: «Boundaries must be developed to prohibit technology from disrupting the sense of self. When neurotechnology connects individuals with digital networks, it could blur the line between a person's consciousness and external technological inputs».

[86] See, for instance, the criticism of F. Remotti, *Contro l'identità*, Bari-Roma, 2001 or Id., *L'ossessione identitaria*, Bari-Roma, 2010.

[87] G. Pino, *L'identità personale*, in S. Rodotà - M. Tallacchini (eds.), *Ambito e fonti del biodiritto*, in *Trattato di biodiritto*, Milano, 2010, 301 ss. Or see the comprehensive framework of issues in E. Lecaldano, *Identità personale. Storia e critica di un'idea*, Torino, 2021.

For cases not covered by the protection of psychic integrity and cognitive liberty, interference in the sphere of personal identity and psychological continuity would still constitute an intrusion into the individual's private domain, with all associated rights, freedoms, and powers, as previously discussed for other classes of rights.

Lastly, equitable access to cognitive enhancement raises concerns about creating a "two-speed" humanity: one group enhanced due to economic and cultural access to neurotechnologies, and another excluded[88]. This scenario could seem particularly dystopian if we do not consider the existing cultural and economic inequalities that already result in highly unequal health distributions both globally and within individual countries. There is a clear and direct link between income and health, known as the social gradient, which is evident not only in developing countries but also in the wealthiest nations.

Therefore, this expectation, primarily highlighted by the NeuroRights Foundation, can be seen as part of the broader category of the (social) right to healthcare. The claim in question should be understood as the right (to equitable access) to a healthcare service (provided or funded by the public sector, depending on jurisdiction), and pertains to treatments that, if not aimed at addressing pathological conditions but rather at improving psycho-physical wellbeing, border on forms of human (neuro) enhancement. This is hardly a novel concept, or at least not new in the specifically "neural" context[89].

Similar considerations arise concerning the protection from bias, which underscores the necessity of countermeasures to combat bias and input from user groups to foundationally address bias in the context of neurotechnologies. Thus, this would not constitute an autonomous class of rights but rather a general expectation towards the quality of technology, protecting various interests (sometimes adhering to the principle of *neminem laedere*, sometimes aimed at combating social inequalities, etc.).

A comprehensive analysis suggests that all new neurorights pertain to two distinct yet partially connected phenomena: informational extraction and manipulative interference with the individual. The need for protection from both forms of access, "outgoing" and "incoming"[90], could be conveniently subsumed—at the cost of overcoming some methodological caution—under a single provisional nomenclature: neuroprivacy.

On one hand, it has been observed that the concept of privacy acts like a veritable "black hole" that engulfs virtually all individual rights, configuring itself as a kind of «general right to self-determination»[91] where privacy, in the strict sense, is just one of

[88] «There should be established guidelines at both international and national levels regulating the use of mental enhancement neurotechnologies. These guidelines should be based on the principle of justice and guarantee equality of access» (NeuroRights Foundation).

[89] On human enhancement, N. Bostrom*, Intensive Seminar on Transhumanism*, New Haven, 2003; P. Benanti*, Postumano, troppo umano. Neurotecnologie e* human enhancement, Roma, 2017, 20 ss.

[90] «Questo tipo di doppia proiezione, in entrata e in uscita, è uno dei pericoli in questo momento più rilevanti», according to O. Pollicino, *Costituzionalismo, privacy e neurodiritti,* cit., 16.

[91] M. Luciani, *Il diritto al rispetto della vita privata: le sfide digitali, una prospettiva di diritto comparato*, Studi del Servizio Ricerca del Parlamento europeo, Bruxelles, 2018, 1. See also A. Cerri, *Riservatezza (Diritto alla), Diritto costituzionale*, in *Enciclopedia giuridica*, XXVII, Roma, 1991. A similar statement yet in W.L.

its specifications, while data protection is considered an autonomous and instrumental right. Undoubtedly, merging different issues into this macro-category has produced broad and fertile reflection, especially true in this initial phase of research[92]. In this sense, leveraging the potential ambiguity and broad semantic scope of the concept of privacy would be beneficial for our purposes.

On the other hand, when identifying a reference level for protecting an individual's intimacy (among neurons, neural processes, cognitive processes, mind, psyche, etc.)[93], one might favour the generic (and perhaps overused) reference to the neural level, in line with the neurohype that characterises many of the various involved disciplines. This "neuro-essentialist" or reductionist option would allow for a "methodological reduction" to a single principle because it places all forms of interference on a secular, bodily level (without the need to invoke all possible levels of reference).

7. Conclusions

The emergence of new rights in response to novel forms of technological interference in cognitive activities occupies a vast and complex domain. These interferences—whether incoming or outgoing—into the psychic sphere encompass neurosurgery, brain imaging, certain pharmacological treatments and narcotics, psychometrics, facial recognition, and the measurement of seemingly non-psychic parameters (such as heart rate, skin conductance, and electrodermal activity), as well as vocal analysis, among others. This spectrum includes various techniques and practices that access the body's dimension, ranging from less to more interventionist, involving heterogeneous data types and diverse instruments with varying levels of prevalence. The array of technologies not only spans a wide scope but also serves a multitude of purposes, from therapeutic to enhancement, and from commercial applications to social control. For example, a particularly relevant aspect of art. 5 of Regulation (EU) 2024/1689 (AI Act) is the prohibition of using AI systems that deploy «subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques with the objective, or the effect, of materially distorting the behavior of an individual or a group of individuals». This prohibition gains particular significance when applied to emotion recognition systems and the protection of vulnerable persons. The provision highlights the risk that technologies designed to detect and exploit emotional states might be used to manipulate intentions and behaviors, exacerbating existing vulnerabilities or creating new ones. This scenario exemplifies the ineffability of suitable legal

Prosser, *Privacy*, cit., 422: «It is evident from the foregoing that, by the use of a single word supplied by Warren and Brandeis, the courts have created an independent basis of liability, which is a complex of four distinct and only loosely related torts; and that this has been expanded by slow degrees to invade, overlap, and encroach upon a number of other fields».

[92] *Ibidem*, 423: «This is not to say that the developments in the law of privacy are wrong. Undoubtedly, they have been supported by genuine public demand and lively public feeling and made necessary by real abuses on the part of defendants who have brought it all upon themselves».

[93] Consequently, we might derive terms such as brain privacy, privacy of mind, mental privacy, neural privacy, psychoprivacy, cognitive privacy, etc.

categories to address these phenomena, particularly in determining the boundary between manipulative techniques and legitimate influences on individual decision-making.

For this reason, in the discussion, it was illustrated how different facets of the same issues can (and should) be examined from multiple perspectives. Within a broader framework that integrates theoretical and doctrinal reflection with the context of positive law, it is possible to assess proposals concerning new neurorights. This involves either aligning them with already recognised rights or highlighting the critical issues posed by their innovative elements, especially the paradoxes they may present.

Lastly, the concept of neuroprivacy, though ambiguous, has proven useful—a notion that, while dogmatically unsatisfactory, serves as a provisional guide for future research. Some of the emphasis placed in the literature on the dystopian image of neurotechnologies likely stems from excessive caution, or even an irrational fear, toward certain emerging technologies. Nonetheless, while questioning the need for new regulations in this sector and an array of new human rights—given that the *corpus juris digitalis* is already overly dense—we recognise that the challenge posed to the law by neuroscience is indeed central. Despite the ineffable confrontation between these two realms and the numerous contradictions arising from their intersection, this remains a critical area. The hope is that this vital space, still largely uncharted today, will not be filled solely with regulations, but also with bridges that connect law and cognitive sciences.

Modelli di regolazione (e supervisione) per l'AI finanziaria: neutralità tecnologica, etica e tutela dell'investitore*

Daniel Foà

Abstract

Nell'ambito del settore finanziario, l'intelligenza artificiale può essere impiegata per varie finalità, molte delle quali caratterizzate da grandi potenzialità; al contempo, l'utilizzo di tali strumenti pone rilevanti sfide. Il contributo intende, in primo luogo, indagare se la regolazione e supervisione finanziaria - nel fronteggiare le insidie poste dalle applicazioni dell'intelligenza artificiale - siano tenute a conformarsi al principio di neutralità tecnologica. In secondo luogo, alla luce del regolamento (UE) 2024/1689 (AI Act) e all'evoluzione dei servizi, viene valutato quale sia l'approccio ottimale per garantire una piena tutela dell'investitore nei confronti dei *roboavisors*, anche nell'ambito di ambienti immersivi caratterizzati da interazioni *phygital*. È, infine, analizzato il caso d'uso dei *virtual worlds* per dimostrare l'esigenza di regolare e supervisionare le applicazioni dell'AI in ambito finanziario adottando un approccio *tech specific*.

Within the financial sector, artificial intelligence may be deployed for various purposes, many of which offer a great extent of capabilities yet posing significant challenges. Firstly, the contribution intends to investigate whether financial regulation (and supervision), when coping with the challenges posed by artificial intelligence applications, are bound to comply with the principle of technological neutrality. Secondly, in the light of the EU Regulation 2024/1689 (AI Act) and the developments in financial services, it is assessed which is the most appropriate approach to grant investor protection vis-à-vis roboavisors, also in the context of immersive environments, charachterised by phygital interactions. Thus, the use case of virtual worlds is analysed to show the urgency of regulating and supervising AI financial applications through a tech-specific approach.

Sommario

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

sors, etica e tutela dell'investitore – 7. *Virtual worlds*: "luogo" in cui l'esigenza di tutela diviene (ancora più) sentita – 8. Conclusioni

Keywords

Intelligenza artificiale – *roboadvisor* – neutralità tecnologica – supervisione – regolazione

1. Introduzione.

1.1 L'intelligenza artificiale[1] ha numerose applicazioni e pone una serie di questioni per il regolatore: anzitutto occorre definirla, comprenderne le insidie, decidere se limitarne l'impiego e/o assoggettare a un penetrante scrutinio coloro che la utilizzino per prestare servizi.

Si tratta di sfide di cruciale rilevanza, rapidamente divenute prioritarie per i legislatori dei paesi più avanzati[2]. In primo luogo, perché l'applicazione di tali tecnologie (generaliste) è potenzialmente orizzontale, potendo essere integrate in qualsiasi settore economico o catena del valore.

Le modalità di regolazione dell'AI passano anzitutto dalla comprensione delle caratteristiche della tecnologia, della sua attitudine trasformativa dei processi, e dei limiti nell'applicazione delle regole già presenti nell'ordinamento. Quando, poi, le tecnologie di intelligenza artificiale siano impiegate nell'ambito del settore finanziario (da parte di istituzioni finanziarie o comunque nell'ambito della prestazione di servizi finanziari)[3], si pongono ulteriori profili di complessità, in quanto occorre garantire la tutela degli investitori e la stabilità del sistema[4]. Difatti, l'uso scorretto – o anche solo incontrollato – di tali strumenti tecnologici potrebbe risultare pregiudizievole sia per gli interessi dei singoli (andando ad incidere anche su diritti costituzionalmente tutelati, quali ad es. quello alla piena e libera autodeterminazione) sia del sistema finanziario complessivamente inteso, potendo perfino porre a repentaglio la fiducia nei mercati finanziari

[1] Per una introduzione ai profili definitori (e giuridici) dell'intelligenza artificiale si veda G. Finocchiaro, *Intelligenza artificiale. Quali regole?*, Bologna, 2023. Più ampiamente sulle questioni definitorie, S. Samoili et al, *AI Watch. Defining Artificial Intelligence Towards an operational definition and taxonomy of artificial intelligence,* JRC Technical Reports, EUR 30117 EN, 2020.

[2] Si pensi, tra le altre, alle iniziative intraprese dall'Unione Europea, dagli Stati Uniti, dalla Cina. Pur adottando approcci molto differenti, sono tutte accomunate dalla percepita esigenza di definire regole (e limiti) per l'utilizzo di strumenti tecnologici dotati caratterizzati da rilevanti potenzialità. Cfr. Financial Times, *The global race to set the rules for AI*, di C. Criddle-J. Espinoza-Q. Liu, 13 settembre 2023.

[3] Cfr. F. Capriglione, *Diritto ed economia. La sfida dell'Intelligenza Artificiale*, in *Rivista Trimestrale Diritto dell'Economia*, 3, 2021, 24 ss.; G. Schneider, *La proposta di regolamento europeo sull'intelligenza artificiale alla prova dei mercati finanziari: limiti e prospettive (di vigilanza)*, in *Responsabilità Civile e Previdenza*, 3, 2023, 1014 ss.

[4] Per una panoramica delle principali applicazioni dell'intelligenza artificiale nel settore bancario e finanziario (e delle sfide che pongono) si veda J.C. Crisanto - C. Benson Leuterio – J. Prenio – J. Yong., *Regulating AI in the financial sector: recent developments and main challenges*, in FSI Insights on policy implementation No. 63, dicembre 2024, spec. 4 ss. In particolare, sul rischio sistemico A. Keller-C. Martins Pereira-M.L. Pires, *The European Union's Approach to Artificial Intelligence and the Challenge of Financial Systemic Risk*, in H. Sousa Antunes et al. (a cura di), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, vol, 58, Cham, 2024, 415 ss.

e negli intermediari che vi operano.
Pertanto, in questo settore, il legislatore e le autorità di vigilanza sono chiamati a fronteggiare i rischi, operativi e finanziari, che l'AI pone avvalendosi in primo luogo dello strumentario della regolazione finanziaria[5]. Ciò però non esaurisce i meccanismi d'intervento, anche in considerazione della natura non intrinsecamente prudenziale dei controlli sull'AI[6] e alla conseguente non riconducibilità alle autorità di vigilanza bancaria e finanziaria di ogni controllo sull'impiego di tali tecnologie da parte delle istituzioni finanziarie.

1.2 Il contributo intende quindi analizzare quale sia l'approccio regolatorio più adeguato a fronteggiare le sfide poste dalle applicazioni di intelligenza artificiale nel settore finanziario, valutando da un lato fino a che punto il principio di neutralità tecnologica limiti il legislatore (e le autorità di vigilanza) nelle proprie scelte e, dall'altro, come ciò impatti sui livelli di tutela degli investitori. A tale scopo, sono anzitutto esaminate le nozioni di intelligenza artificiale impiegate nel regolamento (UE) 2024/1689 (AI Act) e in altri testi ufficiali al fine di valutarne la neutralità (da un punto di vista tecnologico ed etico). Successivamente, ci si sofferma sulle origini e la portata concreta del principio di neutralità tecnologica, prendendo in considerazione i modelli di regolazione e supervisione adottati a livello europeo, con particolare riferimento al settore finanziario. In tale contesto, viene quindi dimostrato a quali condizioni sia possibile derogare a tale principio. Ci si interroga infine su quale sia l'approccio ottimale per garantire una piena tutela dell'investitore nei confronti dei *roboavisors*, anche nell'ambito di ambienti immersivi caratterizzati da interazioni *phygital*. Proprio quest'ultimo caso d'uso dimostra l'esigenza di regolare e supervisionare le applicazioni di AI finanziaria adottando un approccio *tech specific*.

2. Una preliminare questione definitoria

2.1 Intelligenza artificiale è una nozione commerciale, che non corrisponde in modo univoco a una specifica tecnologia[7]. Difatti, nell'accezione comunemente accolta vi rientrano diverse tecnologie (e relative applicazioni) basate sugli algoritmi di *supervised* e *unsupervised machine learning* (ML), di *natural language process* (NLP), *large language models* (LLM), *expert systems*, nonché i modelli di intelligenza artificiale generale (AGI)[8]. E tale

[5] Armour et al., *The Goals and Strategies of Financial Regulation*, in Principles of Financial Regulation, Oxford, 2016, 51 ss.; A. Sciarrone Alibrandi, *Innovazione tecnologica e regolazione dei mercati*, in R. Lener-G. Luchena-C. Robustella (a cura di), *Mercati regolati e nuove filiere di valore*, Torino, 2021, 5 ss.

[6] R. Lener, *Vigilanza prudenziale e intelligenza artificiale*, in *Rivista Trimestrale di Diritto dell'Economia*, 1, 2022, 214.

[7] Si ritiene che sia stata introdotta per la prima volta nello scritto, J. McCarthy-M. L. Minsky-N. Rochester-C.E. Shannon, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, agosto 1955, in *AI Magazine*, *27*(4) 2006, 12.

[8] Più ampiamente sulle nozioni di algoritmo e la loro riconducibilità al concetto di intelligenza artificiale G.F. Italiano, *Le sfide interdisciplinari dell'intelligenza artificiale*, in Analisi giuridica dell'economia, 2019, 9 ss.; N. Cristianini, *Machina sapiens*, Bologna, 2024. Con specifico riferimento all'applicazione di tali tecnologie nell'ambito finanziario, si veda The Alan Turing Institute, *The AI Revolution: Opportunities*

elencazione, che pur si pone ad un livello di dettaglio certamente superficiale, è tutt'altro che esaustiva.

Anche per questo motivo, la Commissione UE nel 2018 aveva proposto una definizione funzionale di intelligenza artificiale che ne evidenziava la caratteristica di «*sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi*»[9]. Nell'ambito della versione finale dell'AI Act[10], che pur aderisce alla medesima impostazione, si è assistito ad una progressiva evoluzione della definizione, che tenta di delineare una nozione ampia e a basso tasso di tecnicismo, che ne permetta una lunga durata, limitando il rischio di rapida obsolescenza[11]: far riferimento all'intelligenza artificiale non significa quindi necessariamente far riferimento ad una (e una sola) tecnologia, ma ad una categoria di tecnologie che hanno alcuni elementi in comune[12].

In altri termini, l'ampia definizione di intelligenza artificiale contenuta nell'AI Act – a cui consegue una marcata disomogeneità tra le tecnologie ricomprese nel suo ambito applicativo – potrebbe suggerire che non si tratti, a dispetto della rubrica, della regolamentazione di una tecnologia[13]. Difatti, il regolamento non disciplina un'attività – quantomeno non in modo diretto – bensì un modo di svolgere un'attività: non un'attività, non un prodotto, ma un processo. E le attività svolte avvalendosi di tali strumenti sono soggette, in base alla valutazione circa il livello di rischio che pongono[14], a una disciplina dedicata (e ulteriore) rispetto alle medesime attività prestate con

and Challenges for the Finance Sector, 2023, 9 ss.; ESMA, *Artificial intelligence in EU securities markets*, 1 febbraio 2023, ESMA50-164-6247.

[9] Commissione UE, *L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, 25 aprile 2018. A tale definizione aderisce anche il *White Paper sull'intelligenza artificiale, Un approccio europeo all'eccellenza e alla fiducia* COM(2020) 65 final, che dà anche conto della definizione – più tecnica – elaborata dal Gruppo di esperti di alto livello, Orientamenti etici per un'IA affidabile, 2019, 45.

[10] Regolamento (UE) 2024/1689 del Parlamento Europeo e Del Consiglio del 13 giugno 2024.

[11] Al riguardo, l'art. 3 (1) dell'AI Act definisce il sistema di intelligenza artificiale come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

[12] Elementi caratterizzanti – non necessariamente omogenei tra le varie categorie di intelligenze artificiali – sono poi quelli relativi alla (mancanza) di esplicabilità del funzionamento, al livello di trasparenza e vulnerabilità. Quest'ultima, in particolare, può essere intesa in varie accezioni: sia in relazione al corretto funzionamento degli algoritmi e alla loro modificabilità sia al livello di sfruttamento delle debolezze degli utenti che sono rese possibili dalle applicazioni di intelligenza artificiale.

[13] Per il vero, gli artt. 5 e 6 dell'AI Act nonché gli allegati al regolamento (In particolare, l'Allegato III) elencano con un certo livello di dettaglio alcune tecnologie soggette alla disciplina del regolamento stesso, e quindi espressamente regolate. Va precisato, però, che le tecnologie individuate nell'Allegato sono indicate facendo riferimento al loro utilizzo e non invece alle caratteristiche tecniche, che consentirebbe un riferimento più univoco.

[14] In base a tale classificazione impone regole diverse e persino divieti, sul punto si veda G. Finocchiaro, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in Il Diritto dell'informazione e dell'informatica, 2, 2022, 303 ss. La specifica attenzione dedicata dai regolatori alle applicazioni di AI può ricondursi anzitutto alla mutevolezza e non staticità – caratteristica di quei sistemi basati su forme di autoapprendimento – che rendono complicato per un soggetto esterno valutare ed approvare definitivamente una determinata tecnologia, in quanto richiederebbe una valutazione circa la possibile evoluzione futura dello strumento. In secondo luogo, poiché casi d'uso diversi pongono sfide

diverse modalità.
Un approccio regolatorio di questo tipo risulterebbe però non tecnologicamente neutrale. Quel che occorre valutare è (i) se ciò sia legittimo e (ii) se sia giustificato.

3. Neutralità tra tecnologia ed etica

3.1 Parlando di neutralità tecnologica nel contesto del *rule-making*, ci si riferisce ad un principio emerso originariamente nel settore delle comunicazioni elettroniche[15], espressione del generale dovere di non discriminazione. Tale principio – che impone di applicare le medesime regole ad attività sostanzialmente analoghe, anche quando siano svolte avvalendosi di tecnologie diverse – ha poi vissuto una *vis expansiva*, trovando applicazione anche in altri settori, ugualmente interessati dall'innovazione tecnologica e dalle conseguenti trasformazioni.
Ebbene, l'estensione dell'ambito applicativo del principio di neutralità tecnologica è probabilmente da ricondurre alla convinzione comune che la tecnologia sia un semplice strumento: se la tecnologia è un mero mezzo, che non attribuisce una connotazione specifica e diversa all'attività che viene così svolta, non vi è ragione per assoggettarla a regole diversificate. In realtà, le tecnologie (quantomeno, alcune di esse) concorrono a definire i connotati delle attività che vengono svolte avvalendosene. Costituiscono sovente un facilitatore di attività che altrimenti non potrebbero essere svolte o che risulterebbero più complesse. Se alcune tecnologie sono di per sé caratterizzate da elementi di positività o negatività perché il loro unico possibile uso è così connotato, per molte altre bisogna interrogarsi se siano neutre[16]. In questo senso, neutralità significherebbe indifferenza dello strumento rispetto ad un fine.
È assai raro però che una tecnologia sia effettivamente neutra, molto più spesso è "*double charged*": in grado di produrre effetti sia positivi sia negativi[17].
Quando osservate nella prospettiva statica anche le tecnologie "a doppia carica" potrebbero sembrare eticamente neutre, salvo poi disvelare le proprie caratteristiche nella prospettiva dinamica. Esattamente così avviene anche con riferimento all'intelligenza artificiale[18].

diverse: le stesse tecnologie possono comportare diversi livelli di allarme sociale e rischi (micro e macro) a seconda del settore in cui vengono impiegate.

15 U. Kamecke-T. Korber, *Technological Neutrality in the EC Regulatory Framework for Electronic Communications: A Good Principle Widely Misunderstood*, in *European Competition Law Review*, 2008, 330 ss.; W. Briglauer – V. Stocker – J. Whalley ., *Public Policy Targets in EU Broadband Markets: The Role of Technological Neutrality*, 29th European Regional ITS Conference, Trento, 2018, 184936.

16 S. Heyndels, *Technology and Neutrality*, in Philosophy & Technology, 2023, 36, 74 ss.

17 L. Floridi, *On Good and Evil, the Mistaken Idea That Technology Is Ever Neutral, and the Importance of the Double-Charge Thesis*, in *Philosophy & Technology*, 36, 60, 2023, 2.

18 Anche guardando specificamente al settore finanziario possono individuarsi esempi di "cariche" positive e meno positive (*rectius*, negative) dell'AI, nel senso di attitudine dello strumento tecnologico a migliorare o peggiorare gli esiti dell'attività umana. F. D'Acunto-N. Prabhala-A. Rossi, *The Promises and Pitfalls of Robo-advising*, 2018 CESifo Working Paper No 6907, 18. ss., studiando un ottimizzatore di portafoglio rivolto al mercato azionario indiano, rileva come il robo-advisor sia stato vantaggioso per gli investitori ex-ante poco diversificati, aumentando la diversificazione del loro portafoglio, riducendo

Come accennato, le applicazioni di intelligenza artificiale rappresentano sovente un efficace ausilio alle attività umane; in questo caso, si ritiene dunque che il "vettore buono" sia molto più forte di quello cattivo[19]. E pertanto, le possibili falle – tali da far emergere criticità e rischi che possono portare agli "*ethical disasters*" – sono da individuare non solo nelle fasi di progettazione e sviluppo, quanto piuttosto nell'impiego della specifica tecnologia[20]. Trattandosi di strumenti che, se correttamente progettati, rappresentano una *force for the good*, sui progettatori (e sui *deployers*) ricade una significativa responsabilità[21]. È quindi necessario fare in modo che non se ne faccia un cattivo uso. E anche questo è il compito della regolazione.
Pertanto, dalla constatazione della non-neutralità della tecnologia si può trarre un'argomentazione a favore della derogabilità del principio di neutralità tecnologica; a supporto di un approccio regolatorio *tech-specific*.

3.2. Nell'ambito del settore finanziario, il principio di neutralità tecnologica trova espressa previsione. Al riguardo, la comunicazione del Parlamento UE del 2017[22] in materia di FinTech affermava che la regolazione e supervisione, nell'ambito della *financial technology* si sarebbe dovuta basare sui seguenti principi «*a) same services and same risks; b) technology neutrality; c) a risk-based approach*». In tale contesto, il principio di neutralità tecnologica appariva essere una pietra angolare della disciplina europea del (allora ancora agli albori) *fintech*. Sembrava infatti che qualsiasi regolazione della tecnologia – anche quando strumento abilitante per prestare determinati servizi, caratterizzante la relativa fattispecie – fosse una discriminazione intollerabile perché avrebbe minato il *level playing field* tra gli operatori economici. Il presupposto, allora condiviso, è che la scelta dello strumento tecnologico sarebbe stata una mera opzione organizzativa, neutra dal punto di vista delle caratteristiche proprie del servizio prestato. A ciò sarebbe conseguito il vincolo per il legislatore di disciplinare in modo uniforme le attività, a prescindere dalle modalità operative concrete, salvo che queste dessero vita a rischi nuovi e disomogenei. Difatti, solo laddove sulla base dell'approccio basato sul rischio

il rischio e aumentando i loro rendimenti. Quanto, invece, agli investitori che già avevano un livello accettabile di diversificazione, il roboadvisor non ha migliorato le loro performances, anzi talvolta le ha peggiorate.

19 Floridi, *On Good and Evil, the Mistaken Idea That Technology Is Ever Neutral,* cit., 3.

20 L. Floridi, *The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities*, Oxford, 2023, passim.

21 P.P. Verbeek, *Morality in Design: Design Ethics and the Morality of Technological Artifacts*, in P.E. Vermaas et al. (a cura di), *Philosophy and Design,* Cham, 2008, 91 ss. le qualità del *deployer* risultano decisive nel caratterizzare i rischi del sistema di AI, in quanto dalla sua condotta potrebbe dipendere la bontà dell'uso del sistema stesso, nonché la prontezza d'intervento a fronte di possibili criticità. P. Benanti, *IA e medicina più dello strumento conta il manico*, in Avvenire.it, 21 marzo 2024, occupandosi in particolare delle applicazioni dell'AI all'ambito sanitario e dei relativi rischi ha riconosciuto come rilevi ben più "il manico" rispetto allo strumento in sé. Similmente E. McCaul, *Technology is neither good nor bad, but humans make it so*, Discorso del membro del supervisory board ECB nell'ambito della conferenza "*The use of artificial intelligence to fight financial crime*", organizzato da Intesa Sanpaolo, Torino 13 luglio 2022. Queste valutazioni paiono sottese anche alla disciplina di cui agli artt. 16 ss. dell'AI Act, i quali prevedono puntuali regole di condotta (e responsabilità) gravanti sul *deployer* del sistema.

22 Parlamento europeo, *FinTech: the influence of technology on the future of the financial sector*, Risoluzione del 17 maggio 2017 (2016/2243(INI)), (2018/C 307/06).

emergessero esigenze di tutela ulteriori, ciò giustificherebbe una deroga legittima al principio della parità di trattamento: le regole diverse e più stringenti deriverebbero dall'esistenza di rischi diversi, e ciò sarebbe del tutto giustificato nel contesto di una procedura di valutazione dell'impatto della regolazione.

Al di fuori di queste ipotesi, che ne rappresentano il confine esterno, la funzione del principio di neutralità tecnologica sarebbe quindi quello di garantire la non discriminazione ma anche la certezza per gli operatori economici, che potrebbero così contare sull'applicazione di regole omogenee per le proprie attività, a prescindere dalle tecnologie impiegate.

Nonostante continuino ad esserci riferimenti al principio di neutralità tecnologica in tutte le principali iniziative europee di disciplina di fenomeni caratterizzati dalla penetrante presenza tecnologica[23], tale principio pare ormai di fatto eroso. Basti pensare al GDPR (che richiama il principio al considerando 15) e al Regolamento MICA (considerando 9) che nonostante tali richiami nei considerando, contengono poi nell'articolato disposizioni che pongono regole dedicate per specifiche applicazioni tecnologiche. Anche nell'ambito della proposta della Commissione europea per l'introduzione di una disciplina dell'intelligenza artificiale si diceva espressamente che tale testo normativo «*aims to be as technology-neutral and future-proof as possible*»[24]. Ciononostante, tale proposito è stato disatteso nell'elaborazione concreta delle disposizioni normative (ad esempio, vietando alcune tipologie di sistemi di AI).

Sembra dunque evincersi - anche dallo svolgimento dell'*iter legis* dei provvedimenti citati - che il principio di neutralità tecnologica sia divenuto più un punto di partenza per il legislatore che un limite oltre il quale non sia lecito spingersi.

Occorre dunque domandarsi se e come un approccio non neutrale – sia regolatorio sia di vigilanza – possa meglio raggiungere gli obiettivi della regolazione finanziaria, consentendo di governare in modo più efficace e consapevole i rischi posti dalle variegate applicazioni dell'AI. In altri termini, si intende valutare quali caratteristiche debba avere la regolazione delle applicazioni finanziarie di AI (nonché i poteri d'intervento delle *supervisory authorities*) al fine di garantire una piena tutela degli obiettivi della regolazione finanziaria, nonché dei diritti fondamentali che potrebbero essere incisi.

Anche alla luce delle premesse sin qui esposte si intende dimostrare come sia sempre più indispensabile adottare un modello di regolazione e supervisione tecnologicamente orientata, che consenta di fronteggiare gli specifici rischi che un'applicazione tecnologica può porre, confrontandosi efficacemente con le evoluzioni della stessa, difficilmente governabili con strumenti neutrali.

---

[23] M.A. Scopellitti, *È ancora possibile la neutralità tecnologica della normativa?,* in V. Falce (a cura di), *Strategia dei dati e intelligenza artificiale*, Torino, 2023, 213 ss.

[24] Proposta di Bruxelles, Regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'unione, 21.4.2021, COM(2021) 206 final, 2021/0106 (COD), p. 12.

4. Regolazione neutrale

4.1 La regolazione tecnologicamente neutrale è quindi quella che ponga regole uniformi, guardando alle attività e ai loro rischi tipici, senza distinguere (e discriminare) in ragione degli strumenti impiegati per lo svolgimento delle stesse. Non preoccupandosi di quanto tali tecnologie possano mutare le modalità operative e abilitare interazioni nuove. La regolazione neutrale può impiegare tecniche normative per principi o *rules based*, così come può fare ampio rinvio a disposizioni attuative. In altri termini, anche rimanendo nel perimetro di una regolazione primaria formalmente neutrale, è possibile tenere in considerazione alcune delle specificità di determinate modalità operative. Ciò a cui sempre più spesso si assiste è quindi un approccio combinato: accanto a regole (tendenzialmente) tecnologicamente neutrali, vengono emanati regolamenti complementari su specifici ambiti di applicazione e soluzioni tecnologiche[25]. Un chiaro esempio in questo senso è la disciplina contenuta nella direttiva sui mercati degli strumenti finanziari (MIFID II) - nonostante contenga essa stessa disposizioni che non sono affatto in linea con la neutralità tecnologica (ad esempio l'art. 17 della MIFID II che regola il trading algoritmico) – che è poi integrata da puntuali regole "*tech-specific*". Si tratta comunque di un approccio non privo di punti deboli: come rilevato dal Presidente della CONSOB in un suo recente intervento, a fronte dell'emersione di tecnologie che caratterizzano i servizi prestati, se si vuole imporre al legislatore di mantenere un approccio tecnologicamente neutrale, deve poi riconoscersi un ruolo più invasivo alle autorità indipendenti, chiamate ad adottare regolamenti per fronteggiare i rischi non adeguatamente presidiati dalle fonti primarie[26]. È evidente, però, che siffatto modello non risulterebbe autenticamente rispettoso del principio di neutralità tecnologica.

4.2 Se il principio di neutralità tecnologica può ancora essere un obiettivo da perseguire in termini generali, questo è invece recessivo con riferimento ad applicazioni che danno luogo a rischi particolari. Tali considerazioni portano ad escludere che sia ancora possibile adottare un approccio pienamente neutrale nella regolazione delle applicazioni dell'AI[27].

Il principio di neutralità tecnologica, pur avendo solide basi negli atti europei – ed essendo un principio animato da finalità meritevoli – sembra ormai rappresentare un approccio idealistico, in grado di orientare il procedimento di formazione delle norme, ma con limitata portata cogente. La dirompente evoluzione tecnologica trasformativa ne ha fortemente ridimensionato la portata applicativa: la sua applicazione letterale rischierebbe di causare danno all'ordinamento, non garantendo sufficiente protezione dinnanzi a rischi derivanti dall'utilizzo di specifiche tecnologie.

---

[25] W. Buczynski et al., *Hard Law and Soft Law Regulations of Artificial Intelligence in Investment Management*, in *Cambridge Yearbook of European Legal Studies*, 24, 2022, 262 ss.

[26] P. Savona, Intervento del Presidente della Consob in occasione dell'assemblea annuale 2023 Assosim Competitività dei mercati italiani, Milano 29 maggio 2023.

[27] In senso contrario, P. Grady, *The AI Act Should Be Technology-Neutral*, Center for data innovation, febbraio 2023.

E tale conclusione appare ancora più evidente quando si calino tali strumenti nel settore finanziario. Difatti, si pone una questione di governabilità delle tecnologie in campo finanziario, non solo di imposizione di requisiti su uno strumento: il mercato finanziario tecnologizzato genera dinamiche diverse rispetto a quello tradizionale analogico, che devono essere previste, monitorate e gestite[28].

Laddove, dunque, sussistano queste esigenze risulta legittima una regolazione che preveda regole dedicate alla tecnologia e che distingua le regole sull'attività anche in base ai procedimenti impiegati per esercitare tali attività. Ciò purché tale non neutralità sia limitata al minimo indispensabile.

Gli approcci non neutrali, peraltro, non sono necessariamente pregiudizievoli per lo sviluppo delle nuove tecnologie e per chi le impieghi: non necessariamente sono introdotti limiti più stringenti alla prestazione delle attività, ma talvolta disposizioni specifiche e più favorevoli, a vantaggio dello sviluppo e del più ampio utilizzo di tecnologie di cui si voglia incentivare l'utilizzo e che siano ritenute in grado di mitigare i rischi di specifiche attività. Ad esempio, nell'ambito del DLT Pilot Regime (regolamento (UE) 2022/858) è previsto che all'utilizzo di infrastrutture di mercato decentralizzate consegua l'applicazione di un regime normativo semplificato rispetto a quello ordinariamente applicabile alle infrastrutture non decentralizzate[29]. Ebbene, la potenziale bidirezionalità della non-neutralità tecnologica della disciplina normativa concorre a riportare una siffatta scelta – quando oggettivamente giustificata – nell'alveo della discrezionalità del legislatore, senza violare principi inderogabili.

5. (segue) …e modelli di supervisione

5.1 Delineate le caratteristiche per una efficace regolazione dell'AI finanziaria, occorre ora soffermarsi sulla supervisione, complementare rispetto alla prima e parimenti fondamentale per un corretto funzionamento del mercato.

Nel contesto della supervisione bancaria e finanziaria, il principio di neutralità tecnologica assume confini più sfumati, operando solamente "di rimbalzo"; non limiterebbe le modalità d'azione, ma piuttosto rappresenterebbe un obbligo di risultato, corollario del principio *same risks, same rules, same supervision*.

Accogliendo una nozione ampia di neutralità tecnologica, questa può assumere almeno due significati in relazione alla supervisione.

In primo luogo, una supervisione neutrale è quella che effettui il medesimo scrutinio (e.g., in relazione agli atti esaminati, ai rischi posti dall'attività, all'organizzazione) nei confronti degli intermediari che prestino analoghi servizi senza operare differenziazioni sulla base di quali siano le tecnologie impiegate per fornirli, nonché ad eventuali specificità che tali servizi possano assumere proprio in ragione del supporto tecnologi-

[28] Al riguardo, assume anche rilievo la disciplina di cui al regolamento (UE) 2022/2554 del Parlamento Europeo e del Consiglio del 14 dicembre 2022 relativo alla resilienza operativa digitale per il settore finanziario e che modifica i regolamenti (CE) n. 1060/2009, (UE) n. 648/2012, (UE) n. 600/2014, (UE) n. 909/2014 e (UE) 2016/1011.

[29] T.N. Poli, *MiCA, Pilot Regime e Decreto Fintech: la regolazione del fenomeno crypto e le difficoltà di inquadramento nel sistema finanziario*, in Dialoghi di diritto dell'economia, dicembre 2023.

co. La debolezza di un siffatto modello è che si rischia di non identificare correttamente le insidie derivanti dall'impiego di tecnologie *disruptive*, sottostimandone gli effetti e limitando gli spazi di intervento delle autorità di controllo. Una seconda accezione, invece, è quella che pone l'attenzione sull'autorità di vigilanza: dunque, è neutrale quel modello di supervisione che utilizzi uno strumentario omogeneo in relazione a tutti i soggetti e attività vigilate, senza impiegare *tech-specific tools*[30] che consentirebbero invece di monitorare in modo più efficace le caratteristiche dell'oggetto della vigilanza, così da intervenire quando ritenuto opportuno.

5.2 La supervisione sulle applicazioni di AI nel settore finanziario[31] può quindi far uso - purché l'autorità di vigilanza sia dotata di idonee competenze tecniche e sufficienti risorse - di uno strumentario tecnologico, sovente *tech-specific*, e può modulare le proprie modalità d'intervento tenendo conto delle specificità dell'oggetto della vigilanza[32].
Tra gli strumenti di SupTech, l'applicazione del modello della *supervision as code*[33] ha rilevanti potenzialità:
la risposta di vigilanza, espressa in termini di codice ed inclusa nella tecnologia impiegata dal prestatore di servizi fa sì che questa possa essere resa eseguibile dalla macchina[34], monitorando eventuali evoluzioni dello strumento, degli output che restituisce e consentendo un intervento tempestivo.
In questo contesto, la non neutralità può consistere nelle indicazioni delle autorità di vigilanza circa il tipo di codice da utilizzare, le caratteristiche degli strumenti tecnologici impiegati e le modalità di interazione tecnologica che possono avvenire tra vigilante e vigilato, risultando potenzialmente particolarmente invasiva della libertà organizzativa del prestatore di servizi.
Si consente così alle autorità di effettuare una vigilanza assai più effettiva ed efficace rispetto ai meri controlli ispettivi e documentali, che rimangono comunque indispensabili in quanto complementari.
Non neutrali potrebbero risultare anche le cd. valutazioni di conformità "*in action*" dei sistemi di AI impiegati dall'intermediario, le quali potrebbero essere condotte anche alimentando i sistemi in sperimentazione con dati selezionati e predisposti dalle au-

[30] A ciò si aggiungano anche le ipotesi di utilizzo da parte delle autorità di vigilanza di strumenti di intelligenza artificiale al fine di svolgere in modo più efficace ed efficiente i propri compiti. Su tali applicazioni di suptech, si veda ad esempio A. Azzutti-P. Magalhães Batista-W.G. Ringe, *Navigating the Legal Landscape of AI-Enhanced Banking Supervision: Protecting EU Fundamental Rights and Ensuring Good Administration*, EBI Working Paper Series 2023 – no. 140, aprile 2023.

[31] G. Siani, *AI-driven bank: Opportunità e sfide strategiche per il sistema finanziario e la vigilanza*, intervento del Capo del Dipartimento Vigilanza bancaria e finanziaria della Banca d'Italia, Banking Insight 2023, Boston Consulting Group Milano, 3 Ottobre 2023.

[32] M. Rabitti-A. Sciarrone Alibrandi, *RegTech e SupTech*, in A. Pajno-F. Donati-A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?,* Quaderni Astrid, Bologna, 2022, 439 ss.

[33] Il modello di riferimento è quello delle *«rules as code»*. Al riguardo cfr., J. Mohun-A. Roberts, *Cracking the Code: Rulemaking for Humans and Machines*, OECD Working Papers on Public Governance No. 42, 2020, 39 ss. In senso critico, J. Oster, *Code is code and law is law—the law of digitalization and the digitalization of law*, in *International Journal of Law and Information Technology*, Volume 29, Issue 2, 2021, 101 ss.

[34] A. Celotto, *Verso l'algoretica. Quali regole per le forme di intelligenza artificiale?*, in V. Falce (a cura di), *Strategia dei dati e intelligenza artificiale*, Torino, 2023, in particolare 28 ss.

torità, consentendo di far emergere possibili debolezze ed effetti discriminatori. Nel caso in cui l'autorità, a conclusione della verifica, riscontrasse errori nelle performance degli algoritmi, dovrebbe avere il potere di imporne la sospensione e l'eventuale correzione del suo funzionamento, anche attraverso l'introduzione di correttivi e filtri da applicare agli output.[35]

Rendendo le regole e gli strumenti di reazione *embedded* nell'algoritmo utilizzato per la prestazione di servizi si rende la supervisione più efficace, meglio in grado di tutelare gli investitori, vigilare sull'osservanza delle disposizioni in materia finanziaria previste dall'ordinamento, al contempo salvaguardando il buon funzionamento del sistema finanziario e la fiducia degli investitori[36].

Anche in questo caso non mancano le controindicazioni. Anzitutto, non ogni regola o aspettativa di vigilanza può essere resa codice, ma solo quelle disposizioni che siano prescrittive ed univoche; d'altra parte, una siffatta modalità operativa delle autorità di vigilanza può risultare molto onerosa poiché richiede specifiche competenze tecniche ed è fortemente legata alle singole tipologie di servizi offerti. Prioritaria, dunque, potrebbe esserne l'applicazione con riferimento a quelle applicazioni di AI che determinano rischi ritenuti più pressanti.

6. *Roboadvisors*, etica e tutela dell'investitore

6.1 Tra le applicazioni di AI al settore finanziario, i servizi di *roboadvice*[37] costituiscono un rilevante terreno di prova. Sia in ragione della loro diffusione - e quindi alla capacità di determinare effetti per ampie fasce di utenti - sia per la loro insidiosità. Grazie alla loro accessibilità e ai costi ridotti rispetto alle alternative tradizionali costituiscono, infatti, anche un rilevante canale di accesso ai servizi finanziari (e, ancor prima, alle informazioni finanziarie) e pertanto usi impropri o fuorvianti possono cagionare pregiudizio agli utenti/investitori.

Come accennato, nell'ambito della ampia categoria di consulenza finanziaria che si avvale di sistemi di intelligenza artificiale si possono annoverare servizi caratterizzati da modelli operativi molto differenti. Tradizionalmente viene operata una distinzione tra *roboadvisors* puri e ibridi[38], utile a fini descrittivi ma anche idonea a rappresentare

[35] La vigilanza che operasse in questo modo applicherebbe di fatto regole di "secondo ordine". cioè filtri inseriti dal programmatore per correggere eventuali malfunzionamenti (ovvero risultati che rischiano di produrre un pregiudizio per l'individuo o di violare una disposizione di legge) del processo di elaborazione dei dati. Sul tema F. Schauer, *Second-Order Vagueness in the Law*, in G. Keil-R. Poscher (a cura di), *Vagueness and Law: Philosophical and Legal Perspectives*, Oxford, 2016, 177 ss.

[36] Art. 5, c. 1, TUF.

[37] Consob, *La digitalizzazione della consulenza in materia di investimenti finanziari*, in N. Linciano-P. Soccorso-R. Lener (a cura di), Quaderni Fintech n. 3, gennaio 2019; F. Sartori, *La consulenza finanziaria automatizzata: problematiche e prospettive*, in *Rivista Trimestrale Diritto dell'Economia*, 1, 2018, 258 ss.; D. Rossano, *Il Robo advice alla luce della normativa vigente*, in F. Capriglione (a cura di), *Liber Amicorum Guido Alpa*, Padova, 2019, 365 ss.

[38] Una puntuale distinzione è delineata da Consob, *La digitalizzazione della consulenza in materia di investimenti finanziari*, cit., 10 ss. A ciò si aggiunga l'ulteriore - e assai diversa - fattispecie del robo4advisor, servizio BtoB, offerto dal prestatore di servizi (consulente professionale) ad altro consulente finanziario.

l'impatto, assai differente, che questi possono avere sugli utenti. I primi costituiscono forme di consulenza completamente automatizzata, basate sull'intelligenza artificiale e veicolate all'investitore attraverso interfacce che non prevedono l'interazione con operatori umani, né attività di filtro operate da questi ultimi; nell'ambito dei servizi di *roboadvisor* ibrido – decisamente più diffusi – per quanto l'apporto della valutazione operata mediante consulenti robotizzati possa essere comunque decisiva, vi è un intervento umano. Il ruolo dell'AI è quindi mediato: alle fasi dell'attività automatizzate se ne alternano altre in cui è prevista l'interazione con persone fisiche[39], che non solo sono chiamate a correggere eventuali "errori" dell'AI, ma anche a seguire il cliente in tutte le fasi della consulenza. Il consulente finanziario umano opera come filtro ed è vincolato da obblighi di professionalità e buona fede[40].

Spesso i servizi di *roboadvice* puro si rivolgono, quale cliente target, alla clientela *retail*, e ciò ne acuisce i profili di rischio. Anzitutto, in relazione alla consapevolezza del cliente circa la tipologia di contratto che sta concludendo, a maggior ragione quando ciò avvenga nell'ambito di piattaforme generaliste.

Ad ogni modo, anche quando sia prestata mediante tali modalità, la consulenza personalizzata ed individualizzata soggiace alle regole di condotta previste dalla disciplina europea e del Regolamento emittenti CONSOB[41]. Quanto all'effettività di tale regime normativo[42], va evidenziato come la mancanza di interazione con un consulente umano – sia pure online – potrebbe impattare negativamente sulla veridicità dei dati raccolti, inducendo il cliente ad una rapida scelta di cui potrebbe non percepire la rilevanza. Nella prospettiva delle regole di comportamento non solo è dunque fondamentale che siano rispettati i medesimi requisiti di correttezza dei consulenti umani, che l'interazione sia svolta in maniera etica e non volta a trarre in inganno la controparte, ma anche che i procedimenti di raccolta di informazioni sul cliente, nell'ambito delle *know your customer rules,* siano effettivi. Se con riferimento ai consulenti finanziari persone fisiche, autorizzati dall'autorità di vigilanza, sono previste penetranti regole volte garantire che questi agiscano nel rispetto dei canoni deontologici, la verifica circa il rispetto delle stesse può risultare più complessa in relazione ai servizi di *roboadvice* basati su algoritmi a cui non è agevole avere accesso. Le autorità di vigilanza si trovano dunque di fronte alla sfida di valutare le caratteristiche tecnologiche del *roboadvisor* al momento della loro autorizzazione, ma anche a monitorarne efficacemente gli sviluppi successivi.

Anche considerando tali elementi, è elevato il rischio che al cliente sia veicolata una informazione fuorviante – quantomeno nelle modalità comunicative – al fine di indurlo a concludere contratti: il *roboadvisor*, pur di raggiugere il proprio obiettivo, potrebbe discostarsi dai modelli "etici" che gli siano stati sottoposti come input in fase di *trai-*

[39] R. Lener, *Intelligenza Artificiale e interazione umana nel robo-advice*, in Rivista Trimestrale Diritto dell'Economia, 3s, 2021, 101 ss.

[40] U. Gasser-C. Schmitt, *The Role of Professional Norms in the Governance of artificial intelligence*, in M.D. Dubber- F. Pasquale-S. Das (a cura di), *The Oxford Handbook of Ethics of AI*, Oxford, 2020, 141 ss.

[41] Regolamento CONSOB n. 20307/2018, artt. 40 ss. che si pone il linea con quanto previsto dal regolamento delegato UE/2017/565 artt. 54 e 55.

[42] Sul punto, P. Maume, *Robo-advisors How do they fit in the existing EU regulatory framework, in particular with regard to investor protection?,* study Requested by the ECON committee, giugno 2021.

ning[43]. Ciò non solo sarebbe pregiudizievole per il cliente, ma violerebbe espressamente l'obbligo gravante sull'intermediario di agire nel *best interest* del cliente[44].

6.2 In relazione al caso d'uso del *roboadvice* è possibile, dunque, dimostrare le argomentazioni finora sostenute.

Il *roboadvice* puro, quello che – sebbene sia raramente offerto nella pratica – maggiormente costituisce una minaccia per gli investitori in quanto "scavalca" l'interazione con soggetti professionali, tenuti a doveri di diligenza, è foriero di pericoli. Questi, solo in parte, sono neutralizzati dall'applicazione delle regole ordinariamente previste per la consulenza finanziaria[45], le quali sono costruite individuando un soggetto responsabile – l'intermediario – destinatario di obblighi. L'utilizzo di tecnologie avanzate può attenuare le forme di possibile controllo da parte di quest'ultimo (nonostante l'intermediario-*deployer* ne rimanga pienamente responsabile sia ai sensi della disciplina finanziaria, sia in base al modello delineato dagli artt. 16 ss. dell'AI Act), aprendo la strada a rischi nuovi.

È dunque più che mai necessario realizzare una *trustworthy AI*, obiettivo espressamente al centro delle strategie UE[46] ed è essenziale che le istituzioni finanziarie, che impieghino tali tecnologie, siano in grado di offrire una tutela piena ed effettiva dei diritti fondamentali dei propri clienti[47].

Ciò passa non solo dalla corretta progettazione della tecnologia, dell'individuazione dei servizi tecnologizzati più adatti per le preferenze della clientela, ma anche dal continuo monitoraggio umano degli stessi[48].

Tali difese consentono anche di fronteggiare il problema delle "allucinazioni"[49] nell'AI generativa, ipotesi in cui il modello genera informazioni che siano falsamente basate su dati reali, particolarmente problematica quando l'AI venga utilizzata in contesti in cui l'accuratezza delle informazioni è fondamentale. Si pensi, ad esempio, quando il servizio di consulenza finanziaria robotizzata produca tali allucinazioni, con effetti fuorvianti per il cliente al quale potrebbero così essere offerti prodotti inadeguati.

---

[43] Similmente a quanto avviene nel caso di comportamenti collusivi autonomamente sviluppati da parte di sistemi di intelligenza artificiale indipendenti, in quanto ciascuno volto a massimizzare il proprio profitto. A. Ezrachi-M.E. Stucke, *Artificial Intelligence & Collusion: When Computers Inhibit Competition*, in *University of Illinois Law Review*, 2017, 1776 ss.

[44] ESMA, *Public Statement On the use of Artificial Intelligence (AI) in the provision of retail investment services*, 30 maggio 2024, ESMA35-335435667-5924, spec. 4.

[45] F. Annunziata, *Servizi e attività di investimento. Definizioni e accesso al mercato*, in *La disciplina del mercato dei capitali*, Torino, 2023, 175 ss.

[46] Ad esempio nell'ambito dei seguenti documenti: Commissione Europea, Communication on Artificial Intelligence for Europe, (COM(2018) 237 final) e Gruppo di esperti di alto livello, *Orientamenti etici per un'IA affidabile*, 2019. In tema, I. Carnat, *Ethics Lost In Translation: Trustworthy Ai From Governance To Regulation*, in *Opinio Juris in Comparatione*, 1(1), 2023, 90 ss.; G. Comandé, *Unfolding the legal component of trustworthy AI: a must to avoid ethics washing*, in *Annuario di diritto comparato*, Napoli, 2020, 39 ss.

[47] O. Pollicino, *AI ACT: «Banche e assicurazioni preparate alla Valutazione di impatto sui diritti fondamentali»*, Conversazione su Democrazia digitale alla ricerca di check and balance, su *Altalex.com*, 29 luglio 2024.

[48] D. Rossano, *L'Intelligenza Artificiale: ruolo e responsabilità dell'uomo nei processi applicativi (alcune recenti proposte normative),* in *Rivista Trimestrale di Diritto dell'Economia*, 3, 2021, 212.

[49] ESMA, *Public Statement On the use of Artificial Intelligence,* cit., 3.

Questi risultati erodono la fiducia del pubblico nei sistemi di AI, ma anche nelle istituzioni finanziarie che utilizzano questa tecnologia: l'"ambiente" in cui operano i sistemi di intelligenza artificiale finanziaria impone una maggiore cautela essendo la fiducia negli intermediari finanziari e nel mercato bancario un prerequisito per il loro funzionamento.

Pertanto, per mitigare le allucinazioni dell'AI è necessaria un'attenta progettazione dei modelli, basata su training con dati di alta qualità e aggiornamenti frequenti con informazioni accurate, accanto ad un'indispensabile supervisione umana[50].

Sempre con riferimento alla tutela dell'investitore, gli intermediari che facciano utilizzo dell'AI per offrire servizi automatizzati - in tutto o in parte - dovrebbero svolgere un *assessment* puntuale dell'impatto che tali strumenti possono avere sui diritti fondamentali dei propri clienti[51] (similmente a quanto previsto dall'art. 27 AI Act), adottando contromisure più consistenti o addirittura giungendo a desistere dall'utilizzo qualora si ritenga che sia messa a repentaglio la libertà di scelta del soggetto o la *gamification* dell'esperienza possa alterare la capacità decisionale dell'individuo (ipotesi già vietata ai sensi dell'AI Act).

In particolare, quanto alle contromisure da adottare, è auspicabile che siano incluse anche contromisure tecnologiche: risulta infatti opportuno includere nello stesso algoritmo meccanismi che consentano al *deployer* e alle autorità di vigilanza un effettivo monitoraggio del funzionamento della tecnologia – sincerandosi che questa non devii da quanto originariamente previsto – potendo così bloccare eventuali evoluzioni indesiderate dello strumento. In questo senso, si tratterebbe certamente di un approccio non tecnologicamente neutrale, ma che appare legittimo perché volto a garantire che la tecnologia operi nell'alveo di quanto autorizzato e non determini rischi (ulteriori rispetto a quelli che deriverebbero dalla prestazione della medesima attività con altre modalità) pregiudizievoli per l'intermediario, per i consumatori e per il mercato.

7. *Virtual worlds*: luogo in cui l'esigenza di tutela diviene (ancora più) sentita

7.1 Una delle più recenti frontiere dell'interazione online è rappresentata dai *virtual worlds*, mondi virtuali caratterizzati da una forte immersività[52].

Calare i servizi di *roboadvice* nell'ambito dei *virtual worlds* fa emergere ancora più chiaramente gli scenari (e le criticità) messi sinora in evidenza e rende manifesta l'esigenza di una decisa reazione dell'ordinamento, adottando un approccio *tech-specific*.

In questo contesto, caratterizzato dall'immersività della esperienza dell'utente, la vul-

[50] Artt. 14, 26 AI Act con riferimento ai sistemi ad alto rischio.

[51] O. Pollicino, *AI ACT: «Banche e assicurazioni preparate alla Valutazione di impatto sui diritti fondamentali»,* cit. Si tratta di una valutazione da condurre non solo in relazione ai servizi di roboadvice o all'utilizzo a sistemi di credit scoring che potrebbero avere l'effetto di escludere un soggetto dall'accesso al credito, ma anche con riferimento ad altri utilizzi dell'AI nell'ambito dell'organizzazione della banca, che potrebbero incidere sui diritti dei clienti.

[52] Sul tema sia consentito rinviare a F. Di Porto-D. Foà-S. Ennis, *Emerging Virtual Worlds: Implications for Policy and Regulation*, CERRE – Centre on Regulation in Europe, febbraio 2024.

nerabilità digitale[53] risulta fortemente accentuata. Ciò in ragione della possibilità di acquisire dati biometrici e cognitivi, anche attraverso l'*eye tracking,* e la presenza di *avatars* (che molto spesso sono la proiezione, anche nei connotati fisici, dell'essere umano che agisce), in uno con la forte personalizzazione dei servizi, possono ridurre la percezione dei rischi.

Ciò può essere determinato anche dall'inconsapevole sottoposizione a tecniche di profilazione e alla soggezione a pratiche commerciali scorrette che incidono in modo sempre più rilevante sulla consapevolezza dell'investitore circa le conseguenze delle proprie scelte e azioni.

All'utente, dunque, dovrà essere garantito un sufficiente livello di comprensione e consapevolezza del tipo di attività che si sta svolgendo: se un servizio appare "come un videogioco", l'investitore potrebbe non essere sufficientemente consapevole del tipo di attività che si sta svolgendo, dei relativi rischi anche in termini di possibile impatto sul proprio patrimonio. Difatti, le specifiche modalità di interazione possono diminuire la comprensione dei fenomeni e dei rischi connessi, incidendo sul processo decisionale dell'utente. In primo luogo, per evitare che ciò accada, è necessario che gli intermediari agiscano sempre in piena applicazione dei principi di correttezza, buona fede e diligenza, cercando così - con ogni mezzo possibile - di sterilizzare i possibili rischi derivanti da tali specifiche vulnerabilità ulteriori[54].

La personalizzazione dei servizi – che risponde anzitutto all'esigenza di fornire prodotti maggiormente in linea con le preferenze degli utenti – consente altresì di trarre vantaggio dei limiti cognitivi dei consumatori, sfruttando le vulnerabilità individuali con un approccio commerciale granulare[55].

Le insidie per l'investitore potrebbero poi aumentare considerevolmente a causa della completa personalizzazione dell'esperienza (si pensi alla possibilità di fornire a ciascun individuo un consulente che sia compatibile con il proprio *background* culturale), in grado di innescare fenomeni di *familiarity bias*. Inoltre, la sensorialità dell'esperienza del metaverso (abilitata dalle interfacce aptiche) può creare situazioni inedite in cui i sensi dei clienti possono essere stimolati in modo tale da spingerli verso una determinata scelta. Si tratta di caratteristiche – peculiari dell'offerta di servizi finanziari in ambienti immersivi *phygital* – che dovranno essere prese in considerazione nei processi di *product oversight and governance* dei produttori e distributori.[56].

Per far fronte ai citati bisogni di tutela emergenti, i confini tra le aree presidiate dalle normative sulle pratiche commerciali scorrette, le regole di condotta, i doveri di informazione e le regole di negoziazione a distanza appaiono sempre più labili[57]; proprio

---

[53] N. Helberger et al., *Choice Architectures in the Digital Economy: Towards a New Understanding of Digital Vulnerability*, in *Journal of Consumer Policy*, 45, 2022, 175 ss.

[54] Sul concetto di vulnerabilità del consumatore digitale si veda F. Lupiáñez-Villanueva et al., *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation*, Final Report European Commission, aprile 2022.

[55] A. Davola, *Fostering Consumer Protection in the Granular Market: the Role of Rules on Consent, Misrepresentation and Fraud in Regulating Personalized Practices*, in Technology and Regulation, 2021, 76–86.

[56] EBA, *Guidelines on product oversight and governance arrangements for retail banking products*, EBA/GL/2015/18, 15 giugno 2015.

[57] F. Annunziata, *Towards an EU charter for the protection of end users in financial markets*, EBI Working

queste ultime sono oggetto di proposte di riforma volte ad introdurre rimedi per mitigare la vulnerabilità dell'utente in ambienti digitali sofisticati[58].

Va evidenziato però come né il rafforzamento delle regole di *disclosure*, né la previsione di rimedi puramente ex post (come il diritto di recesso), paiano idonei a garantire sufficienti livelli di protezione per il cliente e a neutralizzare i rischi insiti in tali modelli di contrattazione. Questi, infatti, potrebbe comunque sperimentare un'esperienza d'uso negativa e lesiva che minerebbe la sua fiducia nel sistema.

In questo contesto, il legislatore e le autorità di vigilanza saranno quindi chiamati a valutare se i rischi posti dalla prestazione di servizi finanziari in ambiente *phygital* e basati su AI siano da considerare intollerabili e richiedano una specifica disciplina: se sia necessario delimitare le tecnologie utilizzabili, le caratteristiche degli ambienti immersivi (in relazione, ad esempio, a colori che possano sollecitare specifiche reazioni) e ai livelli di personalizzazione (e.g., vietando che gli *avatars* antropomorfi riproducano soggetti aventi la medesima etnia del cliente) in modo da evitare che le vulnerabilità del cliente possano essere sfruttate a suo sfavore. Similmente a quanto già previsto nella proposta di modifica della direttiva sui servizi finanziari conclusi a distanza, appare auspicabile limitare la possibilità per la banca di impiegare strumenti – nell'ambito della propria interfaccia online – che potrebbero distorcere o compromettere la capacità dei consumatori di prendere una decisione o una scelta libera, autonoma e informata[59].

Appare ad ogni modo essenziale garantire la trasparenza sul livello di personalizzazione dell'esperienza utente e dei servizi offerti, in modo da consentire all'utente di comprendere le ragioni per cui determinati servizi gli vengono offerti nonché le peculiarità che li differenziano da quelli offerti ad altri utenti.

Quest'ultima costituirebbe un utile elemento complementare, ma certamente insufficiente come unico presidio di tutela. In altri termini, di fronte a tecnologie così avanzate, in grado di distorcere la percezione umana spingendo il consumatore ad assumere decisioni d'investimento, la mera realizzazione del principio del *caveat emptor,* con la consegna di una ricca informativa al cliente lasciando poi a quest'ultimo la scelta se concludere il contratto risulta del tutto inadeguata. Al contrario, un livello adeguato di tutela potrebbe essere ottenuto mediante una regolazione *tech-specific*, che ponga precisi limiti "tecnologici" per la conclusione di contratti finanziari in ambiente immersivo.

Papers Series, no. 128/2022.

[58] È già stata formulata una proposta sulla modifica della disciplina sulla commercializzazione a distanza dei servizi finanziari, cfr. Commissione europea, Proposta di direttiva del Parlamento europeo e del Consiglio che modifica la direttiva 2011/83/UE concernente i contratti di servizi finanziari conclusi a distanza e che abroga la direttiva 2002/65/CE, COM(2022) 204 definitivo 2022/0147 (COD), 11.5.2022.

[59] Potrebbe, al riguardo, risultare utile l'introduzione di strumenti per la valutazione dell'effettiva comprensione da parte del cliente delle implicazioni dell'utilizzo di determinate tecnologie (anche prevedendo effetti bloccanti, similmente a quanto avviene con il test di adeguatezza di cui alla MIFID). Tale ipotesi è stata prospettata da F. Annunziata, *Retail Investment Strategy How to boost retail investors' participation in financial markets*, Study requested by ECON Committee, giugno 2023, 7.

8. Conclusioni

8.1 L'applicazione delle tecnologie di intelligenza artificiale nell'ambito del settore finanziario ha numerose potenzialità, in grado di rendere la prestazione di servizi più efficace ed efficiente; rilevanti sono anche gli effetti positivi in termini di inclusione sociale, rendendo i servizi finanziari accessibili anche alle fasce della popolazione *underbanked*. A ciò si aggiungono rilevanti applicazioni aventi "rilevanza interna" ai fini di una efficiente organizzazione delle istituzioni finanziarie[60].

Al contempo, molti e rilevanti sono i rischi derivanti dalle medesime tecnologie. E, come evidenziato, le insidie sono ben maggiori (e certamente presentano profili di specificità) nel settore finanziario rispetto all'ordinamento generalmente considerato. Tali considerazioni suggeriscono perciò che la regolazione finanziaria assuma idonee contromisure.

In primo luogo, occorre che sia garantita la governabilità dell'innovazione, anche incidendo sulla *governance* del mercato finanziario "tecnologizzato". E per fare ciò, non appare più sufficiente la regolazione basata su attività, che non tenga in debita considerazione anche le specificità dei processi. Questi ultimi sono spesso specifici in relazione alle singole tecnologie.

In tale contesto, è quindi ormai indispensabile che il legislatore disciplini specifiche applicazioni tecnologiche, perché non è indifferente il modo in cui viene svolta una certa attività: pertanto, non sarà sufficiente prevedere una disciplina "generale" per l'AI finanziaria, ma occorrerà valorizzare le specificità poste dalle sue varie applicazioni.

L'adesione fideistica al principio di neutralità tecnologica potrebbe invece pregiudicare i livelli di tutela, consentendo la proliferazione di rischi, aprendo rilevanti *vulnera* per gli utenti e per gli altri operatori del mercato.

Anche nella prospettiva della vigilanza finanziaria, un approccio tecnologicamente neutrale non appare più idoneo a ottenere i risultati che gli sono richiesti: non appare percorribile garantire la sana e prudente gestione dell'intermediario e la tutela degli investitori senza poter operare controlli intrusivi e *tech-specific*.

Come ben evidenziato dal caso d'uso "di frontiera" dei *virtual worlds*, alla vigilanza bancaria e finanziaria deve essere consentito un approccio tecnologicamente orientato che quindi calibra differentemente le regole e strumenti d'intervento in base alle tecnologie impiegate per fornire un servizio, riconoscendo come questi strumenti tecnologici siano in grado di caratterizzare il servizio stesso, mutandone i rischi caratteristici. Pertanto, deve essere incoraggiato l'uso degli strumenti tecnologici nel contesto della vigilanza sulle applicazioni, per operare un efficace monitoraggio sull'utilizzo, limitare gli abusi ed avere la possibilità di intervenire tempestivamente. Inoltre, potrà essere adottato l'approccio del *supervision as code*, spingendosi fino a introdurre *by design* nell'algoritmo "controlimiti". Per poter essere concretamente operative ed efficaci, regole autoeseguibili di questo tipo dovranno avere contenuto prescrittivo e non prevedere spazi di discrezionalità. Dunque, è chiaro che non ogni regola di condotta potrà essere così codificata, ma tale strategia rappresenterò comunque un utile meccanismo com-

[60] M. Pellegrini, *L'intelligenza artificiale nell'organizzazione bancaria: quali sfide per il regolatore?*, in *Rivista Trimestrale di Diritto dell'Economia*, 3, 2021, 422 ss.

plementare alle tradizionali modalità operative della vigilanza bancaria e finanziaria. Per poter ottenere tali risultati sarà in ogni caso preziosa la cooperazione tra industria e autorità di vigilanza nello sviluppo delle soluzioni più appropriate (anche utilizzando strumenti simili alle *sandboxes* regolamentari[61]).

8.2 Tornando all'esempio del *roboadvisor*, impiegando un siffatto strumentario *tech specific,* l'autorità di vigilanza – pur sempre cooperando con l'intermediario che faccia uso della tecnologia – potrà sincerarsi che il *roboadvisor* operi entro i confini di quanto autorizzato. Ciò consentirebbe agli utenti di godere di una tutela paragonabile – e probabilmente superiore – a quella attualmente realizzata nel contesto dei rapporti con consulenti finanziari fisici: la libertà di autodeterminazione, nonché la dignità del cliente potranno essere tutelati in modo più efficace.
La disciplina della prestazione di servizi finanziari mediante intelligenza artificiale nell'ambito di ambienti digitali immersivi potrà essere il luogo di sperimentazione per introdurre una disciplina tecnologicamente specifica, finalizzata a limitare possibili abusi da parte dell'intermediario e a scongiurare gli effetti pregiudizievoli delle sue vulnerabilità.
L'esigenza di introdurre regole di questo tipo nel settore bancario e finanziario è più pressante che in altri settori, proprio per l'esistenza della necessità di salvaguardare – oltre alla libera formazione della volontà del consumatore – altri interessi di rilievo pubblicistico, in quanto il venire meno della fiducia dell'utente non sarebbe limitata ad uno specifico intermediario (o al solo "metaverso bancario"), ma avrebbe con ogni probabilità ricadute sulla fiducia nel sistema bancario e finanziario nel suo complesso.

[61] S. Ranchordas - V. Vinci, *Regulatory Sandboxes and Innovation-friendly Regulation: Between Collaboration and Capture,* in *Italian Journal of Public Law,* 16(1), 2024, 107 ss.

Per una nuova teorica della regolazione "forte" delle piattaforme digitali tra (necessario) intervento pubblico e tutela (necessaria) della libertà di espressione*

Lorenzo Ricci

Abstract

A seguito dell'adozione dei regolamenti 2022/1925 e 2022/2065 sembra aprirsi la strada per un mutamento di paradigma in punto di regolazione delle piattaforme digitali, superando l'idea dell'autosufficienza della regolazione di carattere antitrust. Al contrario, la direzione intrapresa pare suggerire l'avvio di una stagione caratterizzata da una più forte regolazione pubblica, diminuendo così gli spazi di auto-regolazione. Se questa tendenza sembra da salutare con favore, anche e soprattutto al fine di contenere i forti poteri privati digitali, si pone l'esigenza, costituzionalmente imposta, affinché tutto ciò non si traduca in una restrizione della sfera della libertà di espressione. Pertanto, è necessario ragionare sui caratteri di tale nuova forma di regolazione, sul presupposto che ogni riflessione in merito non possa prescindere dalla consapevolezza che il "punto logico di partenza" è la libertà dell'individuo, nelle sue molteplici accezioni, a partire da quella di espressione.

Following the adoption of regulations 2022/1925 and 2022/2065, the way seems to be opening up for a paradigm shift in terms of the regulation of digital platforms, going beyond the idea of the self-sufficiency of antitrust regulation. On the contrary, the direction taken seems to suggest the start of a season characterised by stronger public regulation, thus diminishing the spaces for self-regulation. While this trend appears to be positive, also and above all in terms of containing strong private digital powers, there is a constitutional need to ensure that it does not translate into a restriction of the sphere of freedom of expression. It is therefore necessary to reflect on the characteristics of this new form of regulation, on the assumption that any analysis on the matter cannot disregard the awareness that the "logical starting point" is the freedom of the individual in its multiple meanings, starting with that of expression.

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

Sommario

Keywords

piattaforme digitali – regolazione pubblica – poteri privati – poteri pubblici – libertà di manifestazione del pensiero.

1. Premessa introduttiva

Il tema delle piattaforme digitali riveste sempre più importanza nel dibattito giuridico. È noto, infatti, che larga parte della vita degli esseri umani (almeno con riferimento ai paesi sviluppati) si svolga in "rete" e da essa sia fortemente condizionata.

In questa sede interessa, in particolare, riflettere attorno al tema della regolazione di tali piattaforme, anche e soprattutto alla luce di taluni interventi piuttosto recenti del legislatore europeo che hanno posto l'attenzione sul profilo della necessità di regolare le piattaforme digitali e porre, dunque, un freno ai poteri privati[1] che di tali piattaforme sono i proprietari o, come è stato detto, i "padroni"[2]. Si tratta di poteri particolarmente forti che, proprio in ragione della loro forza e rilevanza, non possono che essere attenzionati e, quindi, regolati, dal diritto, anche e soprattutto per la loro rilevanza per così dire "sociale". Invero, tali poteri finiscono per incidere su una serie di diritti e libertà fondamentali di ciascun individuo, oltre a rivestire un potere di mercato così forte da condizionare la politica e, dunque, di fatto, l'intera società.

La tendenza alla quale si assiste è quella di una sempre maggiore attenzione, da parte dei pubblici poteri, nei confronti delle piattaforme digitali, i quali sembrano ormai concordi circa la necessità di regolare tale settore della società e di ridimensionare, di conseguenza, il relativo potere privato. Se ciò appare positivo non sono, tuttavia, da

[1] Sul rapporto fra poteri privati e regolazione cfr. G. Resta, *Poteri privati e regolazione*, in *Enc. dir., I tematici, Potere e Costituzione*, Milano, 2023, 1023 ss. Sui poteri privati, nella prospettiva del diritto pubblico, cfr. il terzo numero del 2021 della rivista *Diritto pubblico*, oltre che il libro di M.R. Ferrarese, *Poteri nuovi. Privati, penetranti, opachi*, Bologna, 2022, ed il contributo di E. Bruti Liberati, *Poteri privati e nuova regolazione*, in *Diritto pubblico* 1, 2023, 285 ss. Sui poteri pubblici e privati nel mondo digitale cfr. L Torchia, *Poteri pubblici e poteri privati nel mondo digitale*, in *Il Mulino*, 1, 2024, 14 ss. Della stessa A. cfr. anche *Lo Stato digitale. Una introduzione*, Bologna, 2023.

[2] Sul punto cfr. M. Betzu, *I baroni del digitale*, Napoli, 2022, *passim*. In proposito cfr. anche F. Paruzzo, *I sovrani della rete. Piattaforme digitali e limiti costituzionali al potere privato*, Napoli, 2022.

sottovalutare i rischi che si potrebbero verificare e che si identificano qui nella possibile riduzione della libertà di manifestazione del pensiero in rete, la quale già non è sempre pienamente tutelata dalle piattaforme e non può certo correre il rischio di sopportare ulteriori limitazioni da quegli stessi soggetti che, in teoria, dovrebbero, viceversa, tutelarla: ossia i pubblici poteri.

Pertanto, dopo una parte iniziale in cui si dedicherà l'attenzione al modello di regolazione che sembra emergere a livello europeo[3] – non prima di aver rapidamente ricostruito la fase precedente dell'auto-regolazione ed aver, inoltre, cercato di mettere in luce le criticità della c.d. regolazione antitrust –, l'analisi si soffermerà sull'ipotesi di una "regolazione forte" delle piattaforme digitali come risposta necessaria allo strapotere dei colossi del digitale[4], sottolineando tuttavia l'esigenza che tale regolazione sia anche flessibile, tentando, quindi, di mettere in evidenza le caratteristiche del modello di regolazione qui prospettato.

Infine, si rifletterà sulle possibili implicazioni del nuovo potenziale modello regolatorio sulla libertà di manifestazione del pensiero, cercando di mettere in luce, da un lato, l'effetto positivo che una regolazione di questo tipo può comportare rispetto a tale libertà e, dall'altro, i rischi ad essa connessi, che sono quelli di una censura uguale o, addirittura, maggiore di quella che già viene talvolta posta in essere dalle piattaforme private.

Sulla sfondo, taluni possibili aspetti problematici, sia sul fronte del modello di "regolazione forte" in funzione limitativa dello strapotere delle piattaforme digitali, che su quello della necessità di una tutela piena ed effettiva del contenuto sotteso alla libertà di manifestazione del pensiero in rete.

Il punto dal quale non si può in alcun modo prescindere, e che costituisce, dunque, il *prius* di qualsiasi discorso, è che il fine ultimo di ogni ordinamento giuridico sia sempre e comunque la tutela dei diritti e delle libertà dell'individuo, coerentemente con il carattere personalista che informa l'ordinamento europeo e la quasi totalità delle Costituzioni che ne fanno parte.

2. Il potere pubblico (e la sua "democraticità") vs il potere privato (e la sua "autoritarietà") e lo spostamento della "autoritarietà" dal primo al secondo: andata e (rischio di) ritorno

Preme ora sottolineare come, nel corso degli ultimi decenni, con particolare riferimento al mondo del digitale, si sia assistito ad una crescita dei poteri privati che, a tratti, è sembrata (e continua ad apparire) inarrestabile, con un ridimensionamento di quelli pubblici; di conseguenza, pare utile riflettere sulle due differenti tipologie di potere,

[3] In proposito, per un inquadramento piuttosto recente, cfr. F. Donati, *Verso una nuova regolazione delle piattaforme digitali*, in *Rivista della regolazione dei mercati*, 2, 2021, 238 ss.

[4] Sui poteri digitali cfr. O. Pollicino, *Potere digitale*, in *Enc. dir., I tematici, Potere e Costituzione*, Milano, 2023, 410 ss. In tema cfr. anche A. Simoncini, *Sovranità e potere nell'era digitale*, in O. Pollicino-T.E. Frosini-E. Apa-M. Bassini (a cura di), *Diritti e libertà in internet*, Milano, 2017,19 ss.

anche e soprattutto per chiarire meglio i caratteri della regolazione che si va configurando ed i rischi, attuali e futuri, che tutto ciò tende a produrre. Infatti, tale tendenza, anche alla luce del *DMA* e del *DSA*[5], sembra delineare un mutamento del rapporto fra potere pubblico[6] e privato[7] caratterizzato da un maggior protagonismo del primo, e se ciò è da salutare con favore, in termini generali, è comunque opportuno sottolineare i rischi che ne potrebbero derivare, per evitare così la riedizione di situazioni del passato contrassegnate da una indiscussa (e, a tratti, ingiustificata) supremazia del primo (cioè del potere pubblico) sul secondo (ossia sul potere dei privati, non solo imprese ma, anche e soprattutto, cittadini).

L'autoritarietà del potere giuridico, da intendere come capacità di esso di incidere nella sfera giuridica altrui anche in assenza del relativo consenso (producendo, quindi, effetti giuridici di tipo costitutivo, modificativo ed estintivo, con un esito che, per il destinatario, può essere favorevole ovvero sfavorevole), ha rappresentato per decenni un carattere distintivo dei pubblici poteri, un dato che ne qualificava, cioè, la natura dei soggetti titolari (che, appunto, erano pubblici in quanto dotati del potere autoritativo) e che segnava la differenza di *status* con i soggetti privati, contribuendo in maniera decisa a ravvalorarne la posizione di supremazia di cui il potere autoritativo rappresentava, appunto, la manifestazione principale, con una condizione di specialità (che sovente si traduceva in una condizione di privilegio) rispetto a tutti gli altri soggetti dell'ordinamento.

L'idea del pubblico come soggetto "dominante" e del privato quale "dominato", un rapporto, cioè, contrassegnato da una disparità (di partenza), ossia da una posizione di supremazia del primo, è senz'altro ancora attuale ma si è negli ultimi decenni forte-

[5] Tali due regolamenti, assieme a quello sull'*Artificial Intelligence Act*, compongono quella che è stata definita in dottrina come «trilogia regolamentare dell'UE per l'ecosistema digitale», G. Gardini, *Le regole dell'informazione. Pluralismo e libertà nell'era dell'intelligenza artificiale*, VI ediz., Torino, 2024, 300 ss. Con particolare riferimento al *DMA*, in dottrina cfr. M. Libertini, *Il regolamento europeo sui mercati digitali e le norme generali in materia di concorrenza*, in *Rivista trimestrale di diritto pubblico*, 4, 2022, 1069 ss.; M. Orofino, *Il* Digital Market Act*: una regolazione asimmetrica a cavallo tra diritto della protezione dei dati e diritto antitrust*, in Aa.Vv., *La regolazione europea della società digitale*, Torino, 2024, 175 ss. Con specifico riferimento agli obblighi dei *gatekeeper* cfr. G. Afferni, *Gli obblighi dei* gatekeeper, in L. Bolognini-E. Pelino-M. Scialdone (a cura di), Digital Services Act *e* Digital Markets Act*. Definizioni e prime applicazioni dei nuovi regolamenti europei*, Milano, 2023, 315 ss. Sul *DSA* cfr. invece S. Del Gatto, *Il* Digital Services Act*: un'introduzione*, 724 ss.; G. Finocchiaro, *Responsabilità delle piattaforme e tutela dei consumatori*, 730 ss.; E. Longo*, Libertà di informazione e lotta alla disinformazione nel* Digital Services Act, 737 ss.; G. Sgueo, *L'architettura istituzionale del* Digital Services Act, 746 ss., tutti in *Giorn. dir. amm.*, 6, 2023; M. Orofino, *Il* Digital Service Act *tra continuità (solo apparente) ed innovazione*, in Aa.Vv., *La regolazione europea della società digitale*, cit., 134 ss. Avuto riguardo all'AIA, benché non oggetto di trattazione nella presente sede, cfr. A. Iannuzzi-F. Laviola, *I diritti fondamentali nella transizione digitale fra libertà e uguaglianza*, in *Diritto costituzionale. Rivista quadrimestrale*, 1, 2023, 9 ss.; A. Simoncini, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Rivista AIC*, 2, 2023, 1 ss.; M. Bassini, *Intelligenza Artificiale generativa: alcune questioni problematiche*, in questa *Rivista*, 2, 2023, 391 ss.; E. Longo, *La disciplina del «rischio digitale»*, Aa.Vv., *La regolazione europea della società digitale*, cit., 53 ss.

[6] Sul potere pubblico cfr. G. Di Gaspare, *Il potere nel diritto pubblico*, Padova, 1992. Più di recente cfr. A. Carbone, *Potere e situazioni soggettive nel diritto amministrativo, I, Situazioni giuridiche soggettive e modello procedurale di accertamento (Premesse allo studio dell'oggetto del processo amministrativo)*, Torino, 2020.

[7] Di recente, sul potere, tra pubblico e privato, riflette M. Ruotolo, *Il potere, tra pubblico e privato. Tracce per un dialogo tra civilisti e costituzionalisti*, in *Costituzionalismo.it*, 3, 2024, 46 ss. Sul punto cfr. anche R. Spagnuolo Vigorita, *Potere amministrativo, poteri e interessi privati*, in *CERIDAP*, 2, 2024, 133 ss.

mente ridimensionata[8], essendo ormai riferibile solo a determinati ambiti, e con ciò si sono proporzionalmente ridotte le "sacche" di privilegio, pur continuando a sussistere in talune circostanze[9]. Il problema di tale riduzione dell'autoritarietà del pubblico non si è tradotta – come era invece auspicato ed auspicabile – in una maggiore libertà dei cittadini ma è andata a tutto vantaggio di pochi soggetti privati. Su questo la scienza giuridica deve interrogarsi e sembra che – sia pur con non incolpevole ritardo – stia cominciando a farlo.

Con riferimento a quanto si è verificato negli ultimi decenni, invero, non sembra azzardato affermare che il potere (per come inteso sopra), nei fatti, abbia finito per non essere più elemento esclusivo dei pubblici poteri, essendo, ormai, divenuto strumento in mano anche a taluni soggetti privati, per esempio, appunto, ai colossi del digitale. Il mondo del digitale, in altre parole, costituisce l'esempio emblematico della formazione di un potere assai simile a quello pubblico, rispetto al quale si utilizza qui il termine di "autoritarietà" proprio per metterne in luce tale vicinanza rappresentata, in ultima analisi, dalla idoneità di questo potere di incidere nella sfera giuridica di un numero assai elevato di cittadini in assenza, di fatto, del loro consenso, oltre che nella capacità di produrre effetti nei confronti dei tradizionali poteri pubblici, potendo condizionarne le scelte sotto la minaccia del possibile ricatto realizzabile in una molteplicità di modi (in ragione, prevalentemente, delle elevatissime disponibilità economiche e digitali da tali soggetti generalmente possedute).

L'autoritarietà dei privati, dunque, consiste in questo: nella possibilità, cioè, di imporre ad altri soggetti, pubblici poteri compresi, le proprie scelte mediante lo (stra)potere, economico e digitale, di cui godono[10], con una differenza di non poco conto rispetto a quanto accadeva (e continua accadere, specialmente in altri settori) allorché erano i pubblici poteri ad imporre le proprie scelte. Invero, in questi casi, si trattava (e continua a trattarsi) di un'imposizione comunque motivata – in ultima analisi ed al di là di tutto – da un'esigenza di soddisfare bisogni e interessi riferibili ad una collettività, e non di perseguire il legittimo interesse egoistico all'accumulazione – spesso indiscriminata – di ricchezza ed appagare la "sete di controllo" (della società digitale[11]).

---

[8] Si pensi all'ambito del diritto amministrativo, un tempo contrassegnato dal potere autoritativo dell'amministrazione, tratto distintivo e caratteristico della tradizionale teorica del provvedimento (il c.d. potere di imperio). In tale ambito, tuttavia, si assiste, ormai da decenni, ad un processo di forte riduzione dei contenuti autoritativi del potere, incentivando sempre più strumenti "paritari" come gli accordi (art. 11, legge 7 agosto 1990, n. 241) e, più in generale, moduli negoziali, e anche laddove sussistono ancora strumenti autoritativi (tipo, appunto, il provvedimento) si tratta, comunque, di strumenti per così dire "democraticizzati", dove si ha un potere che, per ricorrere ad una figura dalla valenza meramente descrittiva, è dialogante con il privato, il quale ha visto rafforzare i suoi diritti partecipativi e, più in generale, le garanzie previste dalla legge sul procedimento. Inoltre, a seguito del codice del processo amministrativo del 2010, si è assistito anche ad un rafforzamento sul versante degli strumenti processuali, grazie anche al meccanismo di tendenziale atipicità della tutela.

[9] Il riferimento è, per esempio, ad un trattamento non sempre paritario fra amministrazione e cittadino allorché sussista fra di essi una controversia.

[10] È stato osservato, E. Cremona, *L'erompere dei poteri privati nei mercati digitali e le incertezze della regolazione antitrust,* in *Osservatorio sulle fonti*, 2, 2021, 881, che tali poteri privati «detengono posizioni di 'potere', nel senso dell'attitudine che essi hanno ad incidere unilateralmente – nell'ambiente digitale – sulla sfera giuridica dei soggetti che con essi entrano in contatto».

[11] Sulla società digitale, di recente, cfr. E. di Carpegna Brivio, *Pari dignità sociale e* Reputation scoring.

Il principale profilo problematico di quella che si è qui definito nei termini di "autoritarietà", con riferimento alle piattaforme digitali, è che essa non deriva dalla legge, bensì dalle dinamiche della realtà caratterizzate per decenni – come già visto – da una tendenziale assenza dell'etero-regolazione pubblica. Dove non c'è, quindi, il diritto – o, comunque, la sua presenza è assai limitata –, un ambito, cioè, sprovvisto di regolazione, si hanno dinamiche "spontanee" che sono però caratterizzate da differenti rapporti di forza, dove sono i più forti a prevalere (in tal caso i poteri privati digitali)[12]. L'idea della libertà come spazio dove regna l'autonomia dei soggetti è un'idea senz'altro condivisibile in teoria ma presenta evidenti criticità sul piano pratico perché spesso, è noto, si traduce – come è avvenuto nel caso di specie – nel dominio del "più forte" sul "più debole".

Nel mondo del digitale vi sono pochi "forti" (i colossi del digitale) e tanti "deboli", sia pur con diversa intensità, debolezza che opera tanto con riferimento ai pubblici poteri[13], quanto in relazione ai privati, imprese e consumatori[14], salvo non voler continuare ad assumere lo schema del passato, ormai tramontato, per cui sono i poteri pubblici i "forti" ed è dal loro potere che è sempre e comunque necessario difendersi[15]. In questo senso, si è dinanzi ad una autoritarietà dei privati che, non discendendo dalla legge, non può rinvenire in essa la propria fonte legittimante e, dunque, la sua legittimazione[16] democratica, richiamando così alla mente l'insegnamento di Max Weber per cui l'unico potere legittimo è quello pubblico, il quale, appunto – a differenza di quello privato – è democratico (o, almeno, tale dovrebbe essere)[17].

Il potere privato si caratterizza, in negativo, rispetto al potere pubblico, per una serie di elementi che in questa sede possono essere solamente oggetto di un rapido richiamo. Per esempio, il potere privato è un potere ontologicamente più difficilmente controllabile poiché tende a sfuggire dal sentiero tracciato dalla legge, a differenza di quello pubblico, rispetto al quale non solo vi è la legge che ne funzionalizza l'esercizio (verso

Per una lettura costituzionale della società digitale, Torino, 2024. In proposito cfr. anche A. Celotto, *"Sudditi". Diritti e cittadinanza nella società digitale*, Milano, 2023.

12 È risaputo che dove non c'è il diritto a prevalere è il potere del più forte, come noto ed anche fisiologico, e, non a caso, il diritto nasce anche per questo, per limitare il potere del più forte (si pensi al principio di legalità ed al ruolo che esso ha storicamente rivestito).

13 Dove ora sono questi ultimi al fondo della scala gerarchica, e non più al vertice.

14 Anch'essi sempre al fondo, perché non si può certo parlare di rapporti orizzontali fra questi soggetti per il sol fatto che appartengono entrambi all'emisfero privato.

15 Si tratta di uno schema teorico che senz'altro ben descriveva l'assetto del passato ma che ora necessita di essere ripensato poiché il dato empirico ha mostrato come vi siano poteri ben più forti di quelli privati, rispetto ai quali non è sufficiente affermarne l'illegittimità perché tanto essi continuano ad operare nella realtà fattuale, che è quella della quale il diritto (e con esso i giuristi) deve (dovrebbe) occuparsi, ripensando e ridefinendo gli schemi teorici allorché essi non siano più in grado di descrivere compiutamente la realtà sociale di quel determinato periodo storico.

16 Sul potere pubblico e privato nell'ottica della relativa legittimazione non si prescinda da A. Romano Tassone, *A proposito del potere pubblico e privato e della sua legittimazione*, in *Diritto amministrativo*, 4, 2013, 559 ss.

17 Ricorda O. Pollicino, *L'"autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *Federalismi.it*, 19, 2019, 12, che una premessa fondamentale del diritto pubblico è proprio quella per cui l'unico potere legittimo è il potere pubblico.

il perseguimento dell'interesse pubblico) ma sussiste, ormai, un apparato costituzionale che, rafforzandone i limiti, aumenta così le garanzie dei destinatari. Il tema del controllo è, poi, strettamente legato al profilo concernente la trasparenza per le stesse ragioni, di fatto, per cui è di più difficile controllo; non è un caso che in dottrina, con riferimento ai poteri privati, in specie quelli digitali, si sia di recente parlato di "poteri opachi"[18]. Inoltre, vi è un ulteriore elemento che ne segna in maniera evidente la differenza dal potere pubblico, ossia il profilo della sua giustiziabilità. Mentre dinanzi alle decisioni dei pubblici poteri sono esperibili rimedi tanto stragiudiziali (si pensi alle autorità indipendenti e, più in generale, agli ulteriori meccanismi di risoluzione stragiudiziale delle controversie previsti dalla legge, come le conciliazioni) quanto rimedi di ordine giudiziale, rispetto alle decisioni assunte dai poteri privati spesso le piattaforme digitali creano "consessi", a loro evidentemente favorevoli, dove (fingere di) fare giustizia (si pensi, giusto a titolo di esempio, al *Facebook Oversight Board*).

Lo schema tradizionale del diritto pubblico, dunque, quello cioè imperniato sul conflitto fra "autorità" e "libertà", deve essere rivisto in quanto l'autorità non si identifica più solo nei pubblici poteri ma anche ed in taluni casi (come, per esempio, nel caso in esame) nei poteri privati[19], dando luogo ad un autorità a tratti ben più pericolosa di quella del passato[20] (per le solite ragioni di cui sopra: strapotere finanziario, difficoltà di controllo, trasparenza, ecc.), e ciò che un tempo era il nemico da cui proteggersi (i pubblici poteri) parrebbe ora rappresentare l'amico a cui aggrapparsi, lo "strumento" a cui riattribuire poteri per configurare uno spazio caratterizzato da un riequilibrio nei rapporti di forza, dove lo strapotere privato dei potenti gruppi tecnologici incontra il baluardo insuperabile della sovranità popolare rappresentato, anzitutto, dallo Stato e dalla sua sovranità (la sovranità statale) che dalla prima e nella prima rinviene la propria legittimazione democratica.

Si assiste ormai, come noto, ad un ad un governo privato del cyberspazio rispetto al quale risulta assai condivisibile l'osservazione secondo cui «la libertà di espressione e le libertà economiche [sono] minacciate – quanto meno nelle moderne democrazie occidentali – non tanto dai poteri pubblici tradizionalmente considerati, ma dalle posizioni oligopolistiche di potere privato delle grandi società dell'economia digitale, la cui neutralità è un mito che è stato definitivamente superato»[21]. Tuttavia, il ciberspazio e, dunque, il mondo digitale sono, a tutti gli effetti, fenomeni sociali al pari di ogni altro e, di conseguenza – in special modo allorché in esso assumano rilievo diritti e libertà –, necessitano di una qualche regolazione pubblica al fine, in ultima analisi, di garantire il rispetto e l'effettività di quegli stessi diritti e di quelle medesime libertà.

Ciò detto, è da registrare, come già anticipato, una parziale inversione di rotta – ancora in pieno corso – tesa a ridimensionare il potere privato nel settore del digitale, con una

[18] Il riferimento è al lavoro di M.R. Ferrarese, *Poteri nuovi*, cit., *passim*.

[19] Sul punto cfr. M. Betzu, *I baroni del digitale*, cit., *passim*.

[20] Sul rapporto fra potere privato e diritti fondamentali, anche in prospettiva evolutiva, cfr. il lavoro monografico di G. Lombardi, *Potere privato e diritti fondamentali*, Torino, 1970.

[21] Così M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *La Rivista del Gruppo di Pisa*, 2, 2021, 172. Sul punto cfr. anche M. Cuniberti, *Potere e libertà nella rete*, in questa *Rivista*, 3, 2018, 51 ss., in chiave critica rispetto, in particolare, alla presunta neutralità dei *provider*.

(ri)comparsa di quello pubblico. Si tratta di una prospettiva assolutamente necessaria che, tuttavia, non deve, però, tradursi in una restrizione delle libertà, come, per esempio, in una lesione della libertà di manifestazione del pensiero, la quale rappresenta una delle libertà che più rischiano di essere messe in discussione da questo processo di riequilibrio dei rapporti di forza, come si vedrà più avanti.

3. La (auto)regolazione digitale di tipo tradizionale

Il profilo della regolazione del mondo digitale ha, come noto, attraversato una serie di fasi diverse fra loro caratterizzate da un differente approccio del regolatore pubblico – differenza, in larga parte, derivante anche dal contesto in cui tale regolazione si collocava – sulla base di due approcci fra loro concettualmente distinti: quello statunitense e quello europeo.

Inizialmente, il "digitale" e, quindi, il fenomeno di Internet, erano considerati come ambiti in grado di autoregolarsi; anzi, l'intervento regolatorio dei pubblici poteri avrebbe potuto significare una riduzione, financo una compressione, delle libertà economiche dei relativi operatori e, soprattutto, della libertà degli utenti. Il mondo digitale doveva essere quanto più libero possibile, anche e soprattutto dal diritto[22], dove il diritto non arrivava – se non eccezionalmente e, dunque, in funzione meramente sussidiaria – e dove i singoli operatori si sarebbero dotati delle regole necessarie allo svolgimento della loro attività, sul presupposto (anche) del fatto che fossero loro i conoscitori della "materia", a differenza dei pubblici poteri, sprovvisti delle conoscenze tecnico-scientifiche necessarie a regolare fenomeni così complessi e, in particolare, nuovi. Ciò avrebbe altresì significato, dal lato degli utenti, un ampliamento della loro libertà, anzitutto di quella di manifestazione del pensiero: il mondo digitale come ampliamento della sfera di libertà di operatori e utenti e, di conseguenza, non sussisteva alcun motivo, per i pubblici poteri, di intervenire. Insomma, un mondo praticamente perfetto, dove regnavano autonomia (da…) e libertà (di…) e dove Internet era considerato quale «mezzo anarchico per natura, che la società degli uomini fatica a imbrigliare all'interno di una regolazione giuridica»[23]; una rete, dunque, caratterizzata da una «genetica vocazione libertaria» nonché diffidente dalle «forme di regolazione di tipo istituzionale»[24].

L'idea della necessità di un mondo digitale libero, da difendere dal potere (anzitutto pubblico), si inseriva in una fase storica contrassegnata dalla volontà di configurare uno spazio di libertà non sottoposto a restrizioni da parte degli stessi poteri pubblici e/o da altre autorità; di conseguenza, si tentava, anzitutto, di difendersi da una possibile ingerenza ad opera del legislatore. Inoltre, in quella medesima fase, soffiava forte

[22] M. Betzu, *Poteri pubblici e poteri privati*, cit., 166.

[23] Aa.Vv., *Il futuro del diritto pubblico. Il tempo e le sfide*, in *Diritto pubblico*, 1, 2024, 99 ss. Sul punto cfr. anche P. Costanzo, *La democrazia digitale (precauzioni per l'uso)*, in *Diritto pubblico*, 1, 2019, 80, il quale parla di Internet quale ambito «refrattario a interventi regolatori».

[24] P. Costanzo, *La democrazia digitale*, cit., 86. Sul rapporto fra Internet e democrazia cfr. A. Randazzo, *Internet e democrazia: prime note su tre possibili svolgimenti di un rapporto complesso*, in *Consulta online*-Liber amicorum *per Pasquale Costanzo*, 2020, 1 ss.

il vento della retorica dei diritti (a discapito dei doveri) e, di conseguenza, l'ottica era quella di configurare un sistema di diritti e libertà intangibili per il potere pubblico che rappresentasse, cioè, una sorta di "carta dei diritti"[25]. Il nemico, per intendersi, era rappresentato dai pubblici poteri e dal nemico, si sa, è necessario difendersi, senza considerare che, in quella stessa fase – con riferimento al contesto europeo –, pur resistendo ancora retaggi non irrilevanti della visione organicistica[26], essa andava comunque lentamente sgretolandosi, assumendo, viceversa, sempre più vigore la persona come *uti singulus*, coerentemente con l'impostazione antropocentrica delle Carte costituzionali che, dopo decenni di non sempre piena attuazione sul fronte delle libertà, potevano ora vedere (o, almeno, questa era la speranza) la conquista di uno spazio di libertà pressoché assoluto, al riparo da ogni tipo di condizionamento esterno.

In una situazione come questa, quindi, era (e tende a continuare ad essere) il mercato che precede il regolatore[27], il quale non si limita a ciò poiché lo tiene anche "alla distanza", se così si può dire, intravedendo in quest'ultimo un'"entità" di cui non si ha necessità alcuna, considerazione questa condivisa, nei fatti, anche dallo stesso legislatore che non manifestava certo particolare interesse nell'intervenire nella regolamentazione del settore in esame.

Tuttavia, come ormai noto e ampiamente dimostrato dalle condotte dei vari legislatori e, quindi, dai diversi atti normativi adottati negli ultimi anni, il mito dell'auto-regolazione è inesorabilmente naufragato e più nessuno ripete il mantra della sufficienza della sola regolamentazione privata[28], con buona pace, dunque, dell'idea stessa delle capacità benefiche, quasi "salvifiche", dell'auto-regolazione. Questo approccio, almeno con riferimento al contesto europeo, costituiva una peculiarità in quanto si presentava come eccezione all'idea della necessità di regolare il mercato, assunto che riposa sull'esigenza di una regolazione pubblica per assicurare il gioco della concorrenza e creare così il mercato (potenzialmente) perfetto, quello, cioè, concorrenziale, il quale, così organizzato, assicurerebbe di conseguenza il benessere sociale, coerentemente con l'idea, di marca ordoliberale, riassunta nella nota formula dell'economia sociale di mercato fortemente competitiva (art. 3, par. 3, TUE) che è a fondamento della stessa costruzione dell'ordinamento europeo.

Questo approccio di tendenziale indifferenza nei confronti del mondo digitale e, in particolare, del relativo mercato, ha condotto ad una situazione caratterizzata dalla rapida formazione prima ed ascesa poi di veri e propri colossi del digitale, pochi ma assai potenti, in grado di condizionare arbitrariamente l'intero mercato, così come capaci di incidere unilateralmente sui pubblici poteri nonché sugli utenti-consumatori.

Questi soggetti hanno, dunque, finito per godere di quella che – da più parti e ormai

[25] In proposito, per tutti, cfr. S. Rodotà, *Una Costituzione per Internet?*, in *Politica del diritto*, 3, 2010, 342.

[26] Cfr. A. Orsi Battaglini, «*L'astratta e infeconda idea*». *Disavventure dell'individuo nella cultura giuspubblicistica (A proposito di tre libri di storia del pensiero giuridico)*, in *Quaderni fiorentini per la storia del pensiero giuridico moderno*, 17, 1988, 569 ss., ora anche in *La necessaria discontinuità. Immagini del diritto pubblico* (Quaderni di San Martino), Bologna, 1990, 11 ss., nonché in Id., *Scritti giuridici*, Milano, 2007, 1309 ss. Con particolare attenzione alla giuspubblicistica tedesca M. Fioravanti, *Giuristi e costituzione politica nell'ottocento tedesco*, Milano, 1979.

[27] E. Cremona, *L'erompere dei poteri privati nei mercati digitali*, cit., 880 ss.

[28] Sul punto cfr. M. Manetti, *Regolare Internet*, in questa *Rivista*, 2, 2020, 36.

comunemente – si definisce nei termini di "sovranità digitale", una sorta, cioè, di condizione di supremazia nel mondo digitale che porta questi stessi soggetti ad essere i decisori, di fatto indiscussi, delle sorti stesse del mercato e, più in generale, della vita digitale di ciascun individuo.

L'abnorme concentrazione di potere (privato) nelle mani di pochi soggetti che si è così determinata ha reso, quindi, evidente ed ineludibile l'esigenza di apprestare un sistema (sia di regole che di limiti) per far fronte ad un potere, quello privato, munito della capacità di rendersi indipendente ed autonomo da ogni qualsivoglia tipologia di potere, a cominciare da quello pubblico e, per questo, potenzialmente assai pericoloso, anzitutto e soprattutto avuto riguardo ai diritti ed alle libertà fondamentali che, a vario titolo, vengono in rilievo nel mondo digitale e che ben possono essere da esso messe in discussione.

La sovranità digitale di questi poteri fa sì che essi tendano ad assumere «i tratti tipici di un ordinamento giuridico, autonomo dall'ordinamento generale»[29] ed il rischio che si cela dietro tale fenomeno è quello per cui «siano proprio le piattaforme, dopo aver invocato la libertà di espressione come giustificazione per l'assenza di eteroregolazione, ad assumere misure che possono incidere significativamente sulle libertà fondamentali, senza che siano previsti rimedi o correttivi idonei»[30].

Ovviamente, non si può omettere di ricordare i numerosi benefici che, in termini generali, la diffusione del *web* ha comportato per l'intera collettività; invero, la rapida e costante crescita di quest'ultimo «come strumento di comunicazione ha determinato, in senso positivo, un ampliamento degli spazi entro cui gli individui svolgono la propria personalità»[31]. Tuttavia, esso, come noto, «ha altresì ha parallelamente allargato, in senso negativo, la sfera dei comportamenti lesivi delle libertà altrui»[32] ed è per questo che è necessario un intervento dei pubblici poteri in funzione tanto regolamentare quanto regolatoria.

La priorità dei pubblici poteri, dunque – come è stato anche di recente osservato –, (dovrebbe) consiste(re) nel tentativo di «rompere i monopoli dei grandi gatekeepers e assicurare una regolamentazione pubblicistica delle nuove tecnologie digitali che sia in grado di contenere l'espansione illimitata di gruppi tecnologici privati a cui ha condotto il laissez-faire americano»[33].

Per quanto concerne, quindi, il rapporto fra auto-regolazione ed etero-regolazione, si può rilevare che l'auto-regolazione non si limita solamente ad arrivare sempre prima della seconda ma è onnipresente in quanto «consustanziale all'ideazione e alla fabbricazione di un prodotto o di un servizio»[34]. Di conseguenza, la regolazione pubblica «può solo inseguire ma scontando non solo il deficit iniziale di conoscenze specialisti-

[29] L. Torchia, *I poteri di vigilanza, controllo e sanzionatori nella regolazione europea della trasformazione digitale*, in *Rivista trimestrale di diritto pubblico*, 4, 2022, 1104.

[30] *Ibid.*

[31] Aa.Vv., *Il futuro del diritto pubblico*, cit., 101.

[32] *Ibid.*

[33] Ivi, 104 ss.

[34] A. Iannuzzi, *Paradigmi normativi per la disciplina della tecnologia: auto-regolazione, co-regolazione ed etero-regolazione*, in *Bilancio Comunità e Persona*, 2, 2023, 98.

che quanto, in realtà, un'asimmetria continua di informazioni e di saperi tecnici per via anche dei frequenti aggiornamenti che l'utilizzo di questi strumenti comporta»[35]. Ciò detto, rimane da capire se l'etero-regolazione sia in grado di recuperare tale svantaggio poiché, pur arrivando dopo, potrebbe, comunque, servirsi delle conoscenze tecniche dell'auto-regolazione e imporsi nei suoi confronti. Il fatto, poi, che quella regola tecnica (pro)venga originariamente da un soggetto privato, mettendo così in discussione l'eteronomia della fonte regolatoria, non rappresenta certo un problema insormontabile; invero, nel momento in cui interviene la regolazione pubblica si sopperisce a quel possibile aspetto problematico della derivazione privatistica della norma ed il *vulnus* alla tenuta del sistema democratico (che si fonda su una norma posta in essere dai pubblici poteri) è, così, agilmente risolto.

Chiarita l'esigenza di una regolazione pubblica, è opportuno mettere in luce rapidamente le ragioni del perché la regolazione antitrust, da sola, non sembra rispondere a tale esigenza[36]. Infatti, essa si basa sulla teoria del prezzo ma, in tali casi, essendo assente il prezzo – dato che la maggior parte delle piattaforme digitali non prevede né un costo di accesso né un costo circa il relativo utilizzo –, è evidente che la sua applicazione ponga più di un problema; detto altrimenti, non essendovi il prezzo viene quindi meno il presupposto stesso alla base del paradigma su cui si fonda la tradizionale regolazione di carattere antimonopolistico. Ecco, dunque, che si pone, anzitutto, la questione circa l'applicabilità – e, eventualmente, anche il profilo dell'effettività – di una regolazione di questo tipo. Il tema che allora si pone attiene all'esigenza di rivedere il concetto di concorrenza ed il suo contenuto[37], abbracciando l'idea di una regolazione concorrenziale che non si esaurisca nell'obiettivo di assicurare il "mercato perfetto" ma che, viceversa, contempli altri ed ulteriori interessi, come, ad esempio, quello relativo alla *privacy*, come già osservato in dottrina[38], ovvero quello alla libertà di manifestazione del pensiero, su cui si tornerà più diffusamente nel prosieguo.

4. I due regolamenti del 2022: verso un nuovo modello di regolazione

Il legislatore europeo con l'intervento del 2022 in materia di mercati digitali (c.d. *Digi-*

35 *Ibid.*

36 Sul punto, fra i tanti, cfr. G. Vettori, *Sui poteri privati. Interazioni e contaminazioni*, in *Diritto pubblico*, 3, 2022, 829 e ss., oltre che M. Betzu, *Poteri pubblici e poteri privati*, cit., 166 ss., e E. Cremona, *L'erompere dei poteri privati nei mercati digitali*, cit., 880 ss.

37 Sui cui cfr. almeno L. Buffoni, *La "tutela della concorrenza" dopo la riforma del Titolo V: il fondamento costituzionale ed il riparto di competenze legislative*, in *Istituzioni del federalismo*, 2, 2003, 348 ss.; M. Libertini, *La tutela della concorrenza nella Costituzione italiana*, in *Giurisprudenza costituzionale,* 2, 2005, 1429 ss.; Id., voce *Concorrenza*, in *Enc. dir.*, Ann. III, 2010, 195 ss.; F. Trimarchi Banfi, *Il «principio di concorrenza»: proprietà e fondamento*, in *Diritto amministrativo*, 1-2, 2013, 15 ss.; M. Manetti*, I fondamenti costituzionali della concorrenza*, in *Quaderni costituzionali*, 2, 2019, 315 ss.; M. Ramajoli, *Concorrenza (tutela della)*, in *Funzioni amministrative*, diretto da B. Mattarella, M. Ramajoli, *I tematici dell'Enciclopedia del diritto*, vol. III, cit., 292 ss.; C. Buzzacchi, *Tutela della concorrenza*, in L. Cuocolo, E. Mostacci (a cura di), *Il riparto di competenze tra Stato e Regioni. Vent'anni di giurisprudenza costituzionale sul Titolo V*, Pisa, 2023, 53 ss.

38 M. Betzu, *Poteri pubblici e poteri privati*, cit., 184 ss.

tal Markets Act, da ora *DMA*)[39], muovendo dal presupposto per cui la tendenza all'auto-regolamentazione delle piattaforme digitali non abbia portato alla creazione di un mercato concorrenziale, quanto esattamente al suo opposto, ha deciso di intervenire per innestare la concorrenza in un mercato dove si può affermare che, di fatto, ve ne fosse ben poca traccia. Anche il regolamento sui servizi digitali (*Digital Services Act*, *DSA*), di poco successivo[40], pur avendo altro oggetto, si pone in linea di continuità con quello appena menzionato ed entrambi sembrano offrire significativi spunti di riflessione in ordine al modello regolatorio delle piattaforme digitali che appare andarsi configurando. Di conseguenza, saranno ambedue oggetto di esame, con particolare attenzione agli aspetti più rilevanti ai fini del presente contributo.

Il duplice intervento normativo così realizzatosi segna l'abbandono – con non poco ritardo – della precedente disciplina, quella, cioè, dei primi anni del secolo in materia di commercio elettronico[41], non più evidentemente attuale, essendo stata pensata per tutt'altro contesto, assai diverso da quello odierno caratterizzato da una ramificazione ben più estesa delle tecnologie digitali nonché da un loro più sofisticato e sviluppato avanzamento scientifico.

Per quanto concerne il primo, il *DMA*, esso si inserisce in un quadro regolatorio destinato a mutare a livello europeo anche per effetto di taluni interventi nazionali ad esso precedenti, come, ad esempio, quello del Parlamento tedesco che, agli inizi del 2021, ha adottato il decimo emendamento al *Gesetz gegen Wettbewerbsbeschrankungen* (GWB), in base al quale si è attribuito al *Bundeskartellamt* il potere di proibire una serie di differenti condotte poste in essere da quelle imprese più rilevanti nel mercato digitale, senza dover previamente rinvenire una violazione, da parte loro, dell'apparato normativo a difesa della concorrenza[42] o, ancora, quello dell'autorità inglese (*Competition and Market Autority*) che, nell'aprile dello stesso anno, ha deciso di istituire una *Digital Market Unit* (*DMU*) al fine di attuare un nuovo sistema di regolazione pensato per le piattaforme digitali dotate del c.d. *strategic market status*.

Gli obiettivi di tale intervento possono essere sintetizzati, da un lato, nella volontà di assicurare e, prima ancora, promuovere, condizioni di equità all'interno dei mercati digitali ove operano i c.d. *gatekeeper* e, dall'altro, assicurare in questi stessi mercati la loro contendibilità (messa a repentaglio dagli stessi *gatekeeper*). Infatti, in relazione al primo obiettivo, il regolamento prende in considerazione la circostanza (assai probabile) per cui, in ragione di una serie di loro caratteristiche[43], esse arrechino gravi squilibri

[39] Regolamento (UE) 2022/1925.

[40] Regolamento (UE) 2022/2065.

[41] Il riferimento è, ovviamente, alla direttiva (CE) 2000/31, benché il *DSA* affermi espressamente che il «presente regolamento non pregiudica l'applicazione della direttiva 2000/31/CE» (art. 2, par. 3).

[42] Art. 19(a) GWB.

[43] Come, per esempio, il comune denominatore rappresentato dal considerevole potere economico di cui godono che consente a talune piattaforme di avere la capacità di connettere molti utenti commerciali con altrettanti utenti finali mediante i loro servizi, con la conseguenza, così, di poter sfruttare i vantaggi acquisiti in un settore di attività (quali l'accesso a grandi quantità di dati) in un altro settore, ovvero il fatto che esse esercitano un controllo su interi ecosistemi, causando, con ciò, una notevole difficoltà, a livello competitivo, per gli operatori di mercato esistenti o nuovi, indipendentemente dal livello di innovazione o efficienza di tali operatori. Si veda in proposito il punto 3 dei *Considerando*.

a livello anzitutto di potere negoziale, con ciò che ne deriva in punto di pratiche sleali nonché condizioni inique, sia rispetto ai c.d. utenti commerciali che avuto riguardo agli utenti finali dei servizi di piattaforma somministrati dai *gatekeeper*. Tutto ciò, è evidente, si traduce, per esempio, in un potenziale aumento dei prezzi, così come in un possibile deterioramento della qualità dei relativi servizi[44].

Questa situazione, osserva il legislatore, fa sì che, sovente, i processi di mercato non siano in concreto in grado di garantire risultati economici equi rispetto ai servizi di piattaforma di base. Nonostante le previsioni di cui agli articoli 101 e 102 del TFUE si applichino, in teoria, anche al comportamento dei *gatekeeper*, il loro ambito di operatività è, comunque, limitato a talune tipologie di c.d. potere di mercato (come, a titolo di esempio, una posizione dominante in mercati specifici) e di comportamento ritenuto contrario al principio concorrenziale e, soprattutto, la loro applicazione avviene *ex post*, oltre a richiedere un'indagine approfondita, caso per caso, di fatti che, spesso, sono molto complessi da provare in concreto. In aggiunta, si deve considerare che il diritto vigente dell'UE non affronta – o, comunque, lo fa in maniera poco efficace – i problemi concernenti l'efficiente funzionamento del mercato interno, problemi che sono, per lo più, imputabili al comportamento dei *gatekeeper*, la cui posizione, ai fini del diritto della concorrenza, non sempre rileva in termini di posizione dominante.

Per quanto attiene, al contrario, al secondo obiettivo, il legislatore ricorda innanzitutto l'impatto significativo dei *gatekeeper* sul mercato in quanto forniscono i c.d. *gateway* (ossia punti di accesso) ad un numero assai elevato di utenti commerciali al fine di raggiungere così gli utenti finali in tutta l'UE. In ragione sia delle pratiche sleali sempre più frequenti nonché della assai scarsa contendibilità dei servizi di piattaforma di base parte dei legislatori nazionali, come noto, l'UE e, quindi, i singoli Stati, si sono trovati costretti ad intervenire; del resto, tali pratiche causavano problemi al funzionamento del mercato, determinando una serie di ripercussioni negative, tanto a livello economico quanto sociale. Gli interventi normativi a livello statale così realizzati, tuttavia, hanno contribuito a frammentare il mercato unico, accrescendo così il rischio di un aumento generalizzato dei costi in ragione dei differenti requisiti richiesti a livello nazionale ai vari *gatekeeper*[45].

Pertanto, si pone l'esigenza di operare un riavvicinamento delle diverse legislazioni; solamente in questo modo sarà possibile rimuovere gli ostacoli che impediscono una condizione di libertà sotto il profilo della fornitura dei servizi digitali e avere di conseguenza la possibilità di poterne usufruire con maggiore facilità. Dunque, il legislatore intende configurare, a livello europeo, taluni obblighi al fine di garantire che i mercati digitali siano effettivamente equi e contendibili, dove la presenza dei *gateekeeper* non è ostacolata ma, diversamente, configurata in maniera tale da risultare vantaggiosa per l'intera economia europea e, quindi, per gli stessi consumatori finali[46]. Sotto questo profilo, è coerente la previsione per cui «gli Stati membri non impongono ulteriori obblighi ai *gatekeeper* per mezzo di leggi, regolamenti o misure amministrative allo scopo

[44] Punto 4 dei *Considerando*.

[45] Punto 6 dei *Considerando*.

[46] Punto 8 dei *Considerando*.

di garantire l'equità e la contendibilità dei mercati»[47].

In relazione alla questione concernente il tipo di regolazione che il legislatore ha voluto adottare attraverso il presente regolamento, è opportuno sottolineare che con tale disciplina si intende integrare il quadro normativo in materia di concorrenza. Invero, l'introduzione di questa normativa non dovrebbe comportare il venir meno dell'applicazione (una sorta, cioè, di abrogazione implicita) delle disposizioni di cui agli articoli 101 e 102 TFUE, così come, allo stesso modo, delle relative discipline a livello nazionale in materia di concorrenza nonché delle altre norme (sempre a livello nazionale) in materia di concorrenza che prendono in considerazione quei comportamenti unilaterali basati su una valutazione, caso per caso, delle posizioni nonché dei comportamenti di mercato, oltre che quelle relative al controllo delle concentrazioni[48]. Pur non volendo superare la tradizionale disciplina concorrenziale, si tratta, in ogni caso, di un intervento che non può essere inquadrato al suo interno, nel senso, cioè, che si è in presenza di una regolazione apposita per tale mercato e che, come noto, si rivolge solo ad alcuni determinati soggetti che in tale mercato operano (e di cui sono, di fatto, i padroni indiscussi). Non a caso, si afferma espressamente che, se con gli articoli 101 e 102 TFUE e con le altre normative nazionali in materia l'obiettivo è quello di assicurare «la protezione della concorrenza non falsata sul mercato»[49], con il presente regolamento, al contrario, si vuole perseguire un obiettivo che è sì complementare ma che è, comunque, distinto dalla «protezione della concorrenza non falsata su un dato mercato, quale definita in termini di diritto della concorrenza»[50]. Tale obiettivo, prosegue il legislatore, «consiste nel garantire che i mercati in cui sono presenti gatekeeper siano e rimangano equi e contendibili, indipendentemente dagli effetti reali, potenziali o presunti sulla concorrenza in un dato mercato derivanti dal comportamento di un dato gatekeeper contemplato dal presente regolamento»[51]. Di conseguenza, il regolamento in esame si preoccupa di proteggere un interesse giuridico differente rispetto a quello protetto dalle disposizioni che compongono l'apparato tradizionale della normativa concorrenziale e, dunque, «dovrebbe applicarsi senza pregiudicare l'applicazione di queste ultime»[52].

Si tratta di un intervento che attribuisce un ruolo centrale e di particolare rilievo alla Commissione, tanto nell'individuazione dei *gatekeeper* quanto con riferimento agli obblighi alla cui osservanza tali soggetti sono tenuti. Del resto, le caratteristiche che determinano la qualificazione in termini di *gatekeeper*, per come configurate, lasciano inevitabilmente un apprezzabile margine di valutazione alla Commissione circa l'attribuzione o meno di suddetta qualifica. Ai fini dell'applicazione del regolamento in esame si fa, infatti, riferimento a soggetti che hanno «un impatto significativo sul mercato interno» ovvero che forniscono «un servizio di piattaforma di base che costituisce un punto di accesso (gateway) importante affinché gli utenti commerciali raggiungano gli

47 Art. 1, par. 5.

48 Cfr. punto 10 dei *Considerando*.

49 Cfr. punto 11 dei *Considerando*.

50 *Ibid.*

51 *Ibid.*

52 *Ibid.*

utenti finali» e, infine, che detengono «una posizione consolidata e duratura, nell'ambito delle proprie attività, o è prevedibile che acquisisca siffatta posizione nel prossimo futuro»[53]. È evidente come queste caratteristiche si prestino a diverse interpretazioni e gli indici presuntivi ivi previsti, pur riducendo senza dubbio tale ampio margine di apprezzamento, non sono certo in grado di annullarlo o, comunque, di ridurlo significativamente.

Per quanto riguarda, invece, gli obblighi imposti ai *gatekeeper*, essi si possono dividere in due categorie: una prima che è composta da obblighi già compiutamente definiti (che costituiscono, cioè, una sorta di *numerus clausus*) e che non necessitano di alcuna specificazione da parte della Commissione (art. 5), e quelli che, viceversa, sono suscettibili un intervento ad opera di quest'ultima, rispetto ai quali il potere della stessa Commissione è evidentemente maggiore (art. 6).

In relazione alla violazione di tali obblighi, con l'aggiunta di quello specifico relativo all'interoperabilità dei servizi di comunicazione interpersonale indipendenti dal numero (art. 7), si prevede che la Commissione – a seguito di un'indagine di mercato della durata massima di 12 mesi – possa adottare un atto di esecuzione che imponga a tali *gatekeeper* «qualsiasi rimedio comportamentale o strutturale proporzionato e necessario per garantire l'effettivo rispetto del presente regolamento»[54]. Si tratta di un potere configurato in termini particolarmente generici che, probabilmente, consentirà l'assunzione, da parte della Commissione, di una serie di misure fra loro fortemente eterogenee. Sicuramente si è dinanzi ad un potere sanzionatorio che offre una pluralità – potenzialmente assai numerosa – di possibili misure da adottare e che, come tale, è altresì caratterizzato da una evidente dose di flessibilità.

Da tale regolamento sembra, quindi, profilarsi un nuovo modello regolatorio che si basa su tre differenti elementi: *i*) un ruolo centrale della Commissione, *ii*) la predeterminazione delle misure e *iii*) un ampio margine di libertà, in capo alla stessa Commissione – sotto il profilo dell'individuazione dei *gatekeeper* – della specificazione degli obblighi e, da ultimo, delle misure sanzionatorie.

Tutto ciò fa sì che il modello regolatorio che appare emergere possa essere qualificato nei termini di una regolazione c.d. *ex ante*[55] – a differenza di quella tradizionale antitrust che, all'opposto, si fonda su misure *ex post* – e, soprattutto, flessibile, come si avrà modo di osservare meglio più avanti.

L'altro regolamento, il *DSA*, si pone in linea di continuità con il *DMA*. Invero, attraverso tale intervento il legislatore mira ad apprestare una regolamentazione dei servizi digitali, prevedendo una serie di obblighi al fine di tutelare e migliorare il funzionamento del mercato interno, dettando le condizioni necessarie per lo sviluppo nonché l'espansione di servizi digitali innovativi nel mercato interno. Ad avviso del legislatore, uniformare le singole misure nazionali di regolamentazione in materia di obblighi per i prestatori di servizi intermediari costituisce un punto essenziale per risolvere ed evita-

[53] Art. 3.

[54] Art. 18, par. 1.

[55] Per l'idea di una regolazione *ex ante* nei mercati digitali, cfr. A.P. Massaro, *The Rising Market Power Issue and the Need to Regulate Competition: a Comparative Perspective between the European Union, Germany, and Italy*, in *Concorrenza e Mercato*, 2022, 13 ss.

re la frammentazione del mercato interno e garantire la certezza del diritto, riducendo, così, l'incertezza per gli sviluppatori, da un lato, e promuovendo l'interoperabilità, dall'altro. Inoltre, il fatto di ricorrere a prescrizioni tecnologicamente di carattere neutro dovrebbe determinare uno stimolo in punto di innovazione[56].

È, infatti, fondamentale un intervento armonizzante a livello europeo allo scopo di offrire alle imprese la possibilità di fare ingresso in nuovi mercati ed avere l'opportunità di sfruttare i vantaggi derivanti dallo stesso mercato interno, consentendo, al contempo, ai consumatori ed agli altri destinatari dei servizi la possibilità di disporre di una scelta più vasta[57].

Attraverso tale regolamento si tenta di assicurare un comportamento, da parte dei prestatori di servizi intermediari, improntato alla responsabilità ed alla diligenza, comportamento ritenuto dal legislatore essenziale per poter godere di un ambiente *online* sicuro, che sia prevedibile ed affidabile e, inoltre, per poter consentire ai cittadini di esercitare i loro diritti garantiti dalla Carta dei diritti fondamentali dell'Unione europea, a partire dalla libertà di espressione e di informazione[58], così come la libertà di impresa ed il diritto alla non discriminazione, nella prospettiva di conseguire un livello elevato di protezione degli stessi consumatori[59].

Ciò osservato, preme, ad ogni modo, sottolineare che il *DSA* non offra, in realtà, particolari indicazioni circa il modello di regolazione prescelto da legislatore europeo, occupandosi, infatti, in particolare, di questioni procedurali più che sostanziali. Ciò non significa, tuttavia, che sia del tutto assente una disciplina regolatoria, avendo questa – o, comunque, potendo avere – una marcata valenza procedurale[60]; al contrario, significa negare, da questo punto di vista, l'opzione, da parte del legislatore, per un modello specifico di regolazione.

Pertanto, sembra possibile affermare – avuto particolare riguardo al *DMA*, per le ragioni già dette – che l'impianto regolatorio che ne deriva sia il frutto, per così dire, di una combinazione di diversi modelli regolatori, ascrivibili tanto alla classica regolazione pro-concorrenziale quanto ad un tipo di regolazione che risponde ai caratteri della regolazione *ex ante*, ossia una regolazione che ha una funzione preventiva, che tenta, cioè, di guidare l'oggetto da essa regolato, provando, per quanto possibile, ad anticipare fatti e condotte, anziché essere da essi travolta e trovarsi, quindi, sempre in ritardo rispetto all'emersione di nuovi fenomeni o, comunque, dei mutamenti ad essi

[56] Cfr. punto 4 dei *Considerando*.

[57] Punto 2 dei *Considerando*.

[58] Per un inquadramento generale e recente sul tema cfr. Aa.Vv., *Diritto dell'informazione e dei media*, II ediz.,Torino, 2022; P. Caretti-A. Cardone, *Il diritto dell'informazione e della comunicazione nell'era dell'intelligenza artificiale*, II ediz., Bologna, 2024; G. Gardini, *Le regole dell'informazione*, cit.

[59] Punto 3 dei *Considerando*.

[60] Cfr. in proposito E. Chiti, *La disciplina procedurale della regolazione*, in *Rivista trimestrale di diritto pubblico*, 3, 2004, 679 ss., ove si sottolinea altresì la rilevanza del principio di proporzionalità (su cui cfr. anche G. Napolitano, *Servizi pubblici e rapporti di utenza*, Padova, 2001, 664 ss.). Più in generale, sul rapporto fra regolazione e mercato, cfr. M. Antonioli, *Mercato e regolazione*, Milano, 2001, 75 ss.; A. Zito, voce *Mercati (regolazione dei)*, in *Enc. dir.*, ann. III, 2010, 805 ss.; B. Tonoletti, *Il mercato come oggetto della regolazione*, in *Rivista della regolazione dei mercati*, 1, 2014, 5 ss.; E. Bruti Liberati, *La regolazione indipendente dei mercati. Tecnica, politica e democrazia*, Torino, 2019.; M. Clarich, *Alle radici del paradigma regolatorio dei mercati*, in *Rivista della regolazione dei mercati*, 2, 2020, 230 ss.

connessi ovvero conseguenti.

5. La co-regolazione: sì, ma quale?

Il settore delle piattaforme digitali è, come ormai noto, caratterizzato da quel fenomeno di produzione normativa che si è soliti definire nei termini di "co-determinazione", ove pubblico e privato si fondono nella direzione (da taluni auspicata) di un superamento della distinzione fra diritto pubblico e diritto privato[61]. Si assiste ad un fenomeno che, per certi versi, non è del tutto nuovo: basti pensare – come di recente osservato in dottrina[62] e con riferimento al diritto interno – al settore del diritto del lavoro e a come lo strumento della legge, da un lato, e quello contrattuale, dall'altro, abbiano per così dire cooperato (e continuino a farlo) nella disciplina e nella regolazione, ad esempio, dei rapporti di lavoro, compreso – e, per certi versi, soprattutto con riferimento a – quello pubblico privatizzato.

Posta la necessità di configurare una regolazione delle piattaforme digitali, anche alla luce di quanto disposto dai regolamenti da ultimo analizzati – con la conseguenza di abbandonare definitivamente la stagione dell'auto-regolazione –, si tratta di capire quale tipo di regolazione meglio si presta al governo del mondo digitale. In altri termini, si pone l'esigenza di comprendere come e quanto regolare, il sé (ossia la necessità di una sua regolazione) lo si dà per presupposto.

Alla tradizionale macro-distinzione fra auto-regolazione privata, da una parte, ed etero-normazione pubblica, dall'altra – che si tende a tradurre in una vera e propria polarizzazione fra le due[63] –, sembra possibile enucleare, come già osservato in dottrina[64], una terza forma di regolazione, quella che si è soliti definire nei termini di "co-regolazione". Si tratta di una terza via di regolazione che tenta di riprendere gli aspetti postivi delle altre due tipologie di regolazione, accantonando, invece, quei profili che si sono rivelati essere a tutti gli effetti punti di debolezza che non permettono, quindi, di configurare una regolazione efficace ed effettiva di un mondo, quello delle tecnologie digitali, che – per ragioni note ed in parte già viste e per altre che si analizzeranno più avanti –, rappresenta un ambito a sé che necessita di una regolazione *ad hoc*, come, non a caso, fa il *DMA*. Quest'ultimo, del resto, sembra proprio accedere ad un tipo di

[61] Sul rapporto fra "diritto pubblico" e "diritto privato", ossia la tradizionale macro-distinzione del *jure*, cfr. B. Sordi, *Diritto pubblico e diritto privato. Una genealogia storica*, Bologna, 2020. In proposito cfr. anche il secondo numero del 2024 della Rivista *Antologia di diritto pubblico*.

[62] Aa.Vv., *Il futuro del diritto pubblico*, cit., 38 ss.

[63] Sul punto, più in generale, con riferimento, cioè, alle strategie in materia di intelligenza artificiale, con considerazioni, comunque, di più ampio respiro valide anche ai presenti fini, cfr. E. Chiti-B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della regolazione dei mercati*, 1, 2020, 29 ss. Con riferimento, in particolare, alla regolazione della rete cfr. M. Manetti, *Regolare Internet*, cit., 36, la quale sottolinea il parziale superamento di tale polarizzazione laddove osserva «che oggi in tutto il mondo, e non da ultimo negli Stati Uniti, si discute accanitamente non più sulla necessità, ma sui modi con i quali regolare efficacemente l'uso di Internet».

[64] A. Simoncini, *La co-regolazione delle piattaforme digitali*, in *Rivista trimestrale di diritto pubblico*, 4, 2022, 1031 ss. Sul punto cfr., più di recente, G. Pistorio, *La co-regolazione nell'ecosistema digitale tra etero-regolazione e auto-regolazione. Questioni definitorie*, in *Osservatorio sulle fonti*, 1, 2024, 138 ss.

regolazione che è riconducibile al modello della co-regolazione. Tuttavia, rimane da capire come debba essere articolato tale modello e, soprattutto, quale rapporto instaurare fra i pubblici poteri, da una parte, e le piattaforme, dall'altra, e, dunque, quale sia il punto di equilibrio più coerente con il ruolo che l'ordinamento riconosce ad entrambi o, meglio – in termini più ampi–, quello più coerente con l'ordinamento stesso nel suo complesso, a partire dal delicato profilo della legittimità delle relative scelte assunte.

La co-regolazione, come noto, implica un ruolo attivo tanto dei pubblici poteri quanto di quelli privati[65] e riposa sull'idea di una loro stretta cooperazione nella regolazione di un dato fenomeno.

Essa, dal punto di vista dell'ordinamento europeo, sembra rinvenire la propria legittimazione nell'ambito della Strategia sulla qualità della regolazione che l'UE porta avanti da, ormai, quasi un quarto di secolo e che, più di recente, ha visto un ulteriore ed importante sviluppo. Infatti, nell'aprile del 2021, la Commissione[66] ha presentato una strategia per fare il punto della situazione rispetto al tema del "legiferare meglio", mettendo in luce i punti di forza di tale approccio ma anche quelli di debolezza che si sono verificati nel corso degli anni. In essa si ricorda un elemento centrale della *Better regulation Strategy*, vale a dire il coinvolgimento, nel processo decisionale, (anche) della c.d. società civile e, quindi, di cittadini e imprese (ai cui interessi la presente comunicazione è particolarmente attenta). Sotto questo profilo, come ricordato di recente in dottrina[67], utili indicazioni derivano dal manuale di istruzioni per il c.d. *drafting*, scritto anzitutto per fornire indicazioni ai funzionari dell'UE; nell'ultima versione di tale manuale, invero, fra i possibili strumenti cui ricorrere per una migliore regolazione si richiama proprio quello della co-regolazione, strumento che rientra nelle misure di *soft regulation*.

Si tratta, quindi, di un modello di regolazione dove i compiti fra il pubblico ed il privato sono così ripartiti: ai pubblici poteri spetta di stabilire e definire quelli che sono i principi e gli obiettivi che si intende conseguire, mentre ai privati, sempre assieme ai primi (con i quali sono tenuti a collaborare), compete di dettare le disposizioni di carattere esecutivo-attuativo. La ragione di tale assetto è presto detta: mentre la presenza dei pubblici poteri serve in quanto si assicura così una forma di regolazione pubblica, dove, cioè, i pubblici poteri hanno un ruolo di primo piano[68], la presenza dei privati, invece, serve per poter sfruttare le loro conoscenze, il *know-how* di cui godono, senza il quale la regolazione non potrebbe essere efficace.

Ciò, dunque, offre una serie di indicazioni rispetto al profilo che si metteva in luce in precedenza, ossia il rapporto che si instaura fra pubblico e privato; in ogni caso, è comunque opportuno proseguire su questo punto al fine di sottolineare come, benché vi sia una stretta cooperazione, il rapporto non sia e non possa essere "alla pari", e la

[65] In merito, di recente, cfr. E. Catelani, *Modelli di co-regolazione fra diritto interno e UE: l'influenza dei cittadini, delle associazioni, degli* stakeholder, in *Osservatorio sulle fonti*, 1, 2024, 220 ss.

[66] COM (2021) 219.

[67] A. Simoncini, *La co-regolazione*, cit., 1033 ss.

[68] In questo modo si evita di prestare il fianco alla critica di un atteggiamento di totale indifferenza da parte loro, a tutto vantaggio dei forti poteri privati che sarebbero, altrimenti, controllori e controllati allo stesso tempo (facendo il buono ed il cattivo gioco di un settore sempre più cruciale nell'attuale società).

risposta di ciò è da ricercarsi, in ultima analisi, nel concetto stesso di interesse pubblico. La co-regolazione non implica affatto di dover porre sul medesimo identico piano i poteri pubblici e i poteri privati; differente, infatti, è, anzitutto e soprattutto, il fine da essi perseguito: mentre i primi sono istituzionalmente preordinati a curare l'interesse pubblico (cura che costituisce, al fondo, l'essenza e la giustificazione della loro stessa istituzione), i secondi sono legittimamente dediti al conseguimento del proprio interesse egoistico. Pertanto, ritenere equiparabili questi due diversi poteri, in tali casi, finisce per rischiare di sfociare nella messa in discussione del presupposto su cui si fondono le democrazie liberali, le quali, diversamente, si reggono sull'assunto per cui l'interesse pubblico è predeterminato dai pubblici poteri in qualità di rappresentanti dei cittadini e non, invece, oggetto di una contrattazione con i poteri privati, contrattazione che è, inevitabilmente, al ribasso in ragione dell'asimmetria informativa che sussiste fra i due e, più in generale, a causa del forte potere economico di questi ultimi[69]. Una determinazione dell'interesse da perseguire che, se avvenisse sulla base di una contrattazione fra pubblico e privato, farebbe sì che il suo contenuto non sarebbe più quello tradizionale, risolvendo in termini negativi il quesito circa la possibilità di poter ancora parlare, in tali fattispecie, di interesse pubblico. Si tratterebbe, in altri termini, di un interesse compromesso poiché avrebbe alla sua base un vizio di origine: quello di dover rappresentare non tanto e non solo gli interessi della collettività di riferimento, bensì anche e soprattutto interessi che con essi – non essendo stati selezionati dai pubblici poteri – difficilmente sarebbero collimanti; anzi, probabilmente risulterebbero spesso contrastanti con questi stessi interessi, snaturando la funzione propria dei pubblici poteri, con la conseguenza di assoggettarli ad una condizione servente dei forti poteri privati (con tutto ciò che ne deriva in termini di possibili ricatti) e dei loro interessi economici. La co-regolazione, presupponendo un ruolo rilevante dei privati, deve, tuttavia, necessariamente concedere loro qualcosa, non è pensabile che soggetti privati collaborino o cooperino senza trarre da tale attività alcun vantaggio[70]. L'idea di un apporto dei privati che si traduce unicamente in un "*do*" senza implicare anche un "*des*" è un'idea che pecca di ottimismo, a tratti ingenua, smentita, a più riprese, dalla realtà delle cose.

Qual è, quindi, il vantaggio per i privati, posto che non possa non esservi un vantaggio se si vuole la loro collaborazione? Un esempio è rappresentato dall'art. 45 del *DSA*, in base al quale la «Commissione ed il comitato incoraggiano e agevolano l'elaborazione di codici di condotta volontari a livello di Unione per contribuire alla corretta applicazione del presente regolamento, tenendo conto in particolare delle sfide specifiche connesse alla lotta ai diversi tipi di contenuti illegali e ai rischi sistemici, conformemente al diritto dell'Unione». Tale disposizione, come già rilevato, «legittima sul piano positivo l'assunzione da parte delle big tech di inediti compiti para-normativi»[71] e ciò

[69] E. Cremona, *L'erompere dei poteri privati nei mercati digitali*, cit., 903, il quale parla di «compromesso al ribasso» fra le esigenze del mercato, da un lato, e quelle di protezione dei diritti fondamentali, dall'altro, compromesso che finisce spesso per essere in danno del consumatore; tale danno, a detta dell'A., nel mondo digitale, sarebbe perpetrato per mezzo della «"captazione" del consenso al trattamento dei dati personali, che – nell'ambiente online – non è che un simulacro di una consapevole manifestazione di volontà».

[70] Cosa, questa, del resto, fisiologica ed anche comprensibile.

[71] M. Betzu, *Poteri pubblici e poteri privati*, cit., 180.

potrebbe rappresentare un aspetto problematico. Tuttavia, è opportuno considerare che questa previsione risponde all'esigenza di cui si diceva prima, ossia quella di dover inevitabilmente riconoscere qualcosa ai privati (in questo caso contribuire alla formazione dei codici di condotta)[72]. Inoltre, è la dimostrazione della necessità dei privati ai fini della regolazione; non è un caso, invero, che la giustificazione di tale spinta ad adottare codici di condotta risieda nell'esigenza di una corretta applicazione del presente regolamento.

La presenza dei privati nella regolazione del mondo digitale sembra, dunque, rappresentare un "sacrificio" che non pare possibile evitare dato che non solo essa è di particolare importanza ma, al momento, è anche imprescindibile in ogni qualsivoglia regolazione della materia in ragione delle conoscenze tecniche che costituiscono, di fatto, patrimonio (in buona parte) esclusivo di tali soggetti, conosciuto solo in termini superficiali e parziali da parte dei pubblici poteri.

Alla luce di quanto osservato, sembra, quindi, possibile svolgere qualche considerazione conclusiva circa il modello di co-regolazione, per tentare di offrire taluni spunti in ordine alla domanda avente ad oggetto il tipo di co-regolazione. Essa appare costruirsi sulla base di tre elementi che ruotano attorno: *i*) al rapporto fra pubblico e privato (ovvero, se si preferisce, fra pubblici poteri e poteri privati digitali), *ii*) alla combinazione fra *hard-regulation* e *soft-regulation* e, infine, *iii*) all'importanza del coordinamento.

In ordine al primo, pur essendo inevitabile la presenza dei poteri privati, essi non sono e non devono essere posti sul medesimo piano dei pubblici poteri per le ragioni già dette e che possono essere riassunte nell'assenza della loro legittimazione democratica e nella necessità che il contenuto dell'interesse pubblico da perseguire in concreto sia il risultato di una scelta ascrivibile (almeno con riferimento al suo contenuto di ordine generale) agli stessi pubblici poteri che godono, appunto, della legittimazione democratica necessaria ad effettuare le scelte (e, quindi, anche i sacrifici, allorché necessari) che il perseguimento dell'interesse pubblico inevitabilmente richiede di assumere.

Rispetto, invece, al rapporto tra *hard regulation* e *soft regulation*, sembra possibile osservare che il modello di co-regolazione che emerge dai due regolamenti – e che, come visto, è da ascrivere a quest'ultima –, in realtà, sia caratterizzato anche da qualche elemento tipico dell'*hard regulation* che appare, al fondo, necessario in ragione dell'esigenza di una regolazione che, comunque, deve (e sembra voglia essere) incisiva per far fronte allo (stra)potere dei colossi digitali. Invero, tale co-regolazione riprende parte dei tratti tipici delle misure di *hard-regulation*, vale a dire quelle regole dotate del carattere della vincolatività. Si pensi, per esempio, agli obblighi – taluni particolarmente stringenti – che sia il *DMA* che il *DSA* impongono alle piattaforme digitali, ovvero alle misure che la Commissione può adottare allorché dovesse manifestare una violazione di tali obblighi.

Si tratta, dunque, di un tipo di regolazione che combina elementi tipici della *hard-regulation* (per quanto attiene alla vincolatezza ed all'intensità forte della regolazione), con quelli tipici della *soft-regulation*, ossia la flessibilità delle previsioni, non da intendere

[72] Sui codici di condotta, nella prospettiva del contrasto alla disinformazione, cfr. O. Pollicino, *I codici di condotta tra* self-regulation *e* hard law: *esiste davvero una terza via per la regolazione digitale? Il caso della strategia europea contro la disinformazione* online, in *Rivista trimestrale di diritto pubblico*, 4, 2022, 1051 ss.

come derogabilità delle stesse bensì come capacità di mutare la regolazione e adattarla alle nuove esigenze, flessibilità testimoniata dal fatto che – come osservato – buona parte delle disposizioni dei regolamenti attribuiscono un ampio margine di valutazione e, quindi, di intervento, alla Commissione, la quale può calibrare diversamente le singole misure da adottare a seconda delle varie esigenze e dei diversi profili problematici che si trova a dover affrontare.

Infine, con riferimento al terzo elemento, appare evidente la centralità del coordinamento anzitutto fra i due attori della co-regolazione e, più in generale, fra i vari soggetti che sono chiamati alla regolazione delle piattaforme digitali. Ciò è coerente col fatto che all'interno della categoria della *self-regulation*, la casa, cioè, di origine della co-regolazione, è ricompreso il c.d. metodo aperto di coordinamento e l'esigenza del coordinamento è il naturale completamento del profilo della cooperazione – di cui si diceva poc'anzi –, la quale costituisce il presupposto di tale modello regolatorio. Il *DMA*, sotto questo profilo, dedica, non a caso, due articoli alla «cooperazione con le autorità nazionali» (art. 37) ed alla «cooperazione e coordinamento con le autorità nazionali competenti che applicano le norme in materia di concorrenza» (art. 38), senza dimenticare che l'esigenza di cooperazione, alla base del coordinamento, riveste importanza anche in relazione agli organi giurisdizionali nazionali (art. 39).

La co-regolazione flessibile si pone, dunque, in rotta di collisione con l'idea, di derivazione statunitense, per cui l'auto-regolazione sarebbe «sempre preferibile all'eteroregolazione, anche in ragione della flessibilità della prima rispetto alla rigidità della seconda»[73], smentendo l'assunto in forza del quale ove ci sia regolazione pubblica non possa esservi flessibilità della stessa: dipende da come la si configura.

La co-regolazione, quindi, come opzione non ideologica (di cui, talvolta, sono ammantate le ricostruzioni in termini di auto-regolazione ovvero di etero-regolazione) bensì quale modello regolatorio razionale ed anche di buon senso, per una serie di ragioni che si è cercato di delineare e che sono legate, sia pur a vario titolo, alla constatazione per cui – in ultima istanza e come già anticipato – la tecnologia digitale è «caratterizzata da un mix tale di complessità specialistica e rapidità evolutiva che in molti casi solo i destinatari stessi delle norme sono in possesso delle conoscenze necessarie a svolgere il compito normativo»[74] e che rendono, di conseguenza, insopprimibile la loro presenza, sia pur guidata e limitata dai pubblici poteri, i quali mantengono (devono mantenere) il potere circa la decisione finale e, dunque, l'assetto definitivo da attribuire agli interessi in gioco.

6. Regolazione forte delle piattaforme digitali e ruolo della concorrenza

Nelle pagine precedenti si è visto come la regolazione antitrust non si attagli al mondo delle piattaforme digitali, o meglio, il tradizionale modello di regolazione antitrust non sia, da solo, sufficiente a regolare un ambito delicato come quello delle piatta-

[73] L. Torchia, *I poteri di vigilanza, controllo e sanzionatori*, cit., 1103.

[74] Così A. Simoncini, *La co-regolazione*, cit., 1032.

forme digitali, caratterizzato da pochi ma potenti "colossi". In questo senso, infatti, si muovono gli interventi del legislatore europeo che, di recente, ha configurato una regolamentazione, attraverso il *DMA* e il *DSA*, la quale, partendo proprio dal presupposto dell'insufficienza della regolazione antitrust – testimoniata dall'assenza, di fatto, nel settore digitale di un mercato concorrenziale, assistendo, viceversa, ad un mercato oligopolistico – ha introdotto una specifica regolazione che tenta di contrastare lo strapotere delle piattaforme per mezzo dell'introduzione di regole tese ad instaurare un regime concorrenziale.

Sotto questo profilo, assume particolare rilievo il concetto di "servizio di piattaforma di base" fatto proprio dal *DMA*[75]. Si tratta di un concetto che serve per tentare di configurare un sistema di monitoraggio della quota di controllo del mercato digitale da parte dei pochi ma potenti poteri privati in qualità di oligopolisti. Al momento, la Commissione, dopo aver individuato sei *gatekeeper* (*Iphabet*, *Amazon*, *Apple*, *ByteDance*, *Meta* e *Microsoft***)**, ha designato ben ventidue servizi di base forniti dagli stessi *gatekeeper*, ed altri ancora se ne potrebbero aggiungere in futuro. Si va dai *social network* (*TikTok*, *Facebook*, *Instagram* e *LinkedIn*), ai *browser* (*Safari* e *Chrome*), ai sistemi operativi (*Google Android*, *iOs* e *Windows Pc Os*), ai s*oftware* per spazi pubblicitari (*Google*, *Amazon* e *Meta*), fino ai servizi di intermediazione (*Google Maps*, *Google Play*, *Google Shopping*, *Amazon Marketplace*, *App Store* e *Meta Marketplace*) nonché a quelli di messaggistica (*WhatsApp* e *Messenger*), senza considerare che vi rientrano altresì piattaforme quali *Google Search* (in qualità di motore di ricerca) e *YouTube* (nella veste di piattaforma abilitata alla condivisione di video).

Si assiste, quindi, ad un passaggio da un sistema basato sulle autorità antitrust ad uno dove i pubblici poteri, in particolare la Commissione, rivestono un ruolo centrale. Un modello, cioè, di regolazione che – con tutti i limiti del caso ed al netto di una serie di criticità, come si avrà modo di osservare – va verso uno spostamento del potere decisionale: da una regolazione antitrust (e, quindi, in teoria neutrale), infatti, ad una politica, segnando un'inversione di rotta rispetto a quanto accaduto negli ultimi decenni, i quali hanno visto ampliare il raggio di azione delle autorità indipendenti (nelle loro molteplici differenti forme). Invero, sulla base di una tendenziale maggiore efficacia della logica tecnocratica – che accampa la pretesa di informare di neutralità ciò che non sempre lo può essere dal punto di vista ontologico poiché, talvolta, implica inevitabilmente una scelta a fronte di più opzioni allo stesso modo scientificamente fondate e, dunque, sotto questo profilo, valide[76] –, si riteneva che un soggetto indipendente come le autorità fosse meglio in grado di regolare materie dall'alto tasso di tecnicismo e dove sono in gioco interessi sensibili, a discapito così dell'etero-regolazione che ha tradizionalmente nei pubblici poteri il centro propulsore della propria azione.

Il ruolo centrale della Commissione è, inoltre, testimoniato dal potere che le è attribuito al fine di evitare le c.d. *killer acquisition*, vale a dire le acquisizioni (da parte delle aziende dominatrici del mercato digitale) delle società emergenti in quel mercato, per

[75] Art. 2, par. 2.

[76] Si tratta, quindi, di una scelta che rientra, a tutti gli effetti, nell'attività discrezionale, più precisamente nell'ambito della discrezionalità politico-amministrativa dove vi è, quindi, per definizione, un potere di scelta; cfr. D. Sorace-S. Torricelli, *Diritto delle amministrazioni pubbliche. Una introduzione*, Bologna, 2023, 285 ss.

impedire con ciò – è evidente – un'elusione della normativa[77] in questione finalizzata alla riacquisizione della quota di mercato perduta, il che condurrebbe le piattaforme digitali a riacquisire quel forte potere che il legislatore aveva, invece, tentato di ridimensionare o, comunque, contenere.

Il modello di regolazione che sembra, dunque, emergere e che appare richiesto dall'attuale assetto sbilanciato a tutto favore delle piattaforme digitali è, come visto, un modello che risponde al prototipo della co-regolazione, dotato di una non irrilevante dose di flessibilità, la quale è assolutamente necessaria e dalla quale non pare si possa in alcun modo prescindere. La regolazione, in questo settore, se non è flessibile, rischia di configurare un quadro di regole che, in un lasso di tempo piuttosto breve, è assai probabile che si riveli antiquato e non più "al passo con i tempi", ingenerando confusione e incertezza fra gli operatori e, di conseguenza, fra gli stessi consumatori-utenti.

La regolazione *ex ante* mira, quindi, anche a prevenire i possibili – anzi, constatando la realtà di fatto degli ultimi anni, verrebbe da dire assai probabili – fallimenti del mercato.

La precarietà della norma tecnica pubblica[78] può essere un punto di forza del modello della regolazione flessibile, la quale, è chiaro, rende possibile – ed in ciò rinviene buona parte delle ragioni della sua stessa configurazione in termini di flessibilità – quel continuo e puntuale aggiornamento che si pone come necessario per tentare di governare, e quindi in parte anche anticipare, le evoluzioni tecnologiche.

Si tratta, dunque, di una regolazione che si potrebbe definire – anche alla luce di quanto osservato circa i rapporti fra pubblico e privato – come "co-regolazione flessibile non paritaria" che interviene in funzione preventiva per regolare ed orientare il fenomeno del digitale.

È, tuttavia, auspicabile che tale regolazione non sia solamente flessibile ma anche "forte". Si intende con ciò sottolineare la necessità di una regolazione forte non – come si potrebbe in un primo momento ritenere – nel senso di una regolazione preordinata alla contrazione dello spazio (di autonomia) dei privati sulla base di un generale (e da taluni auspicato) ritorno sulla scena dei poteri pubblici in funzione, quindi, anti-mercato, bensì l'esatto opposto: una regolazione che si traduce in un intervento teso ad attuare il regime concorrenziale, ossia una regolazione che si muove entro la logica del mercato e coerente – da questo punto di vista – con l'idea di fondo su cui si regge lo stesso ordinamento europeo, ossia quella che fa capo al(l'economia sociale di mercato) fortemente competitiva (art. 3, par. 3, TUE).

La regolazione attuale che emerge, in particolare, dai più recenti interventi legislativi a livello europeo, è flessibile ma non ancora del tutto forte poiché mira a conseguire gli obiettivi senza tendenzialmente individuare gli strumenti e le modalità per conseguirli, mentre, all'opposto, è necessario che l'orizzonte della riflessione si concentri anche su questo profilo, ossia quello dell'esigenza di più regole prescrittive e di una maggiore predeterminazione dei comportamenti da sanzionare. Un *mix*, cioè, di flessibilità (da intendere come capacità di adattamento della regolazione ai mutamenti della realtà) ed incisività (da declinare nel senso di più regole prescrittive ed un potere pubblico più forte, sia in funzione preventiva che repressiva).

---

[77] Cfr. il punto 70 del *Considerando* del *DMA*.

[78] A. Iannuzzi, *Paradigmi normativi per la disciplina della tecnologia*, cit., 100.

Per esempio, il sistema di sanzioni configurato dal legislatore europeo attraverso il *DMA*, sembra peccare di coraggio, dando luogo ad una serie di sanzioni affette da quella che si potrebbe definire nei termini di "timidezza legislativa". Le ammende del 10% (art. 30, par. 1) e del 20 % (art. 30, par. 2) – la cui percentuale è da calcolare avuto riguardo al fatturato totale del *gatekeeper* realizzato a livello mondiale nel corso del precedente esercizio finanziario – sono condivisibili ma avrebbero ben potuto prevedere ulteriori e più alte percentuali, da parametrare, ovviamente, in rapporto alla gravità della condotta tenuta ed al tipo di obbligo violato (in linea, così, al criterio di proporzionalità che assiste in generale le sanzioni), dato che l'efficacia deterrente di una sanzione, come noto, si misura sulla base della ricchezza posseduta (colpita in quanto detenuta) o, come in tal caso, sul fatturato (colpito in quanto generato).

In altre parole, la percentuale, di per sé, dice poco circa la capacità di una sanzione di impedire quel determinato comportamento poiché ciò che, al fondo, effettivamente rileva, è quanto incide quella specifica misura sul soggetto in questione, senza considerare che quella determinata percentuale non ha lo stesso peso per tutte le imprese. Pertanto, in generale, è preferibile ricorrere ad una differenziazione, prevedendo più ammende e, quindi, diverse percentuali proporzionate alla gravità dell'infrazione commessa. Inoltre, nel caso di specie, essendo sanzioni applicabili ad imprese di grandi dimensioni e dagli elevati fatturati, si sarebbe potuto auspicare maggiore coraggio dal legislatore, magari prevedendone di ulteriori, dato che le due previste – per quanto possano tradursi in ingenti quantità economiche – costituiscono, comunque, una parte residuale del fatturato e, perciò, rispetto ad esse, appare lecito dubitare della relativa forza deterrente.

In tutto ciò, prima di concludere, preme sottolineare il valore che la concorrenza sembrerebbe destinata a (dover) rivestire. Essa, in questi casi, serve per recedere lo strapotere e la connessa condizione di oligopolio in cui tali poteri agiscono liberi ed indisturbati. Una concorrenza, quindi, non come "fine" o, comunque, come strumentale unicamente alla tutela del mercato in sé, ma, diversamente, quale strumento che ha di mira, in ultima istanza, sempre gli operatori economici e, dunque, i consumatori-utenti. Una concorrenza, in altre parole, come strumento di riequilibrio fra le differenti posizioni di forza, al fine di ricondurre esse ad una situazione di maggiore uguaglianza, una concorrenza che, in ultimissima analisi, si prefigge (anche) di assolvere compiti che – utilizzando un'espressione a cui si fa sempre meno ricorso – un tempo rientravano nella finalità di assicurare la giustizia sociale, alla cui attuazione il diritto più in generale dovrebbe tendere: lo impongono, del resto, le stesse Carte costituzionali degli ordinamenti europei, le quali hanno l'ambizione di tenere assieme libertà individuale, da un lato, e giustizia sociale, dall'altro, l'una (quest'ultima) come precondizione per assicurare l'effettiva realizzazione dell'altra (la prima) e, quindi, del pieno sviluppo della personalità di ciascun essere umano, rispetto al quale una componente fondamentale che viene in rilievo nel settore digitale è – come già anticipato – la libertà di manifestazione del pensiero.

7. La libertà di manifestazione del pensiero in rete ed i suoi aspetti problematici: in particolare, il caso delle *fake news* e del *hate speech*

La prima premessa da cui partire – benché a tutti nota – è che la libertà di manifestazione del pensiero rappresenta un diritto fondamentale riconosciuto e tutelato tanto sul piano internazionale quanto su quello interno. La Dichiarazione universale dei diritti dell'uomo afferma che ciascun «individuo ha diritto alla libertà di opinione e di espressione» (art. 19). Tale diritto è, poi, ribadito e specificato dal Patto internazionale sui diritti civili e politici laddove si osserva che ogni individuo «ha diritto alla libertà di espressione», specificando che tale diritto «comprende la libertà di cercare, riceve e diffondere informazioni e idee di ogni genere, senza riguardo a frontiere, oralmente, per iscritto, attraverso la stampa, in forma artistica o attraverso qualsiasi altro mezzo di sua scelta» (art. 19). Anche la Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (CEDU) afferma tale libertà, ribandendo come questo diritto includa «la libertà d'opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera» (art. 10)[79].

A livello interno, invece, la disposizione di riferimento è rappresentata dall'art. 21 Cost.[80], a detta del quale tutti «hanno diritto di manifestare liberamente il proprio pensiero con la parola, lo scritto e ogni altro mezzo di diffusione».

La seconda premessa da cui muovere è che Internet ha completamente rivoluzionato il modo attraverso il quale le persone comunicano e condividono le loro idee, mutando così le stesse forme di espressione della libertà di manifestazione del pensiero[81]; del resto, «il mondo virtuale non è altro che una proiezione di quello reale»[82], riflette quest'ultimo ed in taluni casi, come quello in esame, ne rappresenta una estensione. La rete, infatti, offre una piattaforma globale per la manifestazione del pensiero, consentendo, potenzialmente a chiunque, di esprimere le proprie opinioni, oltre che di diffondere nonché accedere ad una quantità, di fatto illimitata, di informazioni di ogni tipo[83].

È evidente, quindi, come tale libertà comporti anche nuove sfide per il diritto[84]. La diffusione di contenuti *online*, invero, può creare – come, in effetti, ha creato – fenomeni

[79] Sul punto cfr. la prospettiva di A. Cardone, *L'incidenza della libertà di espressione garantita dall'art. C.e.d.u. nell'ordinamento costituzionale italiano*, in *Osservatorio sulle fonti*, 3, 2012, 1 ss. Sulla libertà di espressione, più in generale, cfr. l'approfondito contributo di M. Luciani, *La libertà di espressione, una prospettiva di diritto comparato* -Italia, in *Studio per il Servizio Ricerca del Parlamento europeo*, PE, 2019.

[80] Su cui non si prescinda da C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958.

[81] Sul punto, per un inquadramento generale, cfr. G. Cassano-A. Contaldo, *Internet e tutela della libertà di espressione*, Milano, 2009; M. Betzu, *Regolare Internet. Le libertà di informazione e di comunicazione nell'era digitale*, Torino, 2012.

[82] Aa.Vv., *Il futuro del diritto pubblico*, cit., p. 100.

[83] «Fin dall'inizio si è invero constatato che la Rete, a differenza dell'etere, non presenta limiti di utilizzo, potendo anzi diffondere una quantità di informazioni infinitamente superiore alla capacità dell'essere umano di prenderne visione», così M. Manetti, *Internet e i nuovi pericoli per la libertà di informazione*, in *Quaderni costituzionali*, 3, 2023, 537.

[84] In merito cfr. G. Pitruzzella, *La libertà di informazione nell'era di Internet*, in questa *Rivista*, 1, 2018, 4.

di disinformazione[85], nuove e più agguerrite forme di incitamento all'odio[86], oltre che una maggiore condivisione di informazioni personali da parte di ciascun utente, aprendo così la strada ad inedite forme di violazione della *privacy* degli individui[87].

La (supposta) libertà in rete[88] è predominio delle piattaforme digitali[89]. Sono esse, infatti, che detengono il controllo della libertà di manifestazione del pensiero e, spesso, anche delle controversie che hanno ad oggetto la sua presunta violazione. Ciò pone una serie evidente di aspetti problematici, a partire dalla legittimazione delle piattaforme digitali di incidere su una libertà fondamentale quale quella in esame. Si tratta di problemi noti e che determinano una serie di implicazioni di ordine generale, tanto a livello teorico quanto a livello pratico. L'interesse di tali piattaforme ad un controllo su questa libertà è particolarmente forte poiché, si sa, attraverso di essa si può incidere sull'opinione pubblica e condizionarla (in diverse direzioni e per distinte finalità).

La libertà di manifestazione del pensiero rappresenta, poi, un aspetto particolarmente rilevante poiché essa è strettamente legata alla democraticità del sistema ove la si intende esercitare e sulla base della tutela ad essa riconosciuta si può, dunque, misurare il grado di democraticità di quello stesso sistema[90], in tal caso della rete[91].

---

[85] In proposito, con specifica attenzione alle misure di contrasto di tale fenomeno a livello internazionale, cfr. F. Sciacchitano-A. Panza, Fake news *e disinformazione online: misure internazionali*, in questa *Rivista*, 1, 2020, 102 ss. Sul punto cfr. anche M. Monti, Fake news *e* social network: *la verità ai tempi di* Facebook, in questa *Rivista*, 1, 2017, 79 ss. Più in generale, sulla disinformazione cfr. M. Cuniberti, *Il contrasto alla disinformazione in rete tra logiche del mercato e (vecchie e nuove) velleità di controllo*, in questa *Rivista*, 1, 2017, 26 ss.; S. Sassi, *Disinformazione contro costituzionalismo*, Napoli, 2021 (della stessa A. cfr. anche *L'Unione Europea e la lotta alla disinformazione*, in *Federalismi.it*, 15, 2023, 183 ss.); C. Hassan-C. Pinelli, *Disinformazione e democrazia. Populismo, rete e regolazione*, Venezia, 2022. In proposito, con particolare riferimento all'intelligenza artificiale, di recente, cfr. O. Pollicino-P. Dunn, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2024.

[86] Sul punto cfr. P. Falletta, *Controlli e responsabilità dei* social network *sui discorsi d'odio online*, in questa *Rivista*, 1, 2020, 146 ss. In proposito anche M. Manetti, *Regolare Internet*, cit., 40 ss.

[87] Più in generale, come è stato di recente osservato, l'aumento «dell'utilizzo del web come strumento di comunicazione ha determinato, in senso positivo, un ampliamento degli spazi entro cui gli individui svolgono la propria personalità, ma ha parallelamente allargato, in senso negativo, la sfera dei comportamenti lesivi delle libertà altrui», Aa.Vv., *Il futuro del diritto pubblico*, cit., 101.

[88] In proposito, dove si riflette sulla libera *dalla* rete e sulla libertà *in* rete, cfr. G. De Minico, *Libertà* in *Rete. Libertà* dalla *Rete*, Torino, 2020, *passim*.

[89] Sul punto, con particolare riguardo al rapporto che intercorre tra utenti, intermediari e gestori delle piattaforme, da un lato, e potere pubblico, dall'altro, cfr. M. Cuniberti, *Potere e libertà nella rete*, cit., 39 ss.

[90] La libertà di manifestazione del pensiero, non a caso, è stata – come noto – definita dalla Corte costituzionale quale «pietra angolare dell'ordine democratico», Corte cost., 2 aprile 1969, n. 84, punto 5 del *Considerato in diritto*. Sul punto, di recente, cfr. L. Buffoni, *Sulle libertà. Contro le dicotomie*, in *Osservatorio sulle fonti*, n. 3, 2024, 35, la quale, muovendo dal modello discorsivo di Carta costituzionale prospettato da Habermas «ove il principio del discorso […] poggia sulla co-originarietà di autonomia pubblica e privata, perché "l'autonomia pubblica dei cittadini dello Stato non è immaginabile a prescindere dalla autonomia privata dei membri della società, e viceversa. Entrambe le autonomie si presuppongono reciprocamente", dove "pubblico e privato, potere e libertà, sono, dunque, uniti, intrecciati», e l'esercizio in concreto dei diritti di libertà rileva quale «condizione essenziale della "democrazia pluralista" e della partecipazione del popolo sovrano alla "elaborazione dell'indirizzo politico" e viceversa», osserva che è lo stesso «sostrato teorico che fa della libertà individuale di manifestazione del pensiero e del di ritto all'informazione «la pietra angolare del sistema democratico», nella nota decisione costituzionale n. 84 del 1969».

[91] In proposito cfr. L. Califano, *La libertà di manifestazione del pensiero… in rete; nuove frontiere di esercizio di un diritto antico.* Fake news, hate speech *e profili di responsabilità dei social network*, in *Federalismi.it*, n. 26,

L'idea di Internet e della rete come luogo di libertà ove quest'ultima regnava sovrana, essendo, per definizione, uno spazio privo di ogni forma di controllo e limitazione[92], dando così luogo ad una società libertà ed aperta[93], si è rivelata – come già osservato – una mera illusione. Invero, la libertà che i poteri privati che gestiscono le piattaforme generalmente assicurano, tende ad arrestarsi nel momento in cui il pensiero espresso non sia in linea con il verbo dominante e dietro la giustificazione della rimozione di alcuni contenuti in quanto non veritieri (le c.d. *fake news*[94])[95] ovvero offensivi (*hate speech*)[96], talvolta, si celano valutazioni di tutt'altra natura, rispetto alle quali è difficile non scorgere interessi che, in ultima analisi, sono di natura economica. Infatti, se, da una parte, è vero che «i social media, e più in generale la rete, consentono alla libertà di manifestazione del pensiero di essere esercitata con una facilità che non si *era* mai vista prima», è altrettanto vero, dall'altra, che in tale spazio vengono consentite «forme di controllo del pensiero e di sorveglianza dei cittadini che, egualmente, non si erano mai viste»[97].
Si pensi, giusto a titolo di esempio – e lungi dal voler assumere posizione sul merito di una questione assai delicata troppo spesso strumentalizzata per bieche finalità pro-

2021, 2. Con particolare riferimento ai *social network* ove si riflette, altresì, sulla relativa loro qualificazione giuridica, cfr. M. Bassini, *Libertà di espressione e* social network*, tra nuovi "spazi pubblici" e "poteri privati". Spunti di comparazione*, in questa *Rivista*, 2, 2021, 67 ss.

[92] Dove, per esempio, la libertà di manifestazione del pensiero si sarebbe realizzata compiutamente più che in ogni altro contesto ed in ogni altra forma possibile in ragione dell'assenza, giustappunto, di un controllo esterno.

[93] «Un mondo, quello dei social media che, in realtà, tutto è meno che una società aperta; la promessa di libertà, di disintermediazione, di autonomia che internet sembrava promettere, risulta compromessa da una terra virtuale proprietà di un nuovo sovrano che deve portarci a riflettere, senza peraltro incorrere nel rischio della proposta di elaborazioni e soluzioni inedite, sulla tenuta dei principi elaborati nel tempo dalla più accorta e autorevole dottrina», così L. Califano, *La libertà di manifestazione del pensiero*, cit., 6.

[94] Sul punto, per un inquadramento del fenomeno con riferimento alla libertà di informazione, cfr. F. Donati, Fake news *e libertà di informazione, in Scritti in onore di Giovanni Furgiuele*, tomo I, Mantova, 2017, 125 ss. Con attenzione al rapporto tra *fake news* e democrazia cfr. E. Lehner, Fake news *e democrazia*, in questa *Rivista*, 1, 2019, 93 ss.

[95] Una questione particolarmente complessa attiene al rapporto tra *fake news* e "diritto ad essere informati". Invero, in ragione dell'esistenza di quest'ultimo diritto, si potrebbe sostenere che le *fake news* rappresentino uno strumento attraverso il quale si perpetra la sua violazione. La risposta all'interrogativo dipende a seconda di come si interpreti il contenuto della libertà di manifestazione del pensiero *ex* art. 21 Cost. Invero, la «risposta è positiva ove si ritenga che l'art. 21 Cost. vieti le manifestazioni del pensiero consapevolmente e subiettivamente false», mentre «è di segno opposto ove non si accolga la concezione funzionale della libertà di manifestazione del pensiero, secondo la quale quest'ultima dovrebbe essere garantita solo in quanto vera e pertanto nei limiti della sua utilità rispetto alla preservazione e al consolidamento delle strutture dello Stato democratico-costituzionale, che esigerebbero la verità», così P. Caretti-A. Cardone, *Il diritto dell'informazione e della comunicazione*, cit., 258. Sul rapporto fra verità e libertà di manifestazione del pensiero, per tutti, cfr. C. Pinelli, *"Postverità", verità e libertà di manifestazione del pensiero*, in questa *Rivista*, 1, 2017, 41 ss. Sull'art. 21 Cost. e la sua evoluzione, di recente, cfr. F. Donati, *L'art. 21 della Costituzione settanta anni dopo*, in questa *Rivista*, 1, 2018, 93 ss. Per l'idea della necessità di un intervento (puntuale) di modifica di tale disposizione, cfr. M. Orofino, *Art. 21 Cost.: le ragioni per un intervento di manutenzione ordinaria*, in questa *Rivista*, 2, 2019, 77 ss.

[96] Per un inquadramento generale circa i caratteri assunti dal costituzionalismo europeo in ordine a tali due fenomeni cfr. O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione nell'era di Internet*, in questa *Rivista*, 1, 2018, 48 ss. Sul punto cfr. anche G. Pitruzzella-O. Pollicino-G.S. Quintarelli, *Parole e potere: libertà di espressione,* hate speech *e* fake news, Milano, 2017.

[97] Così G.L. Conti, *Manifestazione del pensiero attraverso la rete e trasformazione della libertà di espressione: c'è ancora da ballare per strada?*, in *Rivista AIC*, 4, 2018, 203.

pagandistiche –, a quanto avvenuto, di recente, a proposito del sanguinoso conflitto fra Israele e Palestina. Uno dei *gatekepeer*, *Meta*, su alcune piattaforme da essa gestite (come *Instagram*, *Facebook*, *WhatsApp* e *Messanger*), ha adottato una serie di misure – dalla cancellazione di contenuti alla rimozione dei profili – nei confronti di quelle manifestazioni tese a divulgare contenuti aventi ad oggetto la Palestina ed il suo popolo.

Si tratta, dunque, di forme di controllo e sorveglianza che si traducono anche in scelte del tutto arbitrarie di rimozione di determinati contenuti, senza possibilità di reale appello per chi quel contenuto si vede rimosso. Alla "privatizzazione del potere" – che rappresenta, sotto questo profilo, la causa – ne consegue la privatizzazione della censura, alla quale si lega una privatizzazione della stessa possibilità di difendersi, con un accentramento in capo a chi ha deciso (la censura) del potere di dirimere la controversia in ordine a quella stessa decisione precedentemente adottata[98]. Come è stato osservato, una siffatta privatizzazione della censura[99] è da considerarsi «irragionevole sul piano dei principi costituzionali: non solo la censura è costituzionalmente inconcepibile se non in casi limitatissimi, negli schemi del diritto costituzionale d'occidente, e oggetto di una riserva di legge e di una riserva di giurisdizione, ma immaginare che si possa affidare la censura a organizzazioni private secondo modelli che ricordano lo scandirsi orizzontale del principio di sussidiarietà appare impossibile»[100].

Due sono i profili in proposito su cui si intende, sia pur rapidamente, soffermare l'attenzione, ricorrendo a due esempi: quello che appare, a tutti gli effetti, una limitazione della libertà di manifestazione del pensiero e quello che rappresenta una disparità di trattamento fra lo schieramento pro-Palestina e quello pro-Israele. Si tratta di due profili che si distinguono solamente per sviluppare meglio il discorso (essendo, infatti, assai correlati): anche quello relativo alla disparità di trattamento si manifesta, al fondo, come una limitazione della libertà di (manifestazione del) pensiero.

Per essere più chiari, la decisione di *Meta* non è stata quella di bandire ogni contenuto avente ad oggetto il conflitto israeliano-palestinese, scelta che avrebbe anch'essa senz'altro rappresentato una lesione alla libertà di manifestazione del pensiero ma, almeno, non avrebbe realizzato una condotta che, per il tramite della disparità di trattamento, ha finito per tradursi in una lesione del principio di uguaglianza. Ciò, inoltre, è la riprova di quanto si sottolineava poc'anzi: dietro le scelte di tali piattaforme vi sono interessi di natura economica, i quali hanno fatto sì che fossero considerati come "leciti" quei contenuti pro-Israele e, invece, illeciti o, comunque, in qualche modo offensivi, quelli pro-Palestina.

Un discorso analogo, che si lega, tuttavia, più al profilo dell'incitamento all'odio, potrebbe essere sviluppato a proposito della guerra fra Russia ed Ucraina[101]. Come, in-

[98] Si pensi, ad esempio, alla creazione, da parte di *Facebook* ed *Instragram*, di un *Indipendent Oversight Board*, una sorta di comitato di vigilanza che si propone di svolgere attività para-giurisdizionali in materia di diritti fondamentali come, giustappunto, rispetto alla libertà di espressione. Sul punto, di recente, cfr. Aa.Vv., *Il futuro del diritto pubblico*, cit., 102.

[99] Sulla censura privata cfr. M. Cuniberti, *Potere e libertà nella rete*, cit., 51 ss.

[100] G.L. Conti, *Manifestazione del pensiero attraverso la rete e trasformazione della libertà di espressione*, cit., 216.

[101] Con particolare riferimento alla decisione del Consiglio dell'Unione europea di sospendere taluni organi di informazione ed agenzie di stampa sottoposti al controllo dello Stato russo cfr. S. Lattanzi, *La lotta alla disinformazione nei rapporti tra Unione e Stati terzi alla luce del conflitto russo-ucraino*, in questa *Rivista*,

fatti, si ricorderà, nel marzo del 2022, *Meta* ha deciso di allentare la propria politica di contrasto all'odio rispetto a quei *post* che avevano, come contenuto, forme di violenza verbale nei confronti dei soldati russi. Invero, *Meta* ha reso possibile, per gli iscritti di *Facebook* ed *Instagram*, postare contenuti offensivi[102] e/o di incitamento alla violenza nei riguardi dell'esercito russo. Tale scelta è stata assunta sulla base della volontà, espressamente manifestata, di concedere, temporaneamente, forme di espressione politica che sarebbero, di regola, vietate in quanto contrarie alle regole stabilite da Meta. Per esempio, sono stati consentiti *post* violenti che avevano ad oggetto l'invocare la "morte agli invasori russi", specificando, tuttavia, il divieto di estendere siffatti contenuti nei riguardi dei civili russi.

È chiaro che questo ponga una serie di criticità ed apra taluni interrogativi come, ad esempio, la domanda circa sulla scorta di quale legittimazione le piattaforme dei *social* decidono di rimuovere determinati contenuti, legittimazione che, se non si ragiona sulla base della "logica proprietaria" – ossia in considerazione del fatto che sono loro i proprietari di quelle stesse piattaforme –, sembra difficile da trovare. Inoltre, la circolazione delle idee e delle opinioni e, quindi, la stessa libertà di manifestazione del pensiero, non costituisce certo lo scopo principale della rete, il quale, piuttosto, è quello di «produrre traffico di qualità che può essere convertito in un valore commerciale attraverso la pubblicità»[103]. Ciò, di conseguenza, rende ancora più forte il rischio di una lesione della libertà di manifestazione del pensiero e rende ancor più indispensabile una regolazione pubblica, con la precisazione e, soprattutto, l'auspicio – come ora meglio si dirà –, che essa non finisca per tradursi nella medesima situazione la cui esistenza ne ha legittimato (e reso necessario) l'intervento: ossia quella di sfociare, a sua volta, in una fonte di restrizione della libertà in esame.

8. I rischi da evitare: una regolazione censoria della libertà di manifestazione del pensiero

Tutto ciò si traduce nella necessità di un controllo pubblico anche, e soprattutto, per garantire l'effettività dei diritti e delle libertà come, per esempio, quella in questione. Ciò, allo stesso tempo, però, non deve significare che il controllo pubblico e, più in generale, il maggior intervento dei pubblici poteri mediante una regolazione più forte ed incisiva delle piattaforme, si traduca in una situazione analoga a quella attuale o, peggio ancora, in una ulteriore restrizione alla libertà di manifestazione del pensiero, opzione non scartabile a priori, anzi, teoricamente possibile, per non dire probabile. È necessario che il controllo in rete sia supervisionato – se così si può dire – dai pubblici poteri ma non ciò non può e non deve tradursi in una diminuzione dello spazio di libertà, altrimenti il problema non sarebbe risolto e la libertà di manifestazione del pensiero continuerebbe a subire ingiustificate lesioni. Si tratta, quindi, di assicurare un

3, 2022, 158 ss.

[102] Sul controllo, da parte delle piattaforme *online*, circa i contenuti pericolosi, cfr. C. Bassu, *Piattaforme online e controllo dei contenuti pericolosi*, in questa *Rivista*, 1, 2020, 230 ss.

[103] G.L. Conti, *La Costituzione al tempo della simbiosi uomo macchina*, in *Osservatorio sulle fonti*, 3, 2024, 113.

controllo pubblico per fronteggiare l'arbitrarietà dei poteri digitali, senza, però, diminuire lo spazio di libertà: la questione, è chiaro, risulta particolarmente complessa; tale rischio, invero, sussiste eccome[104].

L'intervento regolatorio del legislatore, con riferimento al *DSA*, presenta, sotto questo profilo, talune criticità. Il *DSA*, come accennato, ha fra i suoi obiettivi anche quello di tutelare la libertà di espressione e di informazione[105]. Invero, benché l'obiettivo del legislatore europeo attraverso tale intervento (e lo stesso si può sostenere avuto riguardo al *DMA*) sia quello di garantire il buon funzionamento del mercato[106], coerentemente con la centralità che quest'ultimo riveste nell'ordinamento europeo (a partire dall'art. 3, par. 3, TUE), egli si (pre)occupa anche di disciplinare taluni aspetti afferenti al diritto in questione[107]. Infatti, si fa riferimento all'onere, per i prestatori dei servizi, di agire immediatamente per rimuovere le attività illegali (o i contenuti illegali), ovvero per disabilitare l'accesso agli stessi, non appena ne vengano effettivamente a conoscenza (o ne divengano consapevoli), per poter, così, beneficiare dell'esenzione dalla responsabilità per i servizi di memorizzazione di informazioni. Tale rimozione dei contenuti[108], si legge, dovrebbe essere effettuata nel rispetto dei diritti fondamentali dei destinatari del servizio stesso, a muovere da quello alla libertà di espressione e di informazione[109]. Si prevede, poi, che nel progettare, applicare e far rispettare tali restrizioni, i prestatori di servizi intermediari dovrebbero agire in modo non arbitrario e non discriminatorio, oltre a dover tener conto dei diritti e degli interessi legittimi dei destinatari del servizio, compresi i diritti fondamentali sanciti dalla Carta. A titolo di esempio, i fornitori di piattaforme *online* di dimensioni molto grandi dovrebbero in particolare tenere debitamente conto della libertà di espressione e di informazione, compresi la libertà ed il pluralismo dei media[110], aspetto, quest'ultimo, particolarmente importante[111] nell'ottica

---

104 Sul punto, di recente, avuto riguardo alla libertà di informazione, cfr. M. Manetti, *Internet e i nuovi pericoli per la libertà di informazione*, cit., 523 ss.

105 Cfr. il punto 3 dei *Considerando*.

106 Sul punto cfr. P. Caretti-A. Cardone, *Il diritto dell'informazione e della comunicazione*, cit., 267, ove si rileva che a conferma della «matrice strettamente economistica di tali provvedimenti [*DSA* e *DMA*] basti ricordare la loro base legale, ovvero l'art. 114 del TFUE: l'obiettivo principale di queste forme di regolazione è, infatti, quello di garantire il buon funzionamento del mercato interno, in particolare per quanto riguarda la prestazione dei servizi digitali transfrontalieri».

107 In particolare, il profilo relativo alla disinformazione, su cui cfr. A. Gullo, *Contenuti, scopi e traiettoria della ricerca: le nuove frontiere della compliance nel mercato digitale*, 13 ss.; L. D'Agostino, *Disinformazione e obblighi di* compliance *degli operatori del mercato digitale alla luce del nuovo* Digital Services Act, 16 ss.; E. Birritteri, *Contrasto alla disinformazione,* Digital Services *Act e attività di* private enforcement*: fondamento, contenuti e limiti degli obblighi di* compliance *e dei poteri di autonormazione degli operatori*, 52 ss.; S. Sabia, L'enforcement *pubblico del* Digital Services Act *tra Stati membri e Commissione europea: implementazione, monitoraggio e sanzioni*, 88 ss., tutti in questa *Rivista*, 2, 2023. Nella prospettiva di una virtuosa collaborazione fra potere pubblico e piattaforme digitali, nell'ottica di contrastare il fenomeno della disinformazione in rete, si muove lo scritto di G.E. Vigevani, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritto costituzionale. Rivista quadrimestrale*, 1, 2023, 41 ss.

108 La stessa cosa vale per la disabilitazione dell'accesso agli stessi.

109 Cfr. il punto 22 dei *Considerando*.

110 Cfr. il punto 47 dei *Considerando*.

111 Sul pluralismo informativo, anche nell'ottica della sua distinzione "pluralismo esterno"-"pluralismo interno" (su cui cfr. Corte cost., 14 luglio 1988, n. 826), cfr. M. Manetti, *Pluralismo dell'informazione e libertà di scelta*, in *Rivista AIC*, 1, 2012, 1 ss.; M. Cuniberti, *Pluralismo dei media, libertà di espressione e "qualità" della*

di garantire un'informazione libera e quanto più completa possibile, corollario, in un certo qual senso, della democraticità del sistema[112] e del più generale principio pluralista su cui gli stessi sistemi democratici si fondano[113], alla cui salvaguardia – con riferimento, nel caso di specie, in particolare, della rete – può senz'altro contribuire lo strumento concorrenziale.

In questo senso, assumono particolare rilevanza le disposizioni di cui agli artt. 15 e 16, rispettivamente in tema di "Obblighi in materia di relazioni di trasparenza per i prestatori di servizi intermediari" e di "Meccanismi di segnalazione e azione".

Con la prima disposizione si sancisce l'obbligo, per i prestatori di servizi intermediari, di mettere a diposizione del pubblico, secondo un formato che sia leggibile meccanicamente ed in modo che ne risulti facile il relativo accesso, almeno una volta l'anno, relazioni chiare e di agevole comprensibilità sulle attività di moderazione dei contenuti svolte durante il periodo di riferimento. La norma prevede, poi, un elenco delle informazioni che tali relazioni devono contenere e che variano a seconda dei casi.

Attraverso l'art. 16, invece, si prevede la predisposizioni di taluni meccanismi, ad opera dei prestatori di servizi di memorizzazione di informazioni, in grado di consentire a qualsiasi persona (o a enti) di notificare loro la presenza nel servizio da loro stessi offerto di informazioni specifiche che si ritengono portatrici di contenuti di carattere illegale. Il legislatore elenca, quindi, le misure necessarie che i prestatori di servizi di memorizzazione di informazione devono adottare per poter così consentire nonché facilitare la presentazione delle relative segnalazioni.

Il rischio sotteso ad una impostazione di questo tipo è, in particolare, quello di un utilizzo censorio delle piattaforme giustificato dalla necessità di contrastare il fenomeno delle *fake news*[114] e dei contenuti offensivi e/o illeciti[115], con l'avallo, in tal caso, del legislatore europeo, il quale sembra risolvere in maniera – per così dire – in parte eccessivamente frettolosa ed anche semplicistica, una questione assai delicata, informando di giuridicità (con la conseguenza di rendere lecito) il potere delle piattaforme

legislazione: il caso "Centro Europa 7" di fronte alla Corte Europea dei Diritti dell'Uomo, in *Rivista AIC*, 3, 2012, 5 ss.; O. Pollicino, *Tutela del pluralismo nell'era del digitale: ruolo e responsabilità degli* Internet service provider, in *Percorsi costituzionali*, 1, 2014, 45 ss.; G. Avanzini-G. Matucci (a cura di), *L'informazione e le sue regole. Libertà, pluralismo e trasparenza*, Napoli, 2016; R. Borrello, *Alcune riflessioni preliminari (e provvisorie) sui rapporti tra i motori di ricerca ed il pluralismo informativo*, in questa *Rivista*, 1, 2017, 68 ss.; G.E. Vigevani, *I media di servizio pubblico nell'età della rete. Verso un nuovo fondamento costituzionale, tra autonomia e pluralismo*, Torino, 2018.

[112] In merito rimane attuale S. Rodotà, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Roma-Bari, 1997.

[113] In proposito cfr. F.R. De Martino, *L'attualità del principio pluralista come problema*, in *Rivista AIC*, 2, 2019, 569 ss.

[114] Fenomeno che con Internet è indubbiamente aumentato. Spiega le ragioni di ciò G. Pitruzzella, *La libertà di informazione nell'era di Internet*, cit., 30 ss., dove si illustra, altresì, perché con la rete le notizie false acquistino maggiore rilevanza. La prospettiva che risulta più condivisibile, ad avviso di F. Pizzetti, Fake news *e allarme sociale: responsabilità, non censura*, in questa *Rivista*, 1, 2017, 48 ss., è quella di una maggiore responsabilizzazione degli utenti (affinché essi si possano autoproteggere) anziché quella di ricorrere ad una censura (di matrice privatistica o pubblicistica che sia).

[115] «Non si può trascurare […] che il DSA inizia a porre un freno all'uso illegittimo, dannoso, non solo penalmente rilevante, ma anche distorto dell'informazione online che può produrre effetti manipolatori, limitando di fatto il diritto all'informazione», così V. De Santis, *Identità e persona all'epoca dell'intelligenza artificiale: riflessioni a partire dall'*IA act, in *Federalismi.it*, 19, 2024, 150.

di rimuovere contenuti[116] che potrebbero non essere, in realtà, falsi[117] ovvero illeciti ma, semplicemente, sgraditi alla narrazione in quel dato momento storico dominante. Ciò, infatti, potrebbe condurre ad una sorta di co-censura: privata e pubblica. Alle ipotesi di censura da parte delle piattaforme, si aggiungono quelle che derivano da una segnalazione ad opera dei pubblici poteri, aumentando così le ipotesi di rimozione dei contenuti digitali. Se questa prospettiva comporta indubbiamente un rafforzamento sul piano del contrasto alle notizie false ed ai contenuti offensivi e/o illeciti – oltre a poter essere supportata dal rinvenimento di un dovere in tal senso in capo agli stessi pubblici poteri, in particolare con riferimento alle prime[118] –, determina, tuttavia, al contempo, un aumento del rischio di un maggior ricorso a misure di rimozione, le quali, se non opportunamente valutate e calibrate, rischiano, appunto, di tradursi in una restrizione degli stessi spazi di operatività della libertà di manifestazione del pensiero.

Un illustre giurista affermava qualche anno fa che «vi è oggi un oggetto del desiderio degli Stati è esattamente Internet, con interventi molto pesanti, continui, tanto negli Stati autoritari che in quelli democratici. Si può, dunque, comprendere una resistenza volta a garantire il carattere libertario, persino anarchico, di Internet»[119]. Queste parole parrebbero tornare di attualità. Al momento, lo si ripete, è un rischio solamente potenziale che, tuttavia, deve essere scongiurato, evitando di configurare una regolazione eccessivamente intrusiva che comprima di conseguenza lo spazio di libertà, a partire da quella di manifestazione del pensiero. Non è un caso che si tratti di un'osservazione di dieci anni fa quando i poteri privati digitali, pur esistenti e già forti, non avevano ancora acquisito lo strapotere che hanno acquisito più di recente (in termini, anzitutto, di fatturato e di controllo del mercato); in ogni caso, l'affermazione conserva attualità poiché indica un pericolo possibile da tenere ben presente affinché si adottino le contromisure necessarie per evitare che si verifichi in concreto.

Da questo punto di vista, si tratta, dunque, di prospettare talune possibili soluzioni acciocché la maggiore regolazione del digitale non comporti una diminuzione di quello che è un diritto fondamentale, il quale, da una situazione di limitazione ad opera dei

[116] Il potere delle piattaforme di rimuovere i contenuti trova la sua base all'art. 3, lett. *t*). Infatti, con l'espressione «moderazione dei contenuti» il legislatore intende «le attività, automatizzate o meno, svolte dai prestatori di servizi intermediari con il fine, in particolare, di individuare, identificare e contrastare contenuti illegali e informazioni incompatibili con le condizioni generali, forniti dai destinatari del servizio, comprese le misure adottate che incidono sulla disponibilità, sulla visibilità e sull'accessibilità di tali contenuti illegali o informazioni, quali la loro retrocessione, demonetizzazione o rimozione o la disabilitazione dell'accesso agli stessi, o che incidono sulla capacità dei destinatari del servizio di fornire tali informazioni, quali la cessazione o la sospensione dell'account di un destinatario del servizio». Di "rimozione" si parlava già nella parte dei *Considerando*, come visto (il riferimento è al punto 22 degli stessi).

[117] Sulla questione – assai complicata nonché delicata – circa la possibilità di distinguere le *fake news* dalle opinioni ragiona G. Pitruzzella, *La libertà di informazione nell'era di Internet*, cit., 32 ss.

[118] O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione*, cit., 81, a detta del quale, ragionando nell'ottica propria del costituzionalismo europeo, si può, forse, ritenere «doveroso, soprattutto richiamandosi alla scarsità dell'attenzione da parte degli utenti e all'esigenza costituzionale di una loro corretta informazione, un intervento dei pubblici poteri, quale che ne sia la forma, volta a reprimere la circolazione delle false notizie».

[119] S. Rodotà, *Recensione a Giovanna De Minico. Internet. Regola e anarchia*, Napoli, 2012, in *Diritto pubblico*, 1, 2014, 361.

poteri privati, potrebbe attraversare una fase dove quella stessa limitazione è perpetrata o, comunque, in qualche modo agevolata, anche dai pubblici poteri, con la conseguenza di addivenire al medesimo risultato: la restrizione di una libertà che costituisce – e che deve continuare a costituire – uno dei pilastri delle attuali democrazie liberali, la cui tutela è proporzionalmente legata al grado di democraticità dello stesso ordinamento.

9. Questioni insolute e (possibili) prospettive future

Il fenomeno del digitale presenta una serie di aspetti problematici. Ciò detto, conviene anzitutto – anche alla luce di quanto osservato nelle pagine precedenti – impostare il discorso su due piani distinti. Il primo concerne il profilo relativo alla regolazione, il secondo quello, ad esso connesso, che ha ad oggetto la libertà di manifestazione del pensiero.

In ordine al primo, nelle pagine che precedono si è tentato di mettere in luce la necessità di superare il tradizionale modello regolatorio, pena il rischio di "Achille e la tartaruga". I forti poteri privati digitali, infatti, grazie al loro potere economico, sono in grado di dotarsi di tecnologie all'avanguardia ed il diritto con la sua regolazione poco intrusiva, che risente della evidente nonché marcata asimmetria informativa (chi dovrebbe regolare, cioè i pubblici poteri, spesso non hanno in realtà idea di che cosa dovrebbero in concreto fare), finisce per trovarsi perennemente costretto ad inseguire le novità, lasciando alle piattaforme digitali un potere assai ampio nonché pericoloso poiché, come visto, è da esso che, talvolta, dipende, in concreto, la garanzia e l'effettività di una serie di diritti fondamentali, a partire da quello di manifestazione del pensiero. Tuttavia, ciò non è comunque sufficiente. Il problema, invero, è sempre lo stesso: come rendere applicabili le regole e, quindi, in tal caso, come configurare un'etero-regolazione pubblica realmente effettiva.

Il principale aspetto problematico, come noto, è, e continua ad essere, quello che ruota attorno alla forte attenuazione del vincolo territoriale a favore della a-territorialità[120] che caratterizza la rete e, con essa, della capacità regolatoria dei singoli poteri pubblici, a partire dagli Stati[121]. Si tratta di una questione complessa che necessita di soluzioni che debbono inevitabilmente essere pensate su scala globale in quanto – pur essendo essenziali le regolazioni regionali tipo quella da ultimo adottata dal legislatore europeo con il *DMA* ed il *DSA* –, in ragione proprio del labile vincolo territoriale che caratterizza le piattaforme digitali, esse potrebbero comunque eludere parte delle prescrizioni attraverso una serie di *escamotage* volti a far risultare quel vincolo territoriale esistente rispetto ad un Paese che offre una regolamentazione più favorevole per i loro interessi (configurando un fenomeno analogo a quello dell'elusione fiscale). Solamente mediante una serie di accordi e convenzioni internazionali si può pensare di apprestare una regolazione pubblica delle piattaforme digitali in grado di attenuare il loro strapotere e conformando in senso concorrenziale i relativi mercati, attualmente a carattere oli-

[120] In proposito cfr. C. Napoli, *Territorio, globalizzazione, spazi virtuali*, in *La Rivista del Gruppo di Pisa*, 2, 2021, 204 ss.

[121] Sul punto cfr. M. Betzu, *Poteri pubblici e poteri privati*, cit., 168.

gopolistico.
Per quanto riguarda, viceversa, il secondo piano del discorso, l'attenzione si sposta sulla libertà di manifestazione del pensiero nonché sul problema di renderne effettivo il suo contenuto. Il rischio – lo si è già detto – è che l'aumento di regolazione pubblica non risolva il problema attualmente esistente circa il non sempre pieno rispetto di questa libertà in ragione del potere censorio esercitato spesso in maniera arbitraria dai "padroni" del digitale. Il problema, però, è come evitare tale rischio.
Una soluzione potrebbe, forse, essere quella di creare delle commissioni formate da personalità (come, per esempio, giuristi) esperte della rete e dei c.d. diritti digitali, dotata del necessario tasso di autonomia ed indipendenza, fatta sempre salva la possibilità di adire il giudice. La facoltà di ciascun soggetto di adire il giudice a difesa di una propria situazione giuridicamente rilevante che si ritiene essere stata violata, costituisce un altro pilastro fondamentale sul quale si reggono le democrazie liberali e che non può trovare limitazione alcuna, a partire dalla rete, la quale – come già ricordato – è il luogo ove più si esercita ormai tale diritto di libertà. Del resto, è la stessa Carta costituzionale a riconoscere il diritto di azione laddove afferma che tutti «possono agire in giudizio per la tutela dei propri diritti e interessi legittimi» (art. 24 Cost.).
È vero che le piattaforme digitali sono a tutti gli effetti società private e, dunque, hanno la libertà di cancellare tutti i contenuti che credono ma, essendo ormai evidente come anche e soprattutto in rete si svolga la libertà di manifestazione del pensiero (e, quindi, si esprima la personalità di un numero elevatissimo di persone), non pare affatto azzardato richiedere un intervento dei pubblici poteri volto a configurare dei meccanismi per rendere più effettivo tale diritto e, soprattutto, per renderne effettiva la possibilità di una sua tutela allorquando si ritiene che vi sia stata una sua violazione. Si potrebbe dire – anche qui utilizzando una tradizionale dizione non più particolarmente frequente – che la rete finisce per rivestire anche una "funzione sociale" che, come tale, non può essere estranea al diritto ed alla sua regolazione e, dunque, l'intervento pubblico è giustificato dall'esigenza di proteggere e tutelare le libertà individuali ed è in quelle stesse libertà (come quella di manifestazione del pensiero) che rinviene, al tempo stesso, sia la fonte legittimante il suo intervento che il suo limite.
La regolazione pubblica, dunque, dovrebbe preoccuparsi anche di come tutelare i diritti e le libertà in rete e assicurare meccanismi giurisdizionali o simil-giurisdizionali per risolvere le relative controversie, non essendo tollerabile in un ordinamento che si fonda sul concetto di "Stato di diritto" che il contenuto, i limiti e, quindi, più in generale, l'effettività di libertà così centrali nella vita di ciascuna persona, come ad esempio quella che fa capo alla manifestazione del pensiero, finisca per essere rimessa a valutazioni di soggetti estranei alla giurisdizione, i quali tendono a risolvere le controversie non mediante l'applicazione di principi e regole giuridiche bensì sulla base valutazioni al cui fondamento vi sono, il più delle volte, interessi di natura economica o, comunque, valutazioni di mera egoistica opportunità. Anche in tal caso, è evidente come le soluzioni vadano ricercate avendo quale orizzonte la prospettiva globale.
La regolazione, inoltre, incontra ulteriori problemi che possono in questa sede essere solamente elencati e che rispetto ai quali, di nuovo, le soluzioni da intraprendere debbono inevitabilmente collocarsi sul piano di una regolazione a carattere globale. Per

esempio, assume rilievo il tema di una tassazione più elevata dei colossi del digitale, la quale, tuttavia, per poter essere effettiva, deve essere uniforme su scala, appunto, globale, altrimenti si verificheranno fenomeni conosciuti come quello della migrazione verso i ben noti (e sempre più numerosi) "paradisi fiscali"; ciò presuppone, quindi, anche in tal caso, la stipulazione di una serie di accordi internazionali.
Inoltre, la sola regolazione delle piattaforme digitali non è sufficiente a contrastare i forti poteri privati[122], rendendosi pertanto necessaria anche una massiccia produzione, da parte degli Stati, di tecnologie (una sorta di "Stato produttore di tecnologia"), con significativi investimenti pubblici finalizzati alla riduzione della dipendenza che i pubblici poteri hanno nei riguardi dei privati, con ciò che ne consegue in punto di possibilità di essere in qualche modo ricattabili e, in ogni caso, non autonomi.

10. A mo' di conclusione

Si è avuto modo di osservare come l'auto-regolazione delle piattaforme digitali e, quindi, più in generale, del mondo digitale, sia insufficiente per una serie di ragioni e, soprattutto, determini un accentramento del potere di mercato nelle mani di pochi poteri privati, dando luogo ad un sistema di mercato oligopolistico, contrario per definizione al regime concorrenziale su cui si fonda, invece, l'ordinamento europeo.
Del resto, l'auto-regolazione favorisce l'emersione del diritto privato, il quale, tuttavia, funziona ed è in grado di regolare i rapporti[123] fra le parti quando vi è una parità (di partenza) delle stesse, nel momento in cui, cioè, sussiste una consustanziale autodeterminazione delle parti in gioco, poiché tale autodeterminazione è presupposta dallo stesso mondo privatistico, il quale, prefiggendosi di regolare rapporti *inter pares*, non è più adatto – o, comunque, di per sé sufficiente – allorché fra le parti non sussista una parità bensì una disparità e, sotto questo profilo, poco importa se il soggetto più forte non è quello pubblico quanto quello privato, cambia l'aggettivo ma non il sostantivo e la disparità si ha in presenza del primo (del sostantivo, ossia del "potere"), non del secondo (dell'aggettivo, "pubblico" o "privato" che sia). Da questo punto di vista, è necessario abbandonare l'idea per cui il privato sia sinonimo di "libertà" ed il pubblico di "autorità"; invero, quest'ultimo non è solo e necessariamente "autorità" ed il primo non è sempre e comunque equivalente di "libertà".
Tuttavia, anche il modello tradizionale di regolazione del mercato – alternativo a quello dell'auto-regolazione – che si è andato configurando a livello europeo (ma, *mutatis mutandis*, il discorso, in linea generale, vale anche per quello di marca statunitense) tende ad essere costruito attorno ad una regolazione *ex post*, che interviene dopo aver preso atto, nella realtà fattuale, del verificarsi di fenomeni incompatibili, in concreto o in potenza, con la necessità di un mercato concorrenziale. Fra le varie ragioni per cui

[122] Compito del costituzionalismo – oltre a quello di limitare il potere politico – è quello di limitare il potere economico dei soggetti privati, in special modo allorché esso, per le dimensioni raggiunte, rischi di mettere a repentaglio diritti e libertà costituzionalmente riconosciute e garantite; sul punto si rimanda ancora una volta a M. Betzu, *I baroni del digitale*, cit., *passim*, e F. Paruzzo*, I sovrani della rete*, cit., *passim.*

[123] Sul diritto privato regolatorio cfr. R. Natoli, *Il diritto privato regolatorio*, in *Rivista della regolazione e dei mercati*, 1, 2020, 134 ss.

questo modello non è pienamente esportabile nel mercato digitale, vi è quella per cui esso non è in grado di far fronte ad un problema centrale di tale ambito, che consiste in un progresso tecnologico così veloce e sofisticato che, a maggior ragione, necessita più di altri settori di mercato di una regolazione che intervenga anche e, possibilmente, soprattutto, *ex ante*, una sorta cioè di "regolazione anticipante" anzitutto i rischi ed i pericoli che tale innovazione tecnologica determina in punto di tutela di diritti fondamentali (dai dati personali alla libertà di manifestazione del pensiero), di quegli stessi diritti fondamentali e di quelle medesime libertà che la storia del costituzionalismo liberale si è sempre preoccupata di riconoscere e, dunque, tutelare[124].

Il diritto antitrust – rispetto al quale i poteri privati digitali mostrano una certa insofferenza poiché «più che competere nel mercato, fanno concorrenza al mercato, ponendosi «cioè essi stessi come mercato, come infrastruttura tecnologica all'infuori della quale non ci può essere concorrenza»[125] – ha, inoltre (in special modo quello di stampo americano), una pericolosa vocazione protesa alla protezione, più che dei concorrenti e/o dei consumatori, della concorrenza in sé[126], presentando tratti che non sempre appaiono pienamente conformi a quell'idea della persona come centro da cui muovere, dove il diritto è «al servizio dell'uomo»[127], e non viceversa.

La regolazione che sembra emergere e che appare auspicabile, si è detto, è una co-regolazione flessibile (assolutamente necessaria) ed *ex ante*, che sembra trovare – quantomeno in buona parte – fondamento nel *DMA*, il quale si prefigge di introdurre una serie di criteri oggettivi volti a definire le piattaforme *online* di grandi dimensioni che esercitano una funzione di controllo dell'accesso al mercato (*gatekeeper*), con una serie di obblighi nonché di vincoli che possano garantire agli innovatori del "domani" di trovare un proprio spazio e di poter, quindi, competere fra loro liberamente.

È una regolazione anche "forte", incisiva, che è strutturata inevitabilmente nei termini di una regolazione multilivello, dove si è vista l'importanza della cooperazione e, dunque, l'esigenza del coordinamento fra i vari soggetti (pubblici e privati) deputati alla regolazione del fenomeno digitale.

In tale modello si è, altresì, tentato di riflettere attorno alla concorrenza ed al ruolo che essa dovrebbe assumere, la quale non dovrebbe (più) rappresentare il contenuto dell'interesse pubblico (quale valore in sé o, comunque, teso ad identificarsi nell'efficienza del mercato) ma uno strumento per perseguire altri interessi, come la protezione dei dati personali e la libertà di manifestazione del pensiero, i quali, nel modello di regolazione qui avanzato, dovrebbero, dunque, costituire il contenuto dell'interesse pubblico

[124] Sulle sfide che si trova ad affrontare il costituzionalismo nell'epoca della società digitale, con particolare riferimento alla garanzia dei diritti fondamentali, cfr. O. Pollicino, *L'"autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete*, cit., 11.

[125] E. Cremona, *L'erompere dei poteri privati nei mercati digitali*, cit., 901.

[126] Ivi, 888. Sulla compatibilità o meno del fine della tutela del mercato in sé e di quello dei consumatori ad opera del principio di concorrenza, cfr. G. Amato, *Tutela della concorrenza o tutela dei consumatori. Due fini confliggenti?*, in *Mercato, concorrenza, regole*, 2, 2009, 381 ss. Sul punto cfr. anche G. Repetto, *Efficienza economica, libertà e tutela dai poteri privati: a cosa serve il principio di libera concorrenza?*, in *Diritto costituzionale. Rivista quadrimestrale*, 1, 2021, 112 ss.

[127] Per riprendere un'espressione utilizzata da Giovanni Miele nel suo fondamentale scritto *Umanesimo giuridico*, in *Rivista di diritto commerciale*, 1, 1945, 103 ss.

(primario) di partenza o – se si preferisce utilizzare una dizione più tradizionale – del c.d. "interesse pubblico in astratto"[128]. Inoltre, la concorrenza può rivestire particolare importanza nell'ottica di assicurare quel pluralismo informativo di cui si è detto e dal quale non si può certo prescindere, rappresentando un valore da salvaguardare poiché strettamente legato alla democrazia stessa della rete.

Sempre ponendo particolare attenzione al rischio che la maggiore regolazione pubblica determini una contrazione dello spazio della libertà di espressione[129] – la quale rileva anche come limite alla forza del modello di regolazione qui prospettato –, per mezzo, cioè, di un controllo autoritario da parte pubblici poteri in funzione liberticida, è comunque necessario configurare un governo del fenomeno digitale dove, quindi, un ruolo centrale è destinato ad essere assunto dagli stessi pubblici poteri in qualità di rappresentanti democratici della volontà popolare, con una funzionalizzazione del potere pubblico verso la difesa dei diritti[130] e delle libertà che vengono in rilievo nel mondo digitale e, per questo, legittimo. Del resto, «vuoto concetto sarebbe la sovranità popolare, se venisse spogliata della potestà di determinare, per mezzo dei propri rappresentanti democraticamente eletti, l'indirizzo politico generale»[131], il quale contempla anche e soprattutto l'ambito del digitale, che rappresenta, a tutti gli effetti, una delle proiezioni dello sviluppo della personalità di ciascun essere umano, la quale trova in tale ambito un luogo di ulteriore emancipazione e confronto, dove dare avvio ad una libera e sana "battaglia delle idee", linfa vitale per il progresso individuale e collettivo.

Una maggiore regolazione pubblica è, dunque, auspicabile e, si potrebbe aggiungere, doverosa in quanto strumentale ad un governo pubblico (e non privato) dei fenomeni della società e necessario per l'effettiva attuazione ed affermazione dei diritti individuali, coerentemente con il carattere personalista fondamento dell'ordinamento europeo e, più in generale, del diritto liberale alla base delle attuali democrazie. Un diritto, dunque, «non come infrastruttura nelle mani di pochi, ma come scienza legata ai valori e ai diritti di una rule of law capace di incidere sulla forza del capitale e delle imprese»[132], al fine di salvaguardare, in ogni ambito, la persona, la cui tutela costituisce il fine ultimo di ogni ordinamento giuridico.

---

[128] M.S. Giannini, *Il pubblico potere*, Milano, 1986 (dello stesso A. cfr. anche *Istituzioni di diritto amministrativo*, I, Milano, 1993, 75).

[129] Sul punto si rimanda ancora a M. Manetti, *Internet e i nuovi pericoli per la libertà di informazione*, cit., 523 ss.

[130] In una prospettiva che è necessariamente di tutela multilivello. Sulla tutela multilivello dei diritti fondamentali, per un inquadramento generale, cfr. A. Cardone, *La tutela multilivello dei diritti fondamentali*, Milano, 2012.

[131] M. Betzu, *Poteri pubblici e poteri privati*, cit., 170.

[132] G. Vettori, *Sui poteri privati*, cit., 831.

Note a sentenza

La democrazia italiana di fronte al saluto romano. Alcune note a margine di Cass. sez. un. n. 16153 del 2024*

Bruno Pitingolo

Corte di Cassazione, sez. un. penali, 17 aprile 2024

La condotta tenuta nel corso di una pubblica manifestazione consistente nella risposta alla "chiamata del presente" e nel c.d. "saluto romano", rituali entrambi evocativi della gestualità propria del disciolto partito fascista, integra il delitto previsto dall'art. 5 della legge 20 giugno 1952, n. 645, ove, avuto riguardo a tutte le circostanze del caso, sia idonea ad integrare il concreto pericolo di riorganizzazione del disciolto partito fascista, vietata dalla XII disposizione transitoria e finale della Costituzione. A determinate condizioni può configurarsi anche il delitto previsto dall'art. 2 del decreto-legge 26 aprile 1983, convertito, con modificazioni, nella legge 25 giugno 1993, n. 205 che vieta il compimento di manifestazioni esteriori proprie o usuali di organizzazioni, associazioni, movimenti o gruppi che hanno tra i propri scopi l'incitamento alla discriminazione o alla violenza per motivi razziali, etnici, nazionali o religiosi. Tra i due delitti non sussiste rapporto di specialità e possono concorrere sia materialmente che formalmente in presenza dei presupposti di legge.

Sommario

Keywords

saluto romano – apologia al fascismo – libertà di espressione – libertà costituzionali – democrazia italiana.

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

1. Premessa

Il 7 gennaio 2025, circa mille persone hanno sfilato in corteo, esibendo il saluto romano e rispondendo alla "chiamata del presente", in commemorazione di tre giovani militanti del Movimento Sociale Italiano, deceduti quarantasette anni or sono in via Acca Larentia a Roma[1].

La vicenda – non del tutto nuova in verità – ha sollevato ulteriori interrogativi in ordine alla legittimità di comportamenti, di chiara ispirazione fascista, all'interno dell'ordinamento democratico, suscitando alcune perduranti riflessioni circa i limiti che presiedono all'esercizio delle libertà costituzionali.

La ricerca di possibili soluzioni al problema si rende necessaria attraverso l'utilizzo di strumenti di carattere normativo, in luogo di decisioni politico-emotive che potrebbero eventualmente anche condurre ad un maggiore desiderio di inasprimento punitivo[2].

A questo riguardo, il quadro normativo concernente l'apologia del fascismo e le relative condotte si fonda primariamente a livello costituzionale: la XII disposizione transitoria e finale, vieta espressamente la «riorganizzazione, sotto qualsiasi forma, del disciolto partito fascista».

A tale previsione si sono poi affiancate nel corso del tempo la legge 20 giugno 1952, n. 645 (nota come legge Scelba), che ne attua direttamente i principi, e le fattispecie incriminatrici relative alla discriminazione razziale, introdotte dalla legge 13 ottobre 1975, n. 654 (cosiddetta legge Reale), successivamente modificate dalla legge del 25 giugno 1993, n. 205 (denominata legge Mancino), recante "Misure urgenti in materia di discriminazione razziale, etnica e religiosa"[3].

---

[1] Il 7 gennaio 1978, davanti alla sede del Movimento Sociale Italiano, furono assassinati due giovani attivisti del Fronte della Gioventù, Franco Bigonzetti e Francesco Ciavatta. Un terzo militante, Stefano Recchioni, perse la vita poche ore dopo durante gli scontri con le forze dell'ordine nello stesso luogo. Questi eventi sono noti come la Strage di Acca Larentia, avvenuta a Roma nel quartiere Tuscolano. Da allora, il luogo è diventato un simbolo per le manifestazioni neofasciste di militanti, o solo nostalgici, tutti schierati per invocare la "chiamata al presente" e la liturgia fascista del saluto romano.

[2] Altrimenti descritto nei termini di un "*diritto penale dell'emotività*" da D. Piccione, *L'antifascismo e i limiti alla manifestazione del pensiero tra difesa della Costituzione e diritto penale dell'emotività*, in *Giur. cost*., 4, 2017. Da questo punto di vista, l'approccio privilegiato da preservare è quello dell'interprete del diritto; per dirla con le parole di F. Balaguer Callejon, sebbene le «grandi crisi del XXI secolo si collocano al di fuori del contesto culturale del diritto», occorre «superare questa difficoltà mostrando alla società l'importanza del diritto e l'impossibilità di rispondere alle esigenze del nostro tempo senza strumenti giuridici», v. A. Randazzo, *Intervista a Francisco Balaguer Callejon. La Costituzione dell'algoritmo*, in *Nomos. Le attualità del diritto*, 1, 2023, 2.

[3] La legge 13 ottobre 1975 n. 654, anche conosciuta come legge Reale, nel ratificare la Convenzione di New York sull'eliminazione di tutte le forme di discriminazione razziale del 21 dicembre 1965, ha introdotto una disciplina penale nel nostro ordinamento contro le condotte di discriminazione razziale. Secondo quanto affermato dall'art. 3 della l. 654/1975, «Gli Stati contraenti condannano in particolar modo la segregazione razziale e l'"apartheid" e si impegnano a prevenire, vietare ed eliminare sui territori sottoposti alla loro giurisdizione, tutte le pratiche di tale natura». Questa disposizione è stata successivamente convertita, con modifiche, nella legge 25 giugno 1993, n. 205 (c.d. legge Mancino), il cui art. 2 afferma che «Chiunque, in pubbliche riunioni, compia manifestazioni esteriori od ostenti emblemi o simboli propri o usuali delle organizzazioni, associazioni, movimenti o gruppi di cui all'articolo 3 della legge 13 ottobre 1975, n. 654, è punito con la pena della reclusione fino a tre anni e con la multa da lire duecentomila a lire cinquecentomila».
Il d.lgs. 1° marzo 2018 n. 21 ha in seguito abrogato l'art. 3 della legge 13 ottobre 1975 n. 654 e trasfuso la normativa all'art. 604-*bis* che vieta «ogni organizzazione, associazione, movimento o gruppo avente

In tale contesto, la sentenza delle Sezioni Unite della Cassazione n. 16153 del 18 gennaio 2024 si appunta sul significato giuridico che il saluto romano dispiega all'interno del nostro ordinamento, suscitando un interesse che va oltre la necessità di un'interpretazione uniforme della fattispecie e che mette in luce una delle aspirazioni più complesse dei moderni sistemi democratico-pluralistici[4].
La decisione in commento tenta infatti di raggiungere un equilibrio sostenibile tra la tutela dei diritti individuali fondamentali e la connotazione antifascista della Costituzione, quale fondamento ineludibile della Repubblica[5]. Ed è proprio quest'ultima esigenza a richiedere un'attenzione particolare nei confronti di quelle manifestazioni di carattere rievocativo, che sovente si presentano come espressioni di aperto dissenso nei riguardi del patto costituzionale repubblicano.

2. Il caso all'esame delle Sezioni Unite

La decisione della Suprema Corte di cassazione origina dal ricorso proposto avverso la sentenza emessa dalla Corte d'appello di Milano, in data 24 novembre 2022. Con tale decisione il giudice di secondo grado aveva condannato gli imputati, ai sensi della c.d. "legge Mancino", a due mesi di reclusione e a duecento euro di multa, «per aver partecipato alla manifestazione esteriore di un'organizzazione, un movimento o un gruppo che incita alla discriminazione e alla violenza per motivi razziali, etnici, nazionali o religiosi»[6].

tra i propri scopi l'incitamento alla discriminazione o alla violenza per motivi razziali, etnici, nazionali o religiosi».

[4] È tale la prerogativa tradizionalmente attribuita alla Corte di cassazione nell'esercizio della c.d. funzione "nomofilattica" o di "nomofilachia", «cioè la funzione di unificare e coordinare l'interpretazione ed applicazione delle norme, al fine di garantire un'omogenea evoluzione della giurisprudenza» (sul punto cfr., F. P. Luiso, *Diritto processuale civile Volume II: Il processo di cognizione*, Milano, 2019, 428). Essa, inoltre, è ben delineata dall'art. 65 del Regio decreto 30 gennaio 1941, n. 12 sull'ordinamento giudiziario, laddove si dice che «La corte suprema di cassazione, quale organo supremo della giustizia, assicura l'esatta osservanza e l'uniforme interpretazione della legge, l'unità, del diritto oggettivo nazionale».

[5] Piace a questo proposito prendere ad esempio le parole del presidente emerito della Corte costituzionale, Giovanni Maria Flick, secondo il quale «La Costituzione è un patto che esprime la pari dignità sociale, l'eguaglianza e la diversità, la solidarietà… È un patto di reciprocità fra i diritti e i doveri; un patto di garanzia dei diritti inviolabili dei singoli, in sé e nelle formazioni sociali ove si svolge la loro personalità. È un patto che guarda al futuro facendo tesoro della memoria del passato; un patto di inclusione e di partecipazione, non di esclusione e di appartenenza; un vero e proprio manuale di convivenza» (G. M. Flick, *La Costituzione: un manuale di convivenza. Antologia di scritti su Costituzione, dignità, patrimonio*, Milano, 2018).

[6] All'art. 3 l. 654/1975, «Salvo che il fatto costituisca più grave reato, anche ai fini dell'attuazione della disposizione dell'articolo 4 della convenzione, è punito: a) con la reclusione sino a tre anni chi diffonde in qualsiasi modo idee fondate sulla superiorità o sull'odio razziale o etnico, ovvero incita a commettere o commette atti di discriminazione per motivi razziali, etnici, nazionali o religiosi; b) con la reclusione da sei mesi a quattro anni chi, in qualsiasi modo, incita a commettere o commette violenza o atti di provocazione alla violenza per motivi razziali, etnici, nazionali o religiosi; è vietata ogni organizzazione, associazione, movimento o gruppo avente tra i propri scopi l'incitamento alla discriminazione o alla violenza per motivi razziali, etnici, nazionali o religiosi. Chi partecipa a tali organizzazioni, associazioni, movimenti o gruppi, o presta assistenza alla loro attività, è punito, per il solo fatto della partecipazione o dell'assistenza, con la reclusione da sei mesi a quattro anni. Coloro che promuovono o dirigono tali organizzazioni, associazioni, movimenti o gruppi sono puniti, per ciò solo, con la reclusione da uno a

Più in particolare, durante una manifestazione autorizzata, gli imputati avevano esposto uno striscione inneggiante ai "camerati caduti", Enrico Pedenovi, Sergio Ramelli e Carlo Borsani, rispondendo poi alla "chiamata del presente" ed effettuando contestualmente anche il rituale del saluto romano.

In accoglimento dell'impugnazione proposta dal Pubblico Ministero, la sentenza d'appello aveva riformato in senso peggiorativo la pronuncia dal Tribunale, che, in primo grado aveva assolto gli imputati, ritenendo che il fatto non costituisse reato. Nelle more di quel procedimento, il giudice di prime cure aveva ricostruito la fattispecie in esame ai sensi dell'art. 5 della l. 645/1952, ascrivendo il comportamento degli imputati ad una «manifestazione usuale del disciolto partito fascista». Tuttavia, il medesimo Tribunale aveva contestualmente rilevato la sussistenza di un errore scusabile ed aveva pertanto escluso la sussistenza del «pericolo di ricostituzione di organizzazione fasciste», anche sulla scorta di quanto precedentemente statuito dal Giudice dell'udienza preliminare, in relazione ad una manifestazione analoga, tenutasi nel 2014[7].

Per contro, la Corte d'appello di Milano, dopo avere evidenziato che la pubblica ostentazione di tali gesti avrebbe dovuto ritenersi «concretamente idonea alla propaganda e diffusione di idee fondate sulla superiorità o sull'odio razziale ed etnico», aveva condannato gli imputati, ritenendo sussistente l'elemento soggettivo del reato, sulla base della circostanza che – come ribadito anche dalla giurisprudenza di legittimità – nel caso del "saluto romano" debba ritenersi costantemente integrata l'ostentazione di un simbolo proprio di organizzazioni, movimenti o gruppi fascisti *ex* art. 2 della legge Mancino.

Ciò considerato, la prima sezione penale della Cassazione, con ordinanza del 6 settembre 2023, aveva rimesso il ricorso alle Sezioni Unite, rilevando un contrasto interpretativo in merito alla gestualità del saluto romano.

In effetti, secondo un primo orientamento, la condotta in questione integrerebbe il reato di cui all'art. 2 della legge Mancino, mentre, secondo un diverso indirizzo pretorio quel medesimo comportamento ricadrebbe entro la sfera applicativa dell'art. 5 della "legge Scelba"[8]. In estrema sintesi, quindi, la prima fattispecie assegnerebbe maggiore rilievo alla diffusione di idee fondate sulla superiorità e sull'odio razziale[9], mentre la seconda ricostruzione sarebbe tesa a scongiurare la ricostituzione del disciolto partito fascista, vietata in maniera tassativa dalla XII disposizione transitoria e finale della

sei anni».

[7] Nello specifico, il Giudice dell'udienza preliminare aveva escluso la sussistenza del pericolo motivando che la manifestazione si sarebbe svolta «in forma statica senza essere preceduta o seguita da alcun corteo» (*ritenuto in fatto*, punto 1).

[8] Si riportino, ai fini di una maggiore chiarezza espositiva, le due disposizioni incriminatrici di riferimento: «Chiunque con parole, gesti o in qualunque altro modo compie pubblicamente manifestazioni usuali al disciolto partito fascista è punito con l'arresto fino a tre mesi o con l'ammenda fino a lire cinquantamila» (l. 20 giugno 1952, n. 645, c.d. "legge Scelba"); «Chiunque, in pubbliche riunioni, compia manifestazioni esteriori od ostenti emblemi o simboli propri o usuali delle organizzazioni, associazioni, movimenti o gruppi di cui all'articolo 3 della legge 13 ottobre 1975, n. 654, è punito con la pena della reclusione fino a tre anni e con la multa da lire duecentomila a lire cinquecentomila» (l. 25 giugno 1993, n. 205, c.d. "legge Mancino")

[9] Le principali pronunce sostenitrici del primo orientamento richiamate dalla sentenza delle Sezioni Unite sono: Cass. pen., sez. I, n. 21409/2019; sez. I, n. 25184/2009; sez. III, n. 37390/2007.

Costituzione[10].
Attesa questa divergenza interpretativa, la prima sezione penale aveva perciò richiesto alle Sezioni Unite: «se la condotta tenuta nel corso di una pubblica manifestazione consistente nella risposta alla "chiamata del presente" e nel "saluto romano", rituali evocativi della gestualità propria del disciolto partito fascista, sia sussumibile nella fattispecie incriminatrice di cui all'art. 2 del decreto-legge 26 aprile 1993, n. 122, convertito con modificazioni, nella legge 25 giugno 1993, n. 205, ovvero in quella prevista dall'art. 5 della legge 20 giugno 1952, n. 645»[11].

3. La decisione delle Sezioni Unite n. 16153 del 18 gennaio 2024

Le SS. UU. della Cassazione sono state quindi chiamate a pronunciarsi in merito alla rilevanza penale del "saluto romano" ai sensi dell'art. 5 della legge n. 645 del 1952 e dell'art. 2 della legge n. 205 del 1993.
Accanto a tale questione, se ne affiancano altre due: la prima relativa alla configurabilità delle condotte previste dalla c.d. legge Scelba e dalla c.d. legge Mancino come reati di pericolo astratto oppure come reati di pericolo concreto; la seconda concernente invece il rapporto tra i due illeciti quali condotte tra loro effettivamente in concorso.
Con riguardo al quesito principale, il principio di diritto enunciato dalle Sezioni Unite afferma che la condotta, tenuta nel corso di una pubblica riunione, consistente nel cosiddetto "saluto romano" e nella risposta alla "chiamata del presente", integra il delitto previsto dall'art. 5 della legge n. 645/1952, solo ed esclusivamente se, avuto riguardo alle circostanze del caso, sia idonea ad attingere il concreto pericolo di riorganizzazione del disciolto partito fascista, vietata dalla XII disposizione transitoria e finale della Costituzione.
Qualora invece, tenuto significativamente conto del contesto fattuale complessivo, la medesima condotta sia espressiva di una manifestazione propria o usuale delle organizzazioni, delle associazioni, dei movimenti o dei gruppi di cui all'art. 604-*bis*, secondo comma, del codice penale, quello specifico comportamento potrebbe integrare il delit-

[10] Le principali pronunce sostenitrici del secondo orientamento richiamate dalla sentenza delle Sezioni Unite sono: Cass. pen., sez. V, n. 36162/2019; sez. I, n. 11038/2016; sez. I, n. 37577/2014).

[11] Nel trattare il contrasto giurisprudenziale, le Sezioni Unite ritengono di menzionare le sentt. 12 ottobre 2021, n. 7904 e 19 novembre 2021, n. 3806, in qualità di pronunce dotate di una motivazione più puntuale, rispetto alle altre, rispettivamente in adesione del primo e del secondo orientamento.
In particolare, la prima decisione spiega diffusamente che la fattispecie incriminatrice è l'art. 5 della Legge Scelba perché sanziona il pericolo di riorganizzazione "storica" del partito nazionale fascista, a differenza dell'art. 3 della legge Mancino che prescinde da questa caratteristica. Esso, infatti, riferendosi a un gruppo *ex* art. 3 l. 654/1975 esteriorizza manifestazioni razziali *latu sensu*. Mentre, la sentenza della sez. I, n. 3806/2021, ritenuta la pronuncia più "virtuosa" in favore dell'applicazione dell'art. 2 legge cit., è stata richiamata dalle Sezioni Unite per avere ricondotto il rapporto tra l'art. 5 e l'art. 2 alla stregua di un rapporto di specialità: più precisamente, vi sarebbe una presunzione *ex lege*, secondo cui, nelle manifestazioni razziali, discriminatorie, etc., vi sarebbe integrata l'ideologia fascista; tuttavia, l'art. 5 della Legge Scelba è applicabile solo ove vi sia un pericolo concreto di ricostituzione del disciolto partito fascista. Ne deriva che, a fronte della mancata constatazione di questo pericolo, deve ritenersi applicabile l'art. 2 della Legge Mancino.

to, di pericolo presunto, previsto dall'art. 2, c. 1, della legge n. 205/1993.
Secondo la Suprema Corte, dunque, la differenza fondamentale tra le due fattispecie risiede sia nella situazione di fatto in cui la manifestazione esteriore avviene sia nel bene giuridico che si assume violato. In proposito, con riferimento a quanto previsto dall'art. 5 della legge Scelba, il bene protetto sarebbe da individuarsi nella tutela dell'«ordine pubblico democratico»[12], da intendersi – conformemente con la giurisprudenza della Consulta – né come un interesse individuale, né come un mero ordine pubblico materiale, bensì come «ordine pubblico costituzionale»[13] in quanto «ordine legale costituito»[14], ovverosia come identità dello Stato democratico-repubblicano, che confluisce in un vero e proprio «sistema su cui poggia la convivenza sociale»[15].
Tuttavia, data la sua natura di reato di pericolo concreto, per l'applicazione della norma si richiede la sussistenza di «elementi di fatto», quali, a titolo esemplificativo, «il contesto ambientale, la eventuale valenza simbolica del luogo di verificazione, il grado di immediata ricollegabilità dello stesso contesto al periodo storico in oggetto e alla sua simbologia, il numero dei partecipanti, la ripetizione insistita dei gesti»[16].
Tale accertamento è strettamente connesso alla possibile lesione del bene giuridico che la normativa intende salvaguardare e che, espresso sotto forma di esplicito divieto, proibisce la ricostituzione del «disciolto partito fascista», in attuazione della XII disposizione transitoria e finale della Costituzione.
Nella differente ipotesi in cui non siano presenti elementi modali e temporali, tali da integrare la minaccia di cui all'art. 5 della legge n. 645/1952, la medesima condotta – soggiunge la sentenza – potrebbe analogamente configurare il reato di pericolo presunto, di cui all'art. 2 della "legge Mancino". In tale evenienza, il "saluto romano" potrebbe acquisire una propria autonoma rilevanza applicativa, anche in assenza di una concreta offesa nei confronti della Repubblica e delle sue Istituzioni.
Si tratta di un passaggio di particolare rilevanza, poiché, secondo i giudici di Cassazione, la condotta in oggetto potrebbe configurarsi come reato, anche in assenza di un intento diretto alla ricostituzione del disciolto partito fascista, in particolare qualora gruppi, movimenti o aggregazioni, anche di natura estemporanea, compiano atti caratterizzati da discriminazione razziale[17]. Questo si verificherebbe, per esempio, nel caso in cui l'offesa in questione si traduca in una lesione di un bene giuridico «composito», tutelato dalla legge n. 205 del 1993, e finalizzato a proteggere i diritti fondamentali

[12] *Considerato in diritto*, punti 6.2.1, e 6.2.2.

[13] Corte cost., sent. n. 168/1971

[14] Corte cost., sent. n. 87/1966

[15] Corte cost., sent. n. 19/1962. La pronuncia della Cassazione riconosce espressamente il "saluto romano" come una *«naturale» identificazione e richiama anche gli artt. 3 e 9 del regolamento del PNF per affermare che la gestualità è tipica della «liturgia delle adunanze fasciste»,* v. *Considerato in diritto*, punto 8.

[16] *Considerato in diritto*, punto 8.1.

[17] Come spiega la Cassazione, «non è neppure necessario (…) un loro inquadramento in entità espressamente operanti sotto un nome, ovvero dotate di uno statuto (…), ben potendo trattarsi anche di aggregazioni di natura estemporanea, come desumibile dal tenore letterale della norma» (*Considerato in diritto*, punto 9). Così anche S. Curreri, *Costituzione e neo-fascismo: quando il saluto fascista è reato*?, in questa *Rivista*, 1, 2024, 131 ss.

garantiti, tra gli altri, dagli articoli 2 e 3 della Costituzione[18].
Secondo il ragionamento della Corte, pertanto, la condotta posta in essere dagli imputati potrebbe configurarsi anche come delitto ai sensi dell'articolo 2 della legge Mancino, laddove costituisca «lo strumento simbolico di espressione di idee di intolleranza o di discriminazione attualmente proprie degli agglomerati considerati dall'articolo 3 della legge n. 654 del 1975»[19].
Questa duplice possibilità di qualificazione del delitto in questione comporterebbe un restringimento o un ampliamento dell'area penalmente rilevante[20], in grado di illuminare quella "zona grigia" che la giurisprudenza tende a inquadrare con maggiore difficoltà, ossia la rilevanza giuridica del "saluto romano" in contesti quali chiese, sagrati, cimiteri, stadi, consigli comunali, scuole, e così via[21].
Ebbene, con la pronuncia in commento tali esternazioni sarebbero da ricondursi all'ambito applicativo della legge Mancino, mancando in esse gli elementi sintomatici del pericolo di ricostituzione del disciolto PNF, ma connotandosi esteriormente per la manifestazione predominante di idee di natura razziale.
La perseguibilità della condotta, secondo l'art. 2 della legge Mancino, richiede, nondimeno, di considerare il «significativo contesto fattuale complessivo». Da tale inciso si possono far discendere due corollari: il primo contribuisce ad escludere la punibilità della "chiamata al presente" qualora questa si risolva in una fattispecie meramente commemorativa[22], in una "goliardata" o in una semplice "bravata"; il secondo, intrin-

[18] *Considerato in diritto*, punto 6.2.3.

[19] *Considerato in diritto*, punto 9.1.

[20] A. Tesauro, *Le radici profonde non gelano: le manifestazioni fasciste al vaglio delle sezioni unite. Tra storia e diritto*, in *Sistema penale*, 12 gennaio 2024.

[21] Così, F. Spaccasassi, *Le manifestazioni usuali del fascismo tra leggi "Scelba" e "Mancino"*, in *Questione Giustizia*, 15 marzo 2022, 2, secondo cui, questa area grigia sarebbe costituita dalla gestualità del menzionato saluto romano in «chiese, sagrati, cimiteri, stadi, consigli comunali, scuole), intonazione della chiamata del "presente" (in particolare in cerimonie di commemorazione di defunti), sfoggio di bandiere (con fasci littori e aquile, svastiche)».
In questo senso, la pronuncia appare militare in favore del mantenimento della precedente prassi applicativa che, sovente, aveva ricondotto la gestualità del "saluto romano" in assenza degli elementi modali e temporali, utili alla rievocazione del disciolto PNF, alla stregua di una fattispecie punibile ai sensi dell'art. 2 della legge Mancino. Ne sono degli esempi, la sentenza della Cass. pen., sez. I, sent. 25184/2009, attraverso cui i giudici hanno ritenuto tale gesto compiuto in una partita di calcio una «manifestazione (…) fortemente intollerante e discriminante»; e, analogamente, la sentenza di Cass. pen., sez. I, sent. 21409/2019. Cfr., per un maggiore approfondimento, A. Nocera, *Manifestazioni fasciste e apologia del fascismo tra attualità e nuove prospettive incriminatrici*, in *Dir. Pen. Con., 9 maggio* 2018; F. Paruzzo, *La XII Disposizione transitoria e finale: tra garanzia "antirazzista" della legge Mancino e specificità della matrice antifascista*, in *Associazione Italiana Costituzionalisti*, 3, 2024, 118-119; C. Caruso, *Dignità degli altri e spazi di libertà degli intolleranti*, in *Quaderni costituzionali*, 4, 2013, 804-809.
Viceversa, appare, altresì corroborata, la prassi applicativa relativa all' 5 della legge Scelba che ha riguardato casi di movimenti e liste elettorali esclusi dalle tornate elettorali, di fronte al concreto pericolo di riorganizzazione del disciolto PNF, e al differente bene giuridico che la XII dis. trans. e finale Cost., e la l. 645/1952 vogliono tutelare, di cui si dirà più diffusamente *infra* § 4.

[22] Anche quando la Cassazione, al *Considerato in diritto*, punto 8.1., afferma che vada escluso che «la caratteristica commemorativa della riunione possa rappresentare fattore di neutralizzazione degli elementi e, quindi, di "automatica" insussistenza del reato», tale assunto non sembrerebbe da intendersi nel senso che ogni manifestazione anche se dal carattere meramente commemorativo possa essere punita. Meglio, l'insussistenza del reato debba sempre ricercarsi di fronte a un'eventuale inoffensività della condotta rispetto al bene giuridico posto a tutela dalla due normative.

secamente connesso, sottolinea altresì la necessità che il giudice riconduca la rilevanza applicativa delle due disposizioni a una concreta offensività della condotta rispetto ai distinti beni giuridici che le normative intendono tutelare.
Ciò è dimostrato dal fatto che, dopo aver ricondotto le due fattispecie di reato alle rispettive categorie di pericolo concreto e presunto[23], le Sezioni Unite hanno ritenuto tale distinzione «evanescente» rispetto all'accertamento preliminare dell'offensività della condotta in relazione al bene giuridico tutelato[24]. L'approccio in questione rappresenta un elemento innovativo della pronuncia, riducendo – almeno per quanto riguarda la questione oggetto del presente giudizio – la distinzione tra reati di pericolo in concreto e reati di pericolo in astratto a una mera astrazione teorica[25]. In tal senso, i giudici sono chiamati esclusivamente a verificare il rischio effettivo di tali condotte, considerando gli specifici beni giuridici che le due normative intendono tutelare.

4. Alcune note a margine: la rinnovata predilezione per un modello di democrazia fondato sul "*marketplace of ideas*"

Queste riflessioni rappresentano l'aspetto costituzionalmente più rilevante della pronuncia in commento, dal momento che si rivelano funzionali a un inquadramento del modello italiano all'interno dei principali paradigmi di tutela della democrazia, illuminando ulteriormente il rapporto che la Carta fondamentale ha inteso instaurare con i suoi potenziali antagonisti.
A tal proposito, se si analizza il combinato disposto degli artt. 17, 18, 21 e 49 della Costituzione, si può affermare pacificamente che le libertà tutelate da tali norme si contraddistinguono per l'assenza di eccezioni al loro esercizio, fondate esclusivamente su elementi ideali. In altre parole, affinché si possa procedere ad una legittima compressione di tali diritti non è sufficiente un contenuto potenzialmente in contrasto con il dettato costituzionale, ma occorre altresì la comprovata lesione di un altro diritto in concreto[26].

[23] Cfr. *Considerato in diritto*, punto 6.2.2. per il delitto di cui all'art. 5 legge cit. come reato di pericolo in concreto; *Considerato in diritto*, punto 6.2.4., per il delitto di cui all'art. 2 legge cit., come reato di pericolo in astratto.

[24] La pronuncia cita altresì la Corte cost., sent. 225/2008, per evidenziare che «resta affidato al giudice, nell'esercizio del proprio potere ermeneutico «il compito di uniformare la figura criminosa al principio di offensività nella concretezza applicativa», v. *Considerato in diritto*, punto 6.2.4.

[25] La distinzione è nota per inquadrarvi nei reati di pericolo in concreto quelli in cui il giudice deve accertare se nel singolo caso concreto il bene giuridico ha corso un effettivo pericolo: accertamento che è doveroso, sia quando il pericolo è elemento espresso del fatto di reato, sia quando è elemento implicito da ricostruire in via interpretativa; mentre, nei reati di pericolo in astratto, quei reati nei quali il legislatore, sulla base di leggi di esperienza, ha presunto che una classe di comportamenti è, nella generalità dei casi, fonte di pericolo per uno o più svariati beni giuridici. Per un maggiore approfondimento da un punto di vista definitorio, si segnala, per tutti, G. Marinucci - E. Dolcini, *Manuale di Diritto Penale. Parte Generale*, VI Ed., Milano, 2017, 240-244.

[26] A questo proposito, volendo adottare un'interpretazione eminentemente letterale delle norme costituzionali, si evince che, in riferimento all'art. 21 Cost, esso si caratterizza per una portata espansiva e per l'assenza di restrizioni alla manifestazione del pensiero fondate unicamente sui contenuti espressi.

Lo stesso principio sembra emergere da un'interpretazione di massima della XII disposizione transitoria e finale, la quale, nel prescrivere il divieto di ricostituzione del disciolto partito fascista, non può essere considerata un limite alla libertà di manifestazione del pensiero in quanto tale[27].

Da questo punto di vista, la sentenza delle Sezioni Unite si inserisce pienamente in questa tradizione giuridica, declinando un assetto democratico aperto e plurale, che avvicina il paradigma italiano ad una impostazione marcatamente liberale, che storicamente non ha ritenuto delle minacce sufficienti gli episodi di propaganda fascista, benché ripetuti nel tempo, ove connotati da un carattere puramente ed eminentemente commemorativo[28]. In effetti, l'argomentazione esposta dai giudici non consente di

L'unico limite, espresso all'ultimo comma dell'articolo, è quello del buon costume; esso è tradizionalmente collegato alla nozione penalistica di "osceno" (e in particolare all'art. 529 c. che definisce «osceni gli atti e gli oggetti che, secondo il comune sentimento, offendono il pudore»). Tuttavia, come assai noto, è un concetto mutevole nel tempo, con la conseguenza che una manifestazione del pensiero può essere ritenuta immorale solo qualora si ponga in antitesi con «la pluralità delle concezioni etiche che convivono nella società contemporanea» (Corte cost., sent. 293/2000), v. sul punto G. E. Vigevani, *Articolo 21*, in F. Clementi - L. Cuocolo - F. Rosa - G. E. Vigevani (a cura di), *La Costituzione italiana. Commento articolo per articolo*, Bologna, 2018, 150-151.

Analogamente, l'art. 17 della Costituzione si astiene dal definire i fini della riunione, consentendo pertanto l'esercizio della libertà in questione per i più diversi obiettivi; a questo stretto proposito, non possono costituire limiti i fini di natura concettuale attinenti non al fatto della riunione, bensì all'attività finale ivi svolta (v., più approfonditamente, F. Rosa, *Articolo 17*, in *La Costituzione italiana. Commento articolo per articolo*, cit., 122; A. Pace, *Articolo 17*, in G. Branca (a cura di), *Commentario della Costituzione. Rapporti civili. Art. 13-20*, Bologna, 1977, 157.

Lo stesso principio si applica alla libertà di associazione sancita dall'art. 18 della nostra Costituzione, la quale può essere esercitata dal singolo per motivi che egli ritenga meritevoli di considerazione, inclusi quelli di opposizione ai poteri pubblici e privati. Le uniche eccezioni alla libertà sono costituite dalle associazioni segrete e quelle aventi un'organizzazione a carattere militare, entrambe espressamente proibite dalla lettera costituzionale. Per il resto, tale libertà è da considerare della medesima rilevanza costituzionale dell'art. 21 Cost., in virtù del principio espresso dalla Corte costituzionale, secondo cui, non sarebbe possibile vietare alle associazioni quello che non viene vietato ai singoli. Per inciso, «se non è illecito che il singolo svolga opera di propaganda antinazionale, non può costituire illecito neppure l'attività associativa volta a compiere ciò che è consentito all'individuo» (Corte cost., sent. 243/2001), v. sul punto F. Clementi, *Articolo 18*, in *La Costituzione italiana*, cit., 131.

E, in ultimo luogo, anche l'art. 49 Cost., nel proclamare la libertà di associazione partitica, rifiuta di imporre qualsiasi forma di controllo ideologico e di escludere i partiti politici dalle competizioni elettorali in base al programma espresso: non vi è, da questo punto di vista, alcun possibile rimando al congegno costituzionale tipico di una democrazia militante, la quale, sulla falsariga della *wehrhafte e verteidigte Demokratie tedesca*, ammette l'esclusione di quei partiti in grado di negare l'esistenza dello stato di diritto e di sovvertire le sue regole (v., sulla teoria costituzionale il suo massimo esponente K. Loewenstein, *Militant Democracy and Fundamental Rights I*, in *American Political Science Review*, 1937a).

[27] G. E. Vigevani, *Origine e attualità del dibattito sulla XII disposizione finale della Costituzione: i limiti della tutela della democrazia*, in questa *Rivista*, 1, 2019, 29.

[28] Cass., pen., sez. I, sent. 37577/2014, secondo cui i gesti usuali del disciolto partito fascista non rappresentassero una lesione dell'interesse tutelato con l'art. 5 della l. 645/1962 per la «natura puramente commemorativa della manifestazione e del corteo, organizzati in onore di tre defunti, vittime di una violenta lotta politica che ha attraversato diverse fasi storiche»; Cass. pen., sez. I, sent. 11038/2016, la quale, conformemente alla giurisprudenza costituzionale, interpreta la fattispecie sanzionata nell'art. 5 legge cit. da qualificarsi in termini di reato di pericolo in concreto, che non sanziona le manifestazioni di pensiero, ma soltanto «ove le stesse possano determinare il pericolo di ricostituzione di organizzazioni fasciste, da verificarsi in concreto con riguardo al momento ed all'ambiente in cui sono compiute, attentando concretamente alla tenuta dell'ordine democratico e dei valori ad esso sottesi», e, analogamente, App. Milano, 21 settembre 2016; Cass. pen., sez. I, sent. 8108/2018, secondo cui il saluto romano, se fatto con intento commemorativo e non violento, non è penalmente rilevante, in quanto

approdare a un'interpretazione che *ipso facto* conduca a considerare incostituzionali i movimenti neofascisti, né giustifica l'incriminazione delle loro dichiarazioni, anche quando queste siano in evidente contrasto con i principi fondamentali della democrazia repubblicana, in questo conformandosi all'orientamento già tracciato dalla Corte costituzionale, con le sentenze n. 1/1957 e n. 74/1958[29].

Attraverso queste due pronunce, il Giudice delle leggi ha infatti affermato che l'art. 5 della legge 20 giugno 1952, n. 645 non punisce «una qualunque manifestazione del pensiero [...] bensì quelle manifestazioni usuali del disciolto partito, che [...] possono determinare il pericolo di riorganizzazione che la norma ha voluto evitare»[30]. Ne discende, che eventuali restrizioni alla libertà di espressione continuano a richiedere l'accertamento di un pericolo chiaro ed imminente per un interesse pubblico essenziale ("*clear and present danger*")[31], sottolineando la necessità di imprimere eventuali restrizioni alla libertà di espressione solo ove vi siano specifiche circostanze capaci di rappresentare un reale pericolo per la pacifica convivenza sociale generando un incitamento all'odio e alla violenza[32].

---

la legge non punisce «tutte le manifestazioni usuali del disciolto partito fascista, ma solo quelle che possono determinare il pericolo di ricostituzione di organizzazioni fasciste», e di conseguenza, solo «i gesti idonei a provocare adesioni e consensi».

Per un commento più approfondito di queste sentenze e della giurisprudenza richiamata, v. *Sulla rilevanza penale del "saluto romano": non è reato se fatto con intento commemorativo*, in *Giurisprudenza penale*, 21 febbraio 2018; A. Nocera, *Manifestazioni fasciste e apologia del fascismo tra attualità e nuove prospettive incriminatrici*, cit.; A. Galluccio, *Il saluto fascista è reato? L'attuale panorama normativo e giurisprudenziale ricostruito dal Tribunale di Milano, in una sentenza di condanna*, in *Dir. Pen. Con.*, 29 aprile 2019; C. Brusco, *Contrasti giurisprudenziali sull'interpretazione e applicazione delle leggi di contrasto al neofascismo*, in *Questione Giustizia*, 14 maggio 2019.

[29] Parimenti, si potrebbe qui aggiungere, alcuna minaccia seria è stata attribuita al Movimento Sociale Italiano, quale partito neofascista che ha trovato spazio nel contesto politico repubblicano. Per una bibliografia essenziale sulla storia del Movimento Sociale Italiano (MSI) e sulla sua legittimità rispetto alla Costituzione italiana, v. G. Parlato, *Fascisti senza Mussolini: le origini del neofascismo in Italia, 1943–1948*, Bologna, 2006; P. Rosenbaum, *Il nuovo fascismo. Da Salò ad Almirante. Storia del Msi*, Milano, 1975; P. Ignazi, *L'estrema destra in Europa*, Bologna, 2000; P. Ignazi, *Il polo escluso: profilo storico del Movimento sociale italiano*, Bologna, 1998; Id., *Fascists and neo-fascists*, in E. Jones, G. Pasquino (a cura di), *The Oxford Handbook of Italian Politics*, Oxford, 2015; Archivio storico del Senato, *Movimento sociale italiano (Msi) 1946 – 1995*.

[30] Corte cost., sent. 74/1958. Tale orientamento sarebbe stato consolidato negli anni successivi, per ribadire, sotto differenti tonalità, che l'apologia punibile non è la manifestazione di pensiero pura e semplice, bensì quella che si traduce in azione e che, per dirla con le parole della Corte, «per la sua modalità integri un comportamento concretamene idoneo a provocare la commissione di delitti» (Corte cost., sent. 65/1970).

[31] La dottrina è stata introdotta dal giudice Oliver Wendell Holmes Jr. nella sua opinione di maggioranza in Schenck v. United States, 249 U.S. 47 (1919) affermando che «*The question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that the United States Congress has a right to prevent. It is a question of proximity and degree* (…)».

In dottrina è stato evidenziato che, la Corte costituzionale successivamente, emettendo una serie di pronunce, rinvenibile nelle sentt. 71/1978, 87/1966, e 65/1970, avrebbe per l'appunto recepito nella propria giurisprudenza la teoria americana del "*clear and present danger*", quale criterio per la verifica di un rischio effettivo all'ordine pubblico costituito (così, A. Cerri, *Ordine pubblico, II) Diritto costituzionale*, *Enciclopedia giuridica*, XXII, Roma, 1990, 7, 9).

[32] Tale è, dunque, l'interpretazione prevalente delle libertà di espressione ("*free speech*") da ricondurre al Primo emendamento della Costituzione americana, secondo il quale: «*Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances*». Infatti, un differente orientamento porterebbe a una «standardizzazione» delle idee, per opera dei principali gruppi

Con riferimento alle condotte apologetiche che qui interessano, esso potrebbe declinarsi nel senso che, proprio a partire dalla sentenza delle Sezioni Unite, questi pericoli sarebbero da ricondursi: per quanto concerne l'applicazione dell'art. 5 della legge Scelba, alla tutela dell'«ordine democratico-costituzionale», dinanzi a un pericolo che dunque deve leggersi nel segno di una minaccia di tipo sovversivo al modello di Stato delineato nella Costituzione, di cui la componente politico-istituzionale rappresenta la chiave di volta; mentre, con riferimento all'applicazione dell'art. 2 della legge Mancino, deve tradursi in un depauperamento dei principi costituzionali che presiedono allo sviluppo della «dignità umana», a seguito del quale è la componente civica e culturale ad apparire maggiormente coinvolta.

Se questo è vero, sembra potersi così giustificare anche quella prassi applicativa che, nel corso del tempo, ha fatto registrare un rinnovato rigore da parte della giurisprudenza di merito in materia elettorale. Un crescente numero di liste, infatti, sono state escluse dalle competizioni elettorali a causa del loro riferimento fin dal nome del partito, all'ideologia e alla simbologia fascista[33].

Sotto questo specifico punto di vista, anche questo filone giurisprudenziale non sembra essere stato smentito dalla sentenza delle Sezioni Unite: in effetti si potrebbe compendiare che, quanto più si percepisce un fondato timore che alcuni candidati possano accedere al potere e attuare, sulla base del loro programma elettorale, "azioni che modifichino la realtà"[34], tanto più il combinato normativo della XII disposizione transitoria e finale della Costituzione e della Legge Scelba tende a rivelare le sue dinamiche applicative[35].

Si evince, in definitiva, un'oscillazione applicativa della XII disposizione e della sua legge di attuazione, tra la duplice esigenza di tutelare il "puro pensiero" e quella di preservare l'ordine legale precostituito dinanzi a un concreto pericolo di rottura della

dominanti, come alternativa che non è compatibile con il Primo emendamento della Costituzione americana; come puntualmente esorcizzato in *Terminiello v. Chicago*, 337 U.S. 1 (1949), «*For the alternative would lead to standardization of ideas either by legislatures, courts, or dominant political or community groups*».

[33] È stato il caso del movimento dei "Fasci italiani del Lavoro" (TAR Lombardia, sent. 105/2018), e, qualche tempo prima, della lista intitolata "Fascismo e libertà", la cui esclusione dalla tornata elettorale amministrativa celebratesi nel maggio 2012, in Abruzzo, è stata motivata dal Consiglio di Stato, sez. V, sent. 1355/2013, sulla scorta del fatto che la XII Disp. Trans. e finale «fissando un'impossibilità giuridica assoluta e incondizionata, impedisce che un movimento politico formatosi e operante in violazione di tale divieto possa in qualsiasi forma partecipare alla vita politica e condizionarne le libere democratiche dinamiche», cfr. sul punto, G.E. Vigevani, *Origine e attualità del dibattito sulla XII disposizione finale della Costituzione: i limiti della tutela della democrazia*, cit., 39 (nota 46), e più diffusamente F. Blando, *Movimenti neofascisti e difesa della democrazia*, *Costituzionalismo.it*, 1, 2014, 12 – 14.

[34] P. Nuvolone, *Il problema dei limiti alla libertà di pensiero nella prospettiva logica dell'ordinamento*, in Aa. Vv., *Legge penale e libertà di pensiero*, Padova, 1966, 353, e sempre P. Nuvolone, *Le leggi penali e la Costituzione*, Milano, 1953, 46: «[…] il nostro ordinamento s'ispira al principio del pluralismo dei partiti e della convertibilità della maggioranza: tali principi implicano l'inammissibilità di qualsiasi forma di controllo «ideologico» e «politico» che tenda a limitare l'espressione del pensiero nella sua forma associata, ed esigono quindi il pieno riconoscimento e la più rigida garanzia della manifestazione del pensiero in materia politica».

[35] V. anche D. Piccione, cit., 1949, secondo cui, «tali dispositivi si sostanziano in tecniche di tutela già costituzionalmente previste e anche per questo non colpiscono mai la sola manifestazione del pensiero ma gravano, per solito, sull'associazionismo politico o, in casi limite, sull'elettorato politico passivo: esse dunque creano un filtro contro le concezioni che non possono e non debbono farsi valere per la formazione della politica nazionale»

"pace materiale"[36].

Questa tesi, del resto, s'inserisce nel più ampio solco della volontà del Costituente italiano di superare positivamente l'impianto fascista preesistente, inclusa la concezione del rapporto tra cittadino e autorità. L'introduzione di regole costituzionali a tutela del futuro assetto democratico-repubblicano, sotto forma di astrazione meramente teorica, avrebbe infatti replicato le medesime logiche di autoconservazione di un ordine ideale, che, ora, all'interno di una democrazia centrata sulla persona, doveva ritenersi anacronistica. Sicché, di fronte a un quadro normativo così delineato, la democrazia italiana, nel mostrare una certa tolleranza anche nei confronti dei più intolleranti, si sforza di non tradire i propri principi fondamentali e affronta i suoi avversari, per così dire, «con una mano legata dietro la schiena»[37].

Tale impostazione si colloca all'interno di un contesto democratico aperto e pluralistico e si allinea coerentemente con i tratti distintivi dell'utilitarismo liberale di matrice nordamericana promuovendo il concetto di "*marketplace of ideas*"[38] e considerando ogni dottrina emergente dalla società civile come una verità parziale, suscettibile di miglioramento e potenzialmente superabile attraverso un confronto dialettico[39].

In questa chiave, la sentenza della Cassazione non propende per un modello di democrazia protetta, tipico di sistemi come quello tedesco, e dunque si rifiuta di impiegare strumenti repressivi e protettivi nei confronti delle formazioni eversive o di stabilire dei limiti espliciti alla libertà di associazione e di manifestazione del pensiero, dichiarando incostituzionali quei partiti che «per le loro finalità o per il comportamento dei

---

[36] A. Pace, *Problematica delle libertà costituzionali. Parte generale*, III ed., Padova, 2003, 45 ss.

[37] L'espressione è stata utilizzata da S. Curreri, cit.; *Sullo scioglimento di Forza Nuova e, più in generale, delle forze politiche che agiscono con metodo non democratico*, in *LaCostituzione.info*, 19 ottobre 2021; *Perché Forza Nuova va sciolta, ma non per decreto*, in *Il Riformista*, 15 ottobre 2021.

[38] È da questo punto di vista la più ampia circolazione delle più diverse idee, compatibilmente con i principi costituzionali, ad essere stata finanche concepita da C. Esposito la via migliore per l'affermazione dello Stato democratico: «Si vuole solo affermare che non la democraticità dello Stato ha per conseguenza il riconoscimento di quella libertà, sicché possa determinarne la funzione ed i limiti, ma che le ragioni ideali del riconoscimento di quella libertà (e cioè del valore della persona umana) porta tra le tante conseguenze anche alla affermazione dello Stato democratico» (C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958, 12).

Per un approfondimento sul concetto di "*martketplace of ideas*" vedi, J. S. Mill, *On Liberty*, London, 1859; J. Milton, *Areopagitica*, London, 1644. Per i principali casi di giurisprudenza della Corte Suprema USA che hanno consolidato il principio nelle corti statunitensi comuni: *Abrams v. United States*, 250 U.S. 616 (1919); *Terminiello v. Chicago*, 337 U.S. 1 (1949), *Brandenburg v. Ohio*, 395 U.S. 444 (1969). Per una bibliografia più recente sul concetto, v. R. Coase, *Markets for goods and Market for ideas*, *American Economic Review*, 1974; A.I. Goldman, J.C. Cox, *Speech, truth, and the free market for ideas*, Cambridge, 1996; M. N. Browne, J. Rex, D.L. Herrera, *Potential Tension Between a "Free Marketplace of Ideas" and the Fundamental Purpose of Free Speech*, in *Akron Journal of Constitutional Law and Policy*, 2012.

[39] «(…) *though the silenced opinion be an error, it may, and very commonly does, contain a portion of truth; and since the general or prevailing opinion on any subject is rarely or never the whole truth, it is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied*», v. J. S. Mill, *On liberty*, London, 1859, 72, trad. it., *Sulla libertà*, Milano, 2000.

Un certo valore allo scontro dialettico è assegnato dalla Corte suprema americana in *Terminiello v. Chicago*, 337 U.S. 1 (1949), la cui sentenza consegna all'interprete una particolare concezione della libertà di espressione allineata con la teoria del "*marketplace of ideas*", nonché con l'affine considerazione di ogni verità come una verità "parziale": «*Accordingly, a function of free speech under our system of government is to invite dispute. Speech is often provocative and challenging*».

loro aderenti si prefiggono di attentare all'ordinamento costituzionale democratico»[40].

5. … integrato da uno spirito antifascista

Dall'analisi condotta emerge, dunque, come la pronuncia delle Sezioni Unite riaffermi significativamente l'identità antifascista della democrazia costituzionale repubblicana, mantenendone intatto lo spirito pluralista[41]. Se, da una parte, infatti, il pensiero fascista in sé non può essere considerato reato; dall'altra, l'irrilevanza del criterio distintivo tra reati di pericolo concreto e reati di pericolo astratto, sembra qualificare come illecito penale qualunque comportamento che, esternato nella realtà dei fatti, richiami contenutisticamente il Fascismo[42].

Ciò è confermato dalla possibile estensione applicativa della condotta della "chiamata del presente" al reato di cui all'art. 2 della Legge Mancino. Questo, tuttavia, alla condizione che l'applicazione della disposizione incriminatrice non si contrassegni per un limite alla mera manifestazione del pensiero, ma si giustifichi piuttosto con la necessità di tutelare quel bene giuridico «*composito*», custodito dalla legge n. 205 del 1993, e finalizzato a proteggere i diritti fondamentali garantiti dalla Costituzione[43].

In questa ipotesi, la rilevanza penale delle condotte apologetiche persegue un interesse costituzionalmente rilevante, atto a preservare quei «principi fondamentali di ordine legale alla base della convivenza civile»[44]. Da questo punto di vista, al fine di prevenire un turbamento della pacifica convivenza civile[45], si potrebbe persino sostenere che la

[40] Si prendano in considerazione a titolo efficacemente esemplificativo gli artt. 18 e 21 della Legge fondamentale tedesca: il primo, nel tutelare la libertà di espressione afferma che «chi abusa della libertà di espressione delle proprie opinioni (…) per combattere i principi del libero ordinamento democratico, perde questi diritti fondamentali» (art. 18, c. 2, l.f. RFG); mentre, la seconda disposizione sulla libertà di associazione partitica dispone che «I partiti, che per le loro finalità o per il comportamento dei loro aderenti mirino ad attentare al libero e democratico ordinamento costituzionale o a sovvertirlo o a mettere in pericolo l'esistenza della Repubblica Federale di Germania sono incostituzionali».

[41] Efficacemente esemplificativo dell'opposta esigenza è il seguente concetto espresso dagli ermellini che richiama sia l'aspetto identitario, sia l'ultronea necessità dell'accertamento dei presupposti modali che in seguito indicherà come necessari per l'applicazione del citato quadro normativo: «Dunque, proprio per quanto fin qui detto, non la tutela del mero "ordine pubblico materiale" deve ritenersi venire nella specie in gioco […] ma, in una visione di ben più ampio respiro, la stessa tavola dei valori costituzionali e democratici fondativi della Repubblica, efficacemente riassumibili nel bene dell'ordine pubblico democratico o costituzionale, posto in pericolo, a fronte dell'elemento modale – spaziale indicato dalla norma, da possibili consensi o reazioni a tali manifestazioni atti a turbare, anche ma non solo, la civile convivenza» (*Considerato in diritto*, punto 6.2.2.

[42] Se ne riporti a questo riguardo il relativo passaggio della pronuncia dei giudici: «Da tali considerazioni discende che, quanto meno ai fini della presente decisione, la distinzione tra un "pericolo concreto" ed un "pericolo astratto o presunto" finisca, a ben vedere, per divenire, nei fatti, evanescente una volta che si prenda contestualmente atto di come, per quanto appena detto, anche le previsioni contrassegnate da un pericolo presunto debbano coniugarsi con il principio di offensività» (*Considerato in diritto*, punto 6.2.4.).

[43] *Considerato in diritto*, punto 6.2.3.

[44] Corte., cost., sent. 19/1962

[45] A. Nocera, *Manifestazioni fasciste e apologia del fascismo tra attualità e nuove prospettive incriminatrici*, cit., 7. Da questo punto di vista, la sentenza degli ermellini appare richiamare lo stesso tenore espressivo della Cass. pen., sez. I, sent. 3791/1993, secondo la *quale «La diffusione di tali ideologie produce la lesione della*

pronuncia delle Sezioni Unite persegua una volontà pedagogica supplementare, poiché richiama alla memoria collettiva la convinzione antifascista, come patto fondativo su cui si regge lo Stato democratico repubblicano.

Né è un esempio la scelta di spiegarsi anche in termini simbolici, in particolar modo descrivendo il "saluto romano" come una condotta che è chiara a tutti essere una «naturale» identificazione della «liturgia delle adunanze fasciste»[46] e, nell'ambito applicativo di cui all'art. 2 della legge Mancino, un possibile «strumento simbolico di espressione di idee di intolleranza o di discriminazione» [47].

Attesa, dunque, la portata storica di tale gesto, non appare possibile «indulgere nella libertà dell'errore»[48], al cospetto di un atteggiamento che diviene perseguibile penalmente per la sua sostanziale idoneità a rappresentare un pericolo per «la disgregazione dei valori di solidarietà, dignità ed eguaglianza di tutti consociati»[49]. Questo approccio autorizza l'interprete a considerare – senza esitazioni – il tratto identitario «antifascista» come un «principio fondamentale»: ed è questo, se vogliamo, il principale meccanismo di difesa culturale della democrazia che la sentenza di Cassazione tenta di introdurre nel nostro ordinamento.

dignità dell'uomo e delle condizioni di pacifica convivenza democratica, fondate sulla reciproca tolleranza fra popolazioni di differente cultura ed etnia».

[46] *Considerato in diritto*, punto 8.

[47] *Considerato in diritto*, punto 9.1.

[48] C. Mortati, *Costituzionalità nel disegno di legge per la repressione dell'attività fascista*, in Id., *Problemi di diritto pubblico nell'attuale esperienza costituzionale repubblicana*, Milano, 1972, 15-16.

[49] *Considerato in diritto*, punto 6.2.3.

Cronache

La regolamentazione del *deepfake* in Europa, Stati Uniti e Cina*

Alberto Orlando

Abstract

Il vertiginoso aumento dei *deepfake* negli ultimi anni ha portato i regolatori pubblici a interrogarsi sulla necessità di regolamentare il fenomeno, incontrando problematiche simili a quelle che riguardano in generale la regolamentazione dell'IA.
Unione Europea, Stati Uniti e Cina, ossia le potenze che si contendono la posizione dominante in materia di IA, di recente sembrano aver intrapreso la strada della regolamentazione del fenomeno, scegliendo però approcci profondamente diversi tra loro, che riflettono le peculiarità degli ordinamenti e la visione politica e strategica in materia di sviluppo delle nuove tecnologie.
Il presente contributo offre una panoramica delle principali soluzioni normative adottate e mette a confronto i tre approcci, sottolineandone analogie e differenze.

The vertiginous rise of *deepfakes* in recent years has led public regulators to question the need to regulate the phenomenon, encountering similar problems to those affecting AI regulation in general.
The European Union, the United States and China, i.e. the powers that contend for the dominant position in the field of AI, recently seem to have embarked on the path of regulating the phenomenon, choosing, however, profoundly different approaches, reflecting the peculiarities of the legal systems and the political and strategic vision regarding the development of new technologies.
This paper provides an overview of the main regulatory solutions adopted and compares the three approaches, highlighting their similarities and differences.

Sommario

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Keywords

deepfake – intelligenza artificiale – sintesi profonda – regolamentazione dell'IA – disinformazione

1. Introduzione: profili regolatori delle tecnologie di *deepfake*

In una società in cui la circolazione di notizie e contenuti falsi, soprattutto attraverso il *web*, non è certo un fatto nuovo[1], i *deepfake* si ritagliano uno specifico spazio che li distingue da altri fenomeni.

Essi possono essere definiti come output derivanti dall'utilizzo di tecniche di intelligenza artificiale (di seguito, IA) atto a generare audio e video sintetici ma estremamente realistici, soprattutto per quanto riguarda le somiglianze facciali e vocali umane. Poiché la tecnologia è sempre più disponibile per la sperimentazione da parte di chiunque possieda un minimo di competenze, i *deepfake* sono oggi utilizzati in ambiti di applicazione via via crescenti[2], dall'intrattenimento, alla politica, alla pornografia, ecc. Sebbene i contenuti di alta qualità richiedano un'ampia quantità di dati oltre a eccellenti competenze da parte del creatore del contenuto, in realtà anche i *deepfake* di bassa qualità possono rivelarsi ugualmente dannosi[3].

Ben si comprende che, sebbene possano rintracciarsi alcuni utilizzi "positivi"[4], gli impatti negativi sembrano di gran lunga superiori e molto evidenti: ad esempio, danni emotivi, furto di identità, danni alla reputazione, manipolazione politica. Tutti effetti

[1] Si pensi alla questione delle *fake news*, sulla cui regolamentazione il dibattito resta costantemente aperto. Cfr., *ex multis*, M. Bassini – G.E. Vigevani, *Primi appunti su* fake news *e dintorni*, in questa *Rivista*, 1, 2017, 11 ss. e gli altri contributi contenuti nella stessa sezione monografica della rivista.

[2] I contenuti *deepfake* sono generalmente creati utilizzando le reti generative avversarie (GAN), una tecnologia creata da Ian Goodfellow nel 2014. L'ascesa della tecnologia *deepfake* e dell'utilizzo da parte dei consumatori è iniziata nel 2017 sul sito web *Reddit*, quando un utente chiamato appunto "*deepfake*" ha pubblicato materiale pornografico manipolato che scambiava i volti di celebrità e personaggi pubblici con quello di altre persone. Questi post ottenevano grande popolarità, tanto che una pagina *Reddit* specializzata, nota come "*Subreddit*", veniva dedicata esclusivamente ai video *deepfake* e raggiungeva rapidamente centinaia di migliaia di membri della comunità. Nel 2018, un *deepfake* che ritraeva l'ex presidente degli USA Barack Obama intento a utilizzare un linguaggio oltraggioso nei confronti di altri politici fece il giro del mondo ingenerando non poche confusioni sulla sua veridicità e portando alla ribalta le potenzialità e i rischi dei cc.dd. *deepfake* "politici". Dal dicembre 2018 il numero dei deepfake online nei successivi due anni è all'incirca raddoppiato ogni sei mesi, confermando una crescita esponenziale del fenomeno. Cfr. M.B. Kugler – C. Pace, *Deepfake Privacy: Attitudes and Regulation*, in *Northwestern University Law Review*, 3, 2021, spec. 620-621; B. Chesney – D. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in *California Law Review*, 6, 2019, spec. 1758 ss.; L. Whittaker – R. Mulcahy – K. Letheren – J. Kietzmann – R. Russell-Bennett, *Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda*, in *Technovation*, 125, 2023.

[3] Cfr. Y. Geng, *Comparing "Deepfake" Regulatory Regimes in The United States, the European Union, and China*, in *Georgetown Law Technology Review*, 7, 2023, 158.

[4] Cfr. S. Chandler, *Why Deepfakes are a Net Positive for Humanity*, in *forbes.com*, 9 marzo 2020. Tra gli esempi si possono citare il montaggio di video senza riprese, l'esperienza di cose che non esistono più, l'aumento dell'accessibilità per le persone con disabilità, la possibilità di migliorare le pratiche mediche, ecc.

che con buona probabilità arrivano a coincidere con la commissione di condotte illecite, spesso rilevanti anche sul piano penale[5]. In ogni caso, i *deepfake* contribuiscono a quel processo di erosione della fiducia della società nei confronti della tecnologia, del mondo dell'informazione e delle istituzioni politiche. Questa sperequazione tra rischi e benefici connota i *deepfake* rispetto al *mare magnum* dell'IA, laddove la discussione sul rapporto tra vantaggi e svantaggi appare molto più incerta[6].

In questo quadro, il regolatore pubblico deve quantomeno interrogarsi sull'opportunità di regolare il fenomeno del *deepfake*. Tale esigenza si inserisce perfettamente – al netto delle peculiarità appena evidenziate – nella riflessione sulla regolamentazione delle tecnologie di intelligenza artificiale, di cui evidentemente i *deepfake* fanno parte nel momento in cui la loro realizzazione deriva dall'utilizzo di sistemi di IA. Pensare a nuove forme di regolamentazione per i *deepfake* sembra indispensabile anche perché la tutela accordata alla privacy non appare sufficiente: in effetti, pur nelle esperienze assai differenti prese in considerazione in questo contributo[7], sul piano della privacy le condotte illegali si sostanziano in linea generale nella diffusione di dati protetti dalla legge ma comunque veri. Nel caso dei *deepfake*, invece, si assiste ad una parte del contenuto che è vera ma non privata, come ad es. il volto o la voce della persona rappresentata, unita inscindibilmente ad un'altra parte che potrebbe essere privata ma sicuramente non è vera[8].

Al di fuori del contesto della normativa sulla privacy, resta comunque la strada di impedire o limitare i *deepfake* in quanto dati e notizie falsi, in grado di mettere in pericolo interessi pubblici o diritti dei singoli. In questo senso, si pone un problema di necessario bilanciamento con il diritto alla libertà di espressione e tutti i suoi corollari, compreso il diritto di critica e di satira. Dato che la mera falsità non può operare come condizione sufficiente, si potrebbe scegliere di trattare i *deepfake* come contenuti diffamatori, cioè destinati a essere percepiti come veri e dannosi per la reputazione della persona ritratta: in questo caso, l'etichettatura del *deepfake* verrebbe in soccorso, poiché, svelandone la natura, aiuterebbe a negare almeno l'elemento della apparente veri-

---

[5] Esistono anche potenziali aree grigie: ad esempio, l'uso di *deepfake* di celebrità in pubblicità e video di formazione, con o senza il permesso delle celebrità, ha attirato l'interesse di aziende del settore dei media e del marketing per la possibilità di aumentare i profitti a costi inferiori. Cfr. R. Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, in *Georgetown Law Technology Review*, 3, 2019, 368-383.

[6] Cfr., *ex multis*, S. Russell – P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, 2020; W. Barfield – U. Pagallo, *Law and Artificial Intelligence*, Cheltenham, 2020; U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; A. D'Aloia (a cura di), *Intelligenza artificiale e diritto*, Milano, 2021; C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, Speciale, 2019, 101-130; E. Stradella, *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in questa *Rivista*, 1, 2019, 73 ss.

[7] Cfr. A. Di Martino, *Profili costituzionali della privacy in Europa e negli Stati Uniti*, Napoli, 2017; U. Pagallo, *La tutela della privacy negli Stati Uniti d'America e in Europa: modelli giuridici a confronto*, Milano, 2008; D. Clementi, *La legge cinese sulla protezione delle informazioni personali. Un GDPR con caratteristiche cinesi?*, in *Rivista di diritti comparati*, 1, 2022, 189-216; E. Bertolini, *L'"apertura sorveglia": la via cinese alla governance e alla censura di Internet*, in *Diritto pubblico comparato ed europeo*, 3, 2008, 1063-1097.

[8] Cfr. Kugler, *Deepfake Privacy*, cit., 613-614.

dicità. Tuttavia, una tale soluzione non sembra esaurire del tutto le problematiche, dato che il danno alla reputazione o alla dignità della persona potrebbe prodursi ugualmente – per quanto, magari, in misura diversa – nel caso di circolazione di contenuti falsi e dichiarati al pubblico come tali: in questo senso, l'etichettatura non sarebbe risolutiva. Senza contare che, spostandosi dalla ragionevolezza della misura regolatoria alla sua effettività, sembra lecito dubitare della reale efficacia delle etichettature, considerato che queste potrebbero essere comunque eliminate in un momento successivo all'apposizione: ipotesi che, nel mondo di Internet in cui gli utenti sono protetti dall'anonimato e ai *provider* non può essere assegnato con leggerezza il ruolo di censori-gestori dei contenuti, rischia di rendere evanescenti le responsabilità.

In effetti, come accade comunemente in materia di utilizzo di IA[9], proprio la ripartizione delle responsabilità resta un punto controverso. In primo luogo, si potrebbe insistere sulla responsabilità dei creatori di *deepfake*, dato che ad essi può ricondursi la paternità dell'opera e, salvo casi eccezionali, gli effetti conseguenti: tuttavia, tale soluzione rischia di rivelarsi vana a fronte dell'anonimato online e della circolazione virale dei contenuti. In secondo luogo, occorrerebbe riflettere sulle responsabilità di utenti differenti dai creatori, comunque in grado di contribuire alla diffusione, alla trasmissione, al download di contenuti non consentiti. In terzo luogo, il centro del dibattito è probabilmente occupato dalle responsabilità (e dai poteri) che potrebbero essere assegnate ai gestori di servizi online, dovendosi distinguere diverse categorie di operatori, ossia, ad esempio, i fornitori di servizi di sintesi profonda che consentono la creazione di contenuti *deepfake* e le piattaforme online (compresi *social network*) che ne consentono la diffusione. Attribuire responsabilità ai *provider* appare soluzione vantaggiosa perché le competenze tecniche di questi soggetti garantirebbero in maniera efficace un controllo sui contenuti anche in via preventiva. Tuttavia, evidentemente ne scaturirebbero alcuni svantaggi: da un lato, si finirebbe per delegare il delicatissimo controllo sui limiti della libertà di espressione a soggetti privati, spesso operanti in un contesto economico transnazionale; dall'altro, agli stessi soggetti sarebbero richiesti standard e strutture particolarmente gravose da sostenere, a fronte di una esposizione a responsabilità anche rilevanti, in un quadro che potrebbe scoraggiare gli investimenti nel settore (almeno nei contesti geografici caratterizzati da normative più "severe")[10].

Cosa fare, quindi? Verso quale direzione dovrebbero orientarsi i regolatori pubblici? O meglio: quale logica dovrebbero seguire nella ricerca delle soluzioni? Anche qui rischia-

[9] Oltre a rimandare alla bibliografia citata *supra*, nota 6, cfr. anche M. Bassini – L. Liguori – O. Pollicino, *Sistemi di Intelligenza Artificiale, responsabilità e* accountability. *Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 333-371; E. Palmerini – E. Stradella (a cura di), *Law and Technology. The Challenge of Regulating Technological Development*, Pisa, 2013; U. Pagallo, *Intelligenza Artificiale e diritto. Linee guida per un oculato intervento normativo*, in *Sistemi intelligenti*, 3, 2017, 615-636; G. Comandè, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi giuridica dell'economia*, 1, 2019, 173 ss.; S. Lohsse – R. Schulze – D. Staudenmayer (a cura di), *Liability for Artificial Intelligence and the Internet of Things*, Baden-Baden, 2019; R. Brownsword – K. Yeung (a cura di), *Regulating Technologies. Legal Futures, Regulatory Frames and Technological Fixes*, Oxford, 2008.

[10] Cfr. J. Riordan, *The Liability of Internet Intermediaries*, Oxford, 2016; G.E. Vigevani, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritto costituzionale*, 1, 2023, 41-54; L. Albertini, *Sulla responsabilità civile degli internet service provider per i materiali caricati dagli utenti (con qualche considerazione generale sul loro ruolo di gatekeepers della comunicazione)*, in *Medialaws, Law and Media Working Papers Series*, 4, 2020.

mo di trovarci di fronte ad un dilemma. Da un lato, infatti, il fenomeno appare connotato, come accade per tutto ciò che gravita nel *web*, dal carattere della transnazionalità[11], che rende preferibile approntare soluzioni in grado di valere anche al di là dei confini nazionali: questa strada può essere seguita sia per mezzo di un raccordo preventivo tra i legislatori nazionali, sia attraverso lo studio delle normative già vigenti – o in via di discussione – in altri Paesi. Tale elemento appare ancora più centrale se si considera che attorno a queste opzioni legislative si gioca una parte consistente della partita per il primato economico (e culturale) in materia di sviluppo delle nuove tecnologie e segnatamente dell'IA. D'altro canto, il ricorso a regolamentazioni più o meno restrittive deve fare i conti con l'accettazione sociale delle norme introdotte. Da questo punto di vista, l'analisi del contesto valoriale e sociale assume importanza fondamentale, considerato che sovrastimare o sottostimare la criminalizzazione di determinate condotte può sortire l'effetto di degradare la legittimità della legge agli occhi del pubblico e ridurre il rispetto delle regole stabilite[12].

Alla luce di quanto appena detto, il presente contributo, lungi dal prospettare soluzioni definitive rispetto ai problemi proposti, punta proprio a descrivere e analizzare il modo in cui Unione europea, Stati Uniti e Cina – ossia i maggiori *competitor* a livello mondiale in materia di IA[13] – stanno affrontando il fenomeno. Si anticipa sin da ora che in tutti e tre gli ordinamenti negli ultimissimi anni sono emersi, accanto ad orientamenti e studi sui *deepfake*, anche i primi approcci normativi dedicati più o meno specificamente alla materia, i quali rivelano visioni molto diverse, sulla falsariga di quanto sta avvenendo con riguardo alla regolamentazione generale dell'IA[14].

2. Unione europea: tra *tackling* e *risk-based approach*

La legislatura dell'UE conclusa nel 2024 ha affrontato con decisione le sfide regolatorie poste dalle nuovissime tecnologie, scegliendo un approccio unico a livello globale e molto diverso da quello statunitense e cinese: pur non sopprimendo lo spazio per l'autoregolamentazione privata, le Istituzioni europee hanno convintamente intrapreso la strada della regolamentazione "forte", attraverso l'adozione di strumenti di *hard law*

[11] Cfr. O. Pollicino – M. Bassini, *Internet Law in the Era of Transnational Law*, in *EUI Working Papers*, 24, 2011; C. Marsden, *Transnational Internet Law*, Oxford, 2020; G. Teubner, *Nuovi conflitti costituzionali. Norme fondamentali dei regimi transnazionali*, Milano, 2012.

[12] Cfr. Kugler, *Deepfake Privacy*, cit., 615.

[13] Cfr. D. Castro – M. McLaughlin – E. Chivot, *Who Is Winning the AI Race: China, the EU or the United States?*, in *datainnovation.org*, 19 agosto 2019; E. Pisanelli, *Intelligenza Artificiale: battaglia globale per tre*, in *ispionline.it*, 27 ottobre 2022.

[14] Sui differenti approcci in materia di IA, cfr. H. Roberts – J. Cowls – E. Hine – J. Morley – V. Wang – M. Taddeo – L. Floridi, *Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes*, in *The Information Society*, 2, 2023, 79-97; E. Hine – L. Floridi, *Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies*, in *AI & Society*, 1, 2024, 257-278; R. Bal – I.S. Gill, *Policy approaches to artificial intelligence based technologies in China, European Union and the United States*, in *Duke Global Working Paper Series*, 26, 2020.

come i regolamenti: dopo il *Digital Markets Act* (DMA)[15] e il *Digital Services Act* (DSA)[16], al momento in cui si scrive è attesa, dopo lunga gestazione, l'entrata in vigore dell'*Artificial Intelligence Act* (AIA)[17].

Se, quindi, con specifico riferimento al tema dei *deepfake* l'UE non è – o non è ancora – intervenuta con un atto normativo *ad hoc*, il fenomeno non resta comunque sconosciuto alle Istituzioni, tanto che all'interno dei regolamenti appena citati possono trovarsi riferimenti più o meno espliciti, nonché disposizioni più o meno applicabili, le quali si aggiungono alle tutele in materia di privacy stabilite dal GDPR[18].

In effetti, già nel 2021 il *Panel for the Future of Science and Technology* (STOA) rendeva al Parlamento europeo indicazioni sui *deepfake* per mezzo di un report intitolato "*Tackling deepfake in European policy*"[19]. Già dal titolo è possibile evincere il disvalore, cui si accennava in apertura, con cui questi utilizzi di IA sono osservati dalla società e dal regolatore stesso: tanto che essi non vanno regolati, ma "contrastati" (*tackling*). Lo studio dello STOA seguiva di pochi mesi la pubblicazione delle prime proposte del DSA e dell'AIA, delle quali evidentemente teneva conto. In particolare, il testo dell'AIA, nella parte in cui fa riferimento ai *deepfake*, è stato modificato solo leggermente rispetto alla versione del 2021: pertanto, le considerazioni esposte nello studio dello STOA non sono state "superate" dal nuovo regolamento. Inoltre, nel report, accanto ai regolamenti in via di approvazione, vengono considerate anche le disposizioni già vigenti del GDPR.

Innanzitutto, i *deepfake* sono definiti come media audio o video manipolati o sintetici che sembrano autentici e che mostrano persone che sembrano dire o fare qualcosa che non hanno mai detto o fatto, prodotti utilizzando tecniche di IA, tra cui l'apprendimento automatico e il *deep learning*[20]. Al netto di un dibattito sempre aperto sulla definizione di IA[21], tale enunciazione non sembra porre particolari problemi applicativi.

Lo STOA rende poi una serie di raccomandazioni. Primariamente, consiglia di chiarire i casi in cui i *deepfake* possano ricomprendersi, ai sensi dell'AIA, tra le pratiche vietate o tra le applicazioni ad alto rischio, valutando l'opportunità di includerli come regola tra i sistemi ad alto rischio e di prevedere specifici *ban* per utilizzi particolarmente pe-

[15] Regolamento (UE) 2022/1925 del Parlamento europeo e del Consiglio del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali), noto in inglese come *Digital Markets Act*.

[16] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), noto in inglese come *Digital Services Act*.

[17] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), noto in inglese come *Artificial Intelligence Act*.

[18] Cfr. B. Van der Sloot – Y. Wagensveld, *Deepfakes: regulatory challenges for the synthetic society*, in *Computer Law & Security Review*, 46, 2022, 7-8.

[19] STOA, *Tackling deepfakes in European policy*, luglio 2021.

[20] Ivi, 2.

[21] Cfr. C. Muller, *The Impact of AI on Human Rights, Democracy and the Rule of Law*, in CAHAI (Comitato ad hoc sull'IA del Consiglio d'Europa), CAHAI(2020)06-fin, 24 giugno 2020, 4 ss.: «AI remains an essentially contested concept, as there is no universally accepted definition».

ricolosi, come quelli riguardanti la pornografia non consensuale o la disinformazione politica[22]. In secondo luogo, con riferimento all'obbligo di etichettare i contenuti per renderne palese la natura, imposto dall'IA ai creatori (*creators*) di *deepfake*, suggerisce di estendere tale obbligo anche ai fornitori (*providers*)[23]. Inoltre, ammonisce dall'utilizzo di eccezioni troppo ampie rispetto a questo obbligo, dato che la proposta dell'IA esclude l'etichettatura per i contenuti usati a fini di prevenzione dei crimini, ma anche artistici e scientifici e comunque quando necessari per la libertà di espressione[24]. In maniera abbastanza singolare, raccomanda di limitare la diffusione delle tecnologie di rilevamento dei *deepfake*, in modo tale da non consentire ai creatori modalità per eludere i controlli, anche se parallelamente si sconsiglia di riservare la *detection* a una cerchia troppo ristretta di attori[25]; infine, sottolinea l'importanza di investire nelle tecnologie di IA in grado di "difendere" dagli attacchi di *deepfake*, oltre che nell'educazione e nella consapevolezza del fenomeno da parte della società[26].

Sebbene il DSA non menzioni direttamente i *deepfake*, l'impostazione generale del regolamento e alcune disposizioni specifiche sembrano riguardare da vicino questo genere di contenuti. In linea generale, alle piattaforme online e ai motori di ricerca di grandissime dimensioni è imposto l'obbligo di effettuare valutazioni specifiche del rischio e mettere in atto misure ragionevoli, proporzionate ed efficaci per prevenire qualsiasi aspetto negativo effettivo o prevedibile sul discorso civico e sui processi elettorali[27]. Si tratta di un obbligo che teoricamente pare interessare almeno quei *deepfake* incidenti sulla disinformazione politica. Agli stessi soggetti il DSA richiede di adottare idonee misure di attenuazione dei rischi e tra queste elenca esplicitamente «il ricorso a un contrassegno ben visibile per fare in modo che un elemento di un'informazione, sia esso un'immagine, un contenuto audio o video, generati o manipolati, che assomigli notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che a una persona appaia falsamente autentico o veritiero, sia distinguibile quando è presentato sulle loro interfacce online e, inoltre, la fornitura di una funzionalità di facile utilizzo che consenta ai destinatari del servizio di indicare tale informazione»[28]. Il legislatore evita di fare riferimento diretto ai *deepfake*, ma evidentemente ragiona sulla loro circolazione. La scelta di utilizzare una perifrasi tanto ampia si giustifica sia per il fatto di non

[22] STOA, *Tackling deepfakes*, cit., 59.

[23] *Ibid.*

[24] Ivi, 60-61.

[25] Ivi, 59.

[26] Ivi, 60.

[27] DSA, art. 34: «I fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi individuano, analizzano e valutano con diligenza gli eventuali rischi sistemici nell'Unione derivanti dalla progettazione o dal funzionamento del loro servizio e dei suoi relativi sistemi, compresi i sistemi algoritmici, o dall'uso dei loro servizi. […] La valutazione del rischio […] deve comprendere i seguenti rischi sistemici: […] eventuali effetti negativi, attuali o prevedibili, per l'esercizio dei diritti fondamentali, in particolare i diritti fondamentali alla dignità umana […]; eventuali effetti negativi, attuali o prevedibili, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica; […] qualsiasi effetto negativo, attuale o prevedibile, in relazione alla violenza di genere, alla protezione della salute pubblica e dei minori e alle gravi conseguenze negative per il benessere fisico e mentale della persona».

[28] DSA, art. 35, par. 1, lett. k).

vincolare piattaforme e motori di ricerca alla messa in atto di misure troppo specifiche, sia probabilmente perché all'epoca dell'adozione del regolamento (2022) l'espressione "*deepfake*" non era ancora così sdoganata come oggi e rischiava di ingenerare equivoci interpretativi. Tuttavia, come effetto, una disposizione così ampia finisce con il riferirsi non soltanto all'IA generativa su cui si fondano i *deepfake*, ma permette interpretazioni eccessivamente estensive che potrebbero mettere in difficoltà i soggetti obbligati.

Con il regolamento sull'IA, come noto, l'UE ha imboccato la strada di un approccio olistico e basato sul rischio per regolamentare tutte le pratiche di intelligenza artificiale[29]. L'AIA prende in esame l'intero ciclo di vita dei sistemi e classifica gli stessi sulla base dei rischi che presentano, distinguendo tecnologie (*rectius*: impieghi delle tecnologie) comunque vietate, ad alto rischio e a basso rischio e prevedendo per ognuna di queste categorie obblighi differenti in capo agli *AI actors*: in particolare, i sistemi ad alto rischio sono soggetti a un regime di conformità con requisiti dettagliati[30]. Occorre verificare a quale/i di queste categorie – pratiche vietate, sistemi ad alto rischio, altri sistemi – possano appartenere i *deepfake*.

Preliminarmente, è doveroso evidenziare che questo genere di contenuti non è esplicitamente menzionato tra le pratiche comunque vietate, né tra i sistemi ad alto rischio, mentre – come si vedrà – essi vengono richiamati in altre disposizioni, oltre ad essere definiti esplicitamente dal regolamento come «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona»[31]. In ogni caso, la disciplina sulle pratiche vietate e sui sistemi ad alto rischio merita di essere analizzata per verificare se possa escludersi completamente l'appartenenza dei *deepfake* a queste categorie.

Con riferimento alle pratiche vietate ai sensi dell'art. 5 AIA, il regolamento mette al bando sia i sistemi che utilizzano «tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso […]»[32], sia i sistemi che sfruttano «le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo […]»[33]. Non appare impossibile immaginare che i *deepfake* possano presentarsi in queste forme, anche se occorre precisare che il divieto sussiste solo laddove dall'u-

[29] Cfr. per un primo commento della disciplina, nelle more dell'iter legislativo, B. Marchetti – C. Casonato, *Prime osservazioni sulla proposta di Regolamento dell'Unione europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 3, 2021, 415-437; G. Finocchiaro, *The regulation of artificial intelligence*, in *AI & Society*, 3 aprile 2023.

[30] V. AIA, Capo III, spec. sez. 2 (artt. 8-15) e sez. 3 (artt. 16-27).

[31] AIA, art. 3, n. 60.

[32] AIA, art. 5, par. 1, lett. a).

[33] AIA, art. 5, par. 1, lett. b).

tilizzo di questi sistemi derivino, o possano derivare, "danni significativi" a persone[34].
Venendo ai sistemi ad alto rischio, ai sensi dell'art. 6 del regolamento devono essere considerati tali quei sistemi utilizzati in determinati settori e per specifici impieghi, stabiliti tassativamente dall'allegato III[35]. Sebbene in questo elenco non vi sia alcun riferimento esplicito ai *deepfake*, resta comunque possibile che essi possano rientrare tra i settori e gli impieghi ad alto rischio, laddove siano utilizzati nell'ambito di settori "sensibili", quali istruzione e formazione professionale, occupazione lavorativa, servizi essenziali, attività di contrasto ai crimini, processi democratici. A ben vedere, è proprio quest'ultimo il settore che più realisticamente potrebbe essere interessato dai *deepfake*, considerato che l'allegato classifica come ad alto rischio «i sistemi di IA destinati a essere utilizzati per influenzare l'esito di un'elezione o di un referendum o il comportamento di voto delle persone fisiche nell'esercizio del loro voto alle elezioni o ai referendum»[36].
Tuttavia, lo stesso art. 6, al par. 3, sembra mitigare gli effetti di questa classificazione, poiché esclude che un sistema, pur indicato nell'allegato III, debba considerarsi ad alto rischio qualora non presenti «un rischio significativo di danno per la salute, la sicurezza o diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale»[37]. Pertanto, se anche i *deepfake* rientrassero tra i sistemi elencati nell'allegato, comunque potrebbero essere "esentati" dalla stringente disciplina riservata ai sistemi ad alto rischio nel caso in cui si riuscisse a provare l'assenza di un rischio significativo.
D'altronde, tale disciplina si applicherebbe ai soli *deepfake* utilizzati in settori "sensibili": questa scelta appare certamente ragionevole poiché motivata sulla base della maggiore astratta pericolosità. Tuttavia, i *deepfake* utilizzati in questi settori potrebbero essere rivolti al raggiungimento di scopi "meritevoli" (ad es., il contrasto ai crimini), ben più di altri *deepfake*, che, per il sol fatto di essere utilizzati in settori non sensibili e a prescindere dalla "meritevolezza" del loro impiego, sarebbero liberi dai vincoli imposti per i sistemi ad alto rischio.
Comunque la si veda, da questa analisi emerge un assunto chiaro: per il regolatore unionale, i *deepfake* non costituiscono di per sé sistemi ad alto rischio, a meno che non riguardino particolari impieghi in settori sensibili.
Tuttavia, il disvalore associato a questo genere di contenuti ha indotto lo stesso legislatore a ritagliare per essi uno spazio specifico, in modo che possano essere distinti dai semplici sistemi a basso rischio (*rectius*: a minimo rischio), per i quali il regolamento

[34] V. art. 5, par. 1, lett. a) e b), identici nella parte finale della disposizione: «[…] in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo».

[35] L'allegato III indica otto settori, nell'ambito dei quali l'utilizzo di sistemi di IA deve essere considerato ad alto rischio: biometria, infrastrutture critiche, istruzione e formazione professionale, occupazione e gestione dei lavoratori, servizi essenziali, contrasto all'illegalità, migrazione e asilo, giustizia e processi democratici. Per ognuno di questi settori sono segnalati specifici impieghi di IA considerati ad alto rischio.

[36] AIA, allegato III, n. 8, lett. b).

[37] AIA, art. 6, par. 3, primo comma. Tuttavia, come è specificato successivamente, questa disposizione si applica solo qualora sia soddisfatta almeno una di determinate condizioni: v. art. 6, par. 3, secondo comma.

non impone particolari obblighi. Proprio con riferimento ai *deepfake* (e a pochi altri sistemi di IA[38]), l'AIA immagina una categoria intermedia tra i sistemi ad alto e i sistemi a minimo rischio. L'art. 50, infatti, stabilisce specifici obblighi di trasparenza per i fornitori e i *deployer* di «determinati sistemi di IA»: in particolare, i *deployer*[39] di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake» sono obbligati a rendere noto che il contenuto è stato generato o manipolato artificialmente[40]. Tale obbligo subisce comunque delle mitigazioni: intanto, non si applica, come logico, se l'utilizzo è autorizzato dalla legge per il contrasto di reati[41]; inoltre, nel caso in cui il contenuto faccia parte di un'opera o un programma «manifestamente artistici, creativi, satirici o fittizi», allora è sufficiente rivelare l'esistenza dei contenuti generati o manipolati «in modo adeguato», senza che risultino ostacolati «l'esposizione o il godimento» dell'opera[42]. La lettera di questa disposizione lascia qualche dubbio interpretativo. Innanzitutto, non si comprende in cosa differirebbe l'obbligo in questione rispetto a quello previsto per tutti i *deepfake*: in particolare, quando la rivelazione (*disclosure*) potrebbe dirsi effettuata «in modo adeguato»? Sembra che la natura del *deepfake* debba essere rivelata senza che possa essere preclusa «l'esposizione o il godimento» dell'opera: quindi sarebbe possibile rimandare la *disclosure* ad un momento successivo? Inoltre, occorre delimitare il perimetro delle opere e dei programmi «manifestamente artistici, creativi, satirici o fittizi»: alla luce di alcuni di questi aggettivi non sembra potersi escludere del tutto che all'interno di questo perimetro possano essere ricompresi anche contenuti pubblicati sui *social* da semplici utenti, in quanto "protetti"

[38] L'art. 50 AIA stabilisce obblighi per i fornitori di «sistemi di IA destinati a interagire direttamente con le persone fisiche» (par. 1) e per i *deployer* di sistemi «di riconoscimento delle emozioni o di […] categorizzazione biometrica» (par. 3).

[39] Mentre la versione in italiano della proposta del 2021 imponeva l'obbligo agli «utenti» (art. 52, par. 3, Proposta di regolamento COM(2021) 206 final, 21 aprile 2021), il regolamento approvato preferisce fare riferimento ai «*deployer*», scegliendo in modo singolare di mantenere il termine inglese anche nella versione italiana, probabilmente considerandolo di problematica traduzione. Stando all'art. 3, n. 4, il *deployer* deve essere inteso come «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

[40] AIA, art. 50, par. 4. Con riferimento agli obblighi per i fornitori, il par. 2 dispone: «I fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici, garantiscono che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente. I fornitori garantiscono che le loro soluzioni tecniche siano efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche».

[41] Tuttavia, in questo caso, i *deepfake* potrebbero essere assoggettati alla disciplina prevista per i sistemi ad alto rischio alle condizioni stabilite dall'art. 6 e dall'allegato III, n. 4.

[42] Si riporta integralmente il contenuto dell'art. 50, par. 4, AIA: «I deployer di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake» rendono noto che il contenuto è stato generato o manipolato artificialmente. Tale obbligo non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati. Qualora il contenuto faccia parte di un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi, gli obblighi di trasparenza di cui al presente paragrafo si limitano all'obbligo di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera».

dalla loro natura "creativa, satirica o fittizia"[43]. In particolare, il riferimento alle opere "manifestamente fittizie" rischia di ampliare enormemente – e forse eccessivamente – il campo dell'eccezione. Infine, pur tacendo i dubbi che attengono alle modalità di divulgazione dei contenuti, il regolamento non prevede sanzioni chiare in caso di mancata osservanza degli obblighi di trasparenza[44].

Le Istituzioni europee, rompendo gli indugi rispetto a GDPR e DSA, hanno più convintamente avviato il percorso di regolamentazione dei *deepfake*, considerandoli nell'AIA come tecnologie (*rectius*: prodotti di tecnologie) di IA che possono essere classificate sulla base del rischio. Tuttavia, il fatto di aver inserito nel regolamento una disciplina specifica (art. 50) rivela la difficoltà di inquadrare questo genere di contenuti secondo i criteri di rischio previsti per l'IA in generale; in aggiunta, tale disciplina non sembra brillare per chiarezza e completezza. Siamo probabilmente agli albori di una regolamentazione che potrebbe necessitare di nuovi approfondimenti già nel prossimo futuro.

3. Stati Uniti: interventi (statali) per ambiti di utilizzo

Come noto, l'approccio statunitense in materia di regolamentazione dell'IA è profondamente diverso rispetto a quello europeo, soprattutto per il fatto che non pare registrarsi – al netto di alcune proposte di legge presentate al Congresso – la volontà di intervenire con atti di *hard law* operanti a livello federale[45]. Questa scelta ha prodotto l'emersione di normative di livello statale, che sempre più tentano di colmare il vuoto regolatorio con riferimento agli utilizzi di alcune nuove tecnologie[46]. Come sta accadendo, ad esempio, per le tecnologie di riconoscimento facciale[47], anche per i *deepfake*, in assenza di un intervento federale, alcuni Stati sono intervenuti per disciplinare il fenomeno. Prima di dare conto di queste normative, sembra opportuno soffermarsi su alcune recenti proposte di legge in materia presentate a livello federale.

Nel settembre 2023, al fine di proteggere la sicurezza nazionale e garantire mezzi di tutela alle vittime, è stato presentato alla Camera dei rappresentanti il *Deepfakes Accountability Act*[48], che stabilirebbe per i creatori di *deepfake* alcuni obblighi comuni e altri spe-

[43] Vero è che l'avverbio «manifestamente» potrebbe servire a limitare l'ampiezza dell'eccezione.

[44] Cfr. per un commento sulla disciplina dei *deepfake* contenuta nell'AIA, già dai tempi della proposta: A. Fernandez, *"Deep fakes": disentangling terms in the proposed EU Artificial Intelligence Act*, in *UFITA Archiv für Medienrecht und Medienwissenschaft*, 2, 2022, 392-433; M. Łabuz, *Regulating deep fakes in the artificial intelligence act*, in *Applied Cybersecurity & Internet Governance*, 1, 2023, 1-42; M. Łabuz, *Deep fakes and the Artificial Intelligence Act – An important signal or a missed opportunity?*, in *Policy & Internet*, 2024, 1-18.

[45] Cfr. B. Marchetti – L. Parona, *La regolazione dell'intelligenza artificiale: Stati Uniti e Unione europea alla ricerca di un possibile equilibrio*, in *DPCE online*, 1, 2022, 237 ss.; E. Chiti – B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Riv. reg. merc.*, 1, 2020.

[46] S. Parinandi – J. Crosson – K. Peterson – S. Nadarevic, *Investigating the politics and content of US State artificial intelligence legislation*, in *Business and Politics*, 2, 2024, 240-262.

[47] Cfr. J. Spivack – C. Garvie, *A taxonomy of legislative approaches to face recognition in the United States*, in A. Kak (a cura di), *Regulating biometrics: Global approaches and urgent questions*, New York, 2020, 86-95.

[48] *Deepfakes Accountability Act*, H.R. 5586, 20 settembre 2023. Allo stato, questa proposta risulta ferma al momento dell'introduzione, non essendo stata discussa dal Congresso.

cifici a seconda del contenuto (audio/video, solo video o solo audio): in linea generale, si imporrebbe l'inserimento di tecnologie, come quelle di provenienza dei contenuti (*content provenance technologies*), sufficienti a identificare chiaramente il contenuto come composto di elementi alterati o come interamente creato attraverso IA generativa o simili[49]. In particolare, a seconda della natura del contenuto, si pretenderebbero dichiarazioni verbali, scritte o icone da integrare nel *deepfake* atte a prevenire qualsiasi fraintendimento[50]. Alla violazione di questi obblighi sarebbe associata una sanzione differente a seconda della offensività del contenuto, stabilendo la pena della reclusione fino a cinque anni nei casi in cui il *deepfake* sia volto ad arrecare molestie attraverso contenuti sessuali, a interferire in un procedimento ufficiale (comprese le elezioni) purché si tratti di minaccia credibile, a porre in essere frodi o furti di identità, a influenzare un dibattito pubblico interno nell'interesse di una potenza straniera[51]. Le eccezioni risulterebbero abbastanza circoscritte: oltre a consentire contenuti utilizzati dalle forze dell'ordine per la tutela della pubblica sicurezza, verrebbero esentati dalla disciplina i contenuti pubblicati in un contesto tale che una persona ragionevole non potrebbe confondere l'attività falsificata con l'attività effettiva della persona esposta, come nel caso di parodie, rievocazioni storiche o programmi radiofonici, televisivi o cinematografici manifestamente fittizi[52]. Interessanti anche gli strumenti di prevenzione prospettati: Intanto, l'istituzione di una *task force* all'interno del Dipartimento di sicurezza nazionale[53]; inoltre, l'obbligo imposto agli sviluppatori di tecnologie utili alla creazione di *deepfake* di garantire la capacità tecnica del prodotto di inserire la provenienza dei contenuti, unitamente al dovere di rivolgere all'utente una informativa sugli obblighi attinenti alla creazione di *deepfake*[54]; infine, la pretesa che i fornitori di piattaforme online garantiscano non solo la capacità tecnica per l'indicazione della provenienza del contenuto, ma che si dotino di un sistema per il rilevamento dei *deepfake*[55].

All'inizio del 2024, è stato presentato un altro progetto di legge noto come *DEFIANCE Act*[56], approvato dal Senato nel mese di luglio con emendamenti, ai sensi del quale sarebbe garantita la tutela in via giudiziaria alle vittime di *deepfake* "intimi" diffusi senza il loro consenso: si tratta di estendere anche al caso dei *deepfake* la tutela già prevista per la divulgazione non consensuale di *intimate images*[57].

Recentissima, infine, la proposta nota come *COPIED Act* (*Content Origin Protection and*

[49] Ivi, sec. 2(b).

[50] Ivi, sec. 2(c)(d)(e). Singolare la disposizione che per i deepfake audio di durata superiore a due minuti obbligherebbe i creatori a inserire una dichiarazione verbale per chiarire la natura del contenuto ad intervalli di due minuti.

[51] Ivi, sec. 2(f)(1). Per le sanzioni civili e i mezzi a tutela delle vittime, v. ivi, sec. 2, §1041(f)(2)(g)(h).

[52] Ivi, sec. 2(j).

[53] Ivi, sec. 7(a).

[54] Ivi, sec. 10(a).

[55] Ivi, sec. 10(b).

[56] *Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024* (*DEFIANCE Act*), S. 3696, 30 gennaio 2024.

[57] Cfr. 15 USC § 6851.

Integrity from Edited and Deepfaked Media Act)[58], volta a semplificare l'identificazione dei contenuti generati con l'IA e la tutela del diritto d'autore. Il progetto, abbastanza ambizioso, riporta la questione sul piano generale dei sistemi di IA e richiede la presenza di elementi distintivi che consentano di identificare il lavoro dell'IA da quello di un essere umano in modo rapido e standardizzato, ad esempio attraverso l'apposizione obbligatoria di filigrane digitali inerenti all'origine e all'autenticità del contenuto. A protezione del diritto d'autore, la legge vieterebbe l'addestramento di modelli di IA attraverso materiali protetti da copyright senza il consenso dei titolari dei diritti, introducendo sanzioni per le violazioni. L'iter di questo progetto merita di essere osservato con attenzione in riferimento al nostro tema, dato che tali disposizioni, pensate per tutti i contenuti IA, riguarderebbero in primo luogo proprio i *deepfake*.

Se sul piano federale si resta nel campo delle proposte di legge, come accennato, alcuni Stati si sono già dotati di una disciplina in materia. Da questo punto di vista, si possono rintracciare due chiare tendenze: alcuni interventi normativi, infatti, si focalizzano sui *deepfake* pornografici, mentre altri prendono in esame la questione dei *deepfake* politici.

La California ha guidato la carica, approvando nel 2019 due leggi in materia, dedicate appunto alle due applicazioni specifiche anzidette, ossia l'influenza delle campagne politiche[59] e l'utilizzo in ambito pornografico[60]. Occorre evidenziare che la legge sui *deepfake* politici era, in realtà, sottoposta a una *sunset clause* che ne prevedeva l'automatica abrogazione al 1° gennaio 2023, salvo ulteriori determinazioni del legislatore, che non sono però intervenute. Questa legge vietava, nei sessanta giorni precedenti una elezione, la distribuzione di "supporti audio o visivi materialmente ingannevoli", avvenuta con effettiva malizia (*actual malice*) e con l'intento di danneggiare la reputazione di un candidato o di ingannare gli elettori[61]. Le sanzioni non trovavano comunque applicazione nel caso di contenuti satirici o parodistici né per gli operatori dell'informazione che, anche attraverso internet, diffondevano i *deepfake* affermandone però con chiarezza la natura ingannevole[62]. Come detto, tale legge non risulta più in vigore, mentre resta pienamente applicabile la legge sui *deepfake* pornografici, che prevede il diritto di agire in giudizio contro chi distribuisce intenzionalmente *deepfake* di foto o video aventi natura intima o sessuale senza il consenso della persona ritratta[63]. Le eccezioni, pur previste, sembrano abbastanza difficili da realizzarsi, dato che dovrebbe dimostrarsi l'interesse pubblico alla divulgazione, oppure il valore politico o giornalistico del contenuto o, in

[58] La presentazione del *Content Origin Protection and Integrity from Edited and Deepfaked Media Act of 2024 (COPIED Act)* è stata annunciata in data 11 luglio 2024 con un comunicato stampa reperibile su *commerce.senate.gov*.

[59] California, AB no. 730, chap. 493, 10 aprile 2019.

[60] California, AB no. 602, chap. 491, 3 ottobre 2019.

[61] California, AB 730, cit., sec. 4(a). In particolare, per «materially deceptive audio or visual media» si intendeva un'immagine o registrazione audio o video dell'aspetto, del discorso o della condotta di un candidato intenzionalmente manipolata in modo tale che fossero soddisfatte entrambe le seguenti condizioni: il contenuto doveva apparire falsamente autentico a una persona ragionevole; poteva indurre una persona ragionevole ad avere una comprensione o un'impressione sostanzialmente diversa rispetto a quella che avrebbe avuto ascoltando o vedendo la versione inalterata.

[62] California, AB 730, cit., sec. 4(d)(4)(5).

[63] California, AB 602, cit., sec. 1(b). Come altri legislatori statali, anche quello californiano preferisce utilizzare il termine "*digitization*" rispetto a "*deepfake*" nell'ambito della pornografia non consensuale.

generale, la protezione costituzionale dell'attività realizzata[64].

Altri Stati stanno hanno seguito l'esempio californiano, nel senso di una o dell'altra direzione regolatoria. Con riferimento ai *deepfake* pornografici, nel 2019 in Virginia è stata approvata una legge che, nel tentativo di prevenire il c.d. *revenge porn*, sanziona penalmente la distribuzione di *deepfake* pornografici se idonei a costringere, molestare o intimidire una persona[65]. In Florida, dal 2022 le sanzioni penali relative alla pedopornografia e alla pornografia non consensuale sono state estese, fatti esplicitamente salvi gli internet *provider*, a "chi promuove" («*A person who […] promotes*») *deepfake*, a nulla valendo eventuali filigrane o etichettatura del contenuto: sebbene il termine *deepfake* non sia utilizzato esplicitamente, è chiaro che contenuti di tal genere risultino inclusi nella amplissima definizione di "immagini create, alterate, adattate o modificate con mezzi elettronici, meccanici o di altro tipo"[66]. In senso simile hanno provveduto anche Louisiana[67], South Dakota[68] e Washington[69].

In relazione ai *deepfake* politici, il Texas ha approvato nel 2019 una legge, simile alla AB730 della California, che impedisce la distribuzione di *deepfake* politici entro trenta giorni dalle elezioni[70]. Dal 2024 in Mississippi è previsto il reato di diffusione di "*digitization*" (termine preferito dal legislatore rispetto al più comune deepfake) nei novanta giorni precedenti le elezioni, nel caso in cui, mancando il consenso della persona ritratta, chi diffonde il materiale ne conosce la natura e mira a influenzare il dibattito elettorale[71]. In New Mexico, nel 2024, è stata introdotta una legislazione, molto simile alla proposta federale nota come *Deepfakes Accountability Act*, che impone di etichettare i contenuti in modo da renderne chiara la natura, oltre a sanzionare penalmente la condotta di chi diffonde o si accorda con altri per diffondere "*materially deceptive media*"[72]. Simili discipline, con riferimento all'obbligo di etichettatura dei contenuti, sono state approvate anche in Indiana[73] e in Oregon[74].

A margine della bipartizione appena analizzata, merita di essere segnalato l'*Ensuring Likeness, Voice and Image Security (ELVIS) Act*, recentemente approvato in Tennessee, che aggiorna e sostituisce il *Personal Rights Protection Act* del 1984, prevedendo una sanzione civile a carico di chi rende disponibile al pubblico "*voice or likeness*" senza autorizzazione del titolare del diritto[75]. Sebbene concepito per la tutela della proprietà intellettuale, tale provvedimento impatta sulla legittimità della creazione di *deepfake*. Proprio per questo sono previste importanti eccezioni alla disciplina per utilizzi, protetti dal

64 California, AB 602, cit., sec. 1(c)(1).

65 Virginia, H.B. 2678, 18 marzo 2019.

66 Florida, S.B. 1798, 24 giugno 2022.

67 Louisiana, S.B. 175 (Act 457), 28 giugno 2023.

68 South Dakota, S.B. 79, 13 febbraio 2024.

69 Washington, S.B. 1999, 6 giugno 2024.

70 Texas, S.B. 751, 25 maggio 2019.

71 Mississippi, S.B. 2577, 30 aprile 2024.

72 Nex Mexico, H.B. 182, 5 marzo 2024.

73 Indiana, H.B. 1133, 12 marzo 2024.

74 Oregon, S,B, 1571, 27 marzo 2024.

75 Tennessee, H.B. 2091, 26 marzo 2024 (*ELVIS Act*).

Primo emendamento, connessi a fini giornalistici, informativi, educativi, satirici e parodistici[76].

L'analisi del quadro normativo federale e statale negli Stati Uniti lascia la sensazione di una tendenza a regolamentare solo debolmente i *deepfake*. Per quanto debba essere notato che il dibattito e gli interventi normativi siano in clamorosa crescita negli ultimi anni, comunque l'attenzione è circoscritta a contesti specifici, come i procedimenti elettorali o la pornografia. Questo orientamento risente probabilmente della notoria rilevanza che nel diritto costituzionale statunitense è riconosciuta al Primo emendamento, del quale i *deepfake* costituirebbero in linea generale una forma di espressione. Infatti, alla luce del caso *Alvarez*, deciso nel 2012 dalla Corte Suprema[77], anche le dichiarazioni false possono ricevere la protezione del Primo emendamento, a meno che non provochino danni gravi legalmente riconoscibili: in questo senso si spiega l'attenzione dei legislatori statali per le interferenze elettorali o la pornografia non consensuale. Resta pertanto controverso un intervento normativo che limiti il ricorso a questi contenuti, a meno che essi non mettano seriamente in pericolo beni di innegabile rilevanza, come la sicurezza pubblica, la dignità umana, il pudore sessuale e la democraticità delle elezioni.

In aggiunta, il diritto statunitense, più marcatamente rispetto al diritto dell'UE e a quello cinese, si caratterizza per l'alto grado di protezione che assicura agli *internet service provider*, ai motori di ricerca e alle piattaforme online, attraverso cui circolano comunemente i *deepfake*: questi soggetti sono infatti generalmente esentati da responsabilità per i contenuti presenti sui loro siti e app. Tale orientamento sembra confermato dalle leggi statali sui *deepfake*, mentre a livello federale occorre evidenziare il differente atteggiamento del *Deepfake Accountability Act* – comunque fermo al momento della presentazione della proposta –, che quantomeno immagina determinati obblighi per questi soggetti. Sul punto era attesa la pronuncia della Corte suprema nel recente caso *Gonzalez v. Google*[78], che avrebbe potuto rivedere l'orientamento sulla irresponsabilità dei *provider*: invece, nonostante le grandi aspettative, i giudici hanno concisamente ribadito che questi soggetti non possono rispondere come responsabili per i contenuti offensivi pubblicati sui propri siti, manifestando una sorta di disagio istituzionale per essere stati chiamati a pronunciarsi su un tema che richiederebbe l'intervento del Congresso.

4. Cina: *deepfake, deep synthesis* e *"deep control".*

Come noto, il governo cinese ha sviluppato una visione strategica mirata a consolidare la propria "sovranità informatica", esercitando un controllo rigoroso sul cyberspazio

[76] *Elvis Act*, cit., sec. 10(a).

[77] Corte Suprema USA, *United States v. Alvarez*, 567 U.S. 709 (2012).

[78] Corte Suprema USA, *Gonzalez v. Google LLC*, 598 U.S. 617 (2023). A seguito dell'attentato terroristico avvenuto al Bataclan di Parigi nel 2015, i parenti di una vittima statunitense, Nohemi Gonzalez, intentavano causa contro *Google* sostenendone la responsabilità indiretta, in quanto l'organizzazione terroristica ISIS aveva potuto diffondere i propri pericolosi messaggi sulla piattaforma *Youtube*, gestita da *Google*.

e sulle tecnologie emergenti, in particolare l'intelligenza artificiale (IA)[79]. Questo approccio si manifesta attraverso l'implementazione di normative che regolano l'uso e lo sviluppo dell'IA, con l'obiettivo di garantire che tali tecnologie siano allineate con gli interessi nazionali e i valori sociali cinesi. Così si giustificano anche i noti controlli governativi sulle aziende tecnologiche, atti ad evitare, tra le altre cose, che possano emergere figure imprenditoriali con potenziale influenza politica[80].

La determinazione della Cina nel consolidare la propria sovranità informatica, perseguendo una politica che integra sviluppo economico, sicurezza nazionale e influenza geopolitica nel contesto digitale globale, non risparmia neanche il fenomeno dei *deepfake*, a cui, anzi, il regolatore pubblico ha rivolto presto la propria attenzione.

Anche in Cina, come in California, il fermento intorno alla regolamentazione dei *deepfake* ha avuto inizio nel 2019, probabilmente a causa dell'aumento della popolarità delle app – in particolare, *ZAO* – che consentivano la creazione di questi contenuti, soprattutto attraverso lo "scambio di volti", e suscitavano polemiche per i profili relativi alla raccolta dei dati[81]. Le autorità cinesi, non soddisfatte dell'adeguamento delle proprie politiche privacy portato avanti dalla stessa ZAO, hanno presto cominciato a ragionare sulla possibilità di regolamentare, se non vietare drasticamente, i *deepfake*. Così, entro tre mesi dal lancio di ZAO, la *Cyberspace Administration of China* pubblicava diversi documenti in cui si discuteva della necessità di regolamentare l'IA e gestirne lo sviluppo[82]. Entro la fine dell'anno, con entrata in vigore a partire dal 2020, sono stati approvati i "*Regulations on the Administration of Online Audio and Video Information Services*", che hanno stabilito nuove regole e responsabilità, anche sul piano penale, sia per i fornitori che per gli utenti, introducendo di fatto un divieto ampio e generalizzato con riguardo all'uso di immagini, audio e video generati attraverso tecnologie di sintesi profonda per creare o diffondere notizie false[83]. Tale scelta è stata giustificata sulla base della considerazione che l'utilizzo di nuove tecnologie può turbare l'ordine sociale e violare gli interessi delle persone, generando rischi politici e un impatto negativo sulla sicurezza nazionale e la stabilità sociale. Queste norme si rivolgono in larga parte alle piattaforme online, chiamate a rafforzare l'autoregolamentazione nel settore, a istituire un sistema di responsabilità editoriale, a garantire la sicurezza informatica, a verificare la reale identità degli utenti, a segnalare adeguatamente contenuti non reali, a interrompere la circolazione dei contenuti non consentiti e, in generale, ad accettare consapevolmente il controllo sociale.

Si tratta di atti normativi in perfetta sintonia con la direzione intrapresa dal governo cinese, completamente opposta a quella statunitense, di esercitare il proprio controllo

[79] Cfr. G. Santoni, *La Cina e lo spazio digitale. Questioni di governance nello spazio digitale globale*, in *OrizzonteCina*, 3, 2020, 70-75; M. Kolton, *Interpreting China's pursuit of cyber sovereignty and its views on cyber deterrence*, in *The Cyber Defense Review*, 2, 2017, 1 ss.; H. Gu, *Data, big tech, and the new concept of sovereignty*, in *Journal of Chinese political science*, 2023, 591-612; L. Formichella, *La disciplina normativa sulla protezione dei dati e delle informazioni personali in Cina: emersione di un nuovo paradigma nel contesto internazionale*, in *Mondo Cinese*, 3, 2021, 85-94.

[80] Cfr. A.H. Zhang, *High Wire: How China Regulates Big Tech and Governs Its Economy*, Oxford, 2024.

[81] Cfr. Geng, *Comparing "Deepfake" Regulatory Regimes*, cit., 167 ss.

[82] *Ibid.*

[83] *Administrative Regulations on Online Audio and Video Information Services*, 18 novembre 2019. Cfr. *China issues regulation for online audio, video services*, in *english.www.gov.cn* 30 novembre 2019.

sulle nuove tecnologie e sui loro effetti attraverso interventi statali che richiedono una "cooperazione", spesso sotto forma di veri e propri obblighi, alle grandi imprese operanti nel settore[84].

Nel pieno perseguimento di questa politica, nel corso del 2022, a seguito di una procedura di consultazione pubblica e prevedendo l'entrata in vigore per l'anno successivo, la stessa *Cyberspace Administration of China* ha introdotto un'altra serie di norme, note come *Regulations on Deep Synthesis Management of Internet Information Services*, volta a regolamentare le attività di sintesi profonda su Internet in un'ottica di promozione dei "valori socialisti fondamentali", oltre che di protezione della sicurezza nazionale e dei diritti dei cittadini[85]. Queste norme, quindi, riguardano più specificamente i *deepfake*, anche se la definizione utilizzata – "tecnologie di sintesi profonda" – sembra riferirsi a un più ampio spettro di tecnologie, tra cui potrebbero rientrare, ad esempio, ambienti virtuali e *chatbot*. Di fatto, al fine di operare un controllo più incisivo, l'attenzione è posta non sui risultati dell'utilizzo della tecnologia (quali sono i *deepfake*), ma sulla tecnologia stessa. Vero è che il regolamento non stabilisce un divieto generalizzato alla creazione di *deepfake*: tuttavia, i confini dei contenuti non consentiti sembrano particolarmente vaghi. Da un lato, i servizi di sintesi profonda non devono essere utilizzati da alcuna organizzazione o individuo per produrre, riprodurre, pubblicare o trasmettere informazioni proibite da leggi o regolamenti amministrativi, o per intraprendere attività proibite da leggi e regolamenti amministrativi, come quelle che mettono a repentaglio la sicurezza e gli interessi nazionali, danneggiano l'immagine della nazione, danneggiano l'interesse pubblico della società, disturbano l'ordine economico o sociale o danneggiano i diritti e gli interessi legittimi di altri[86]. Dall'altro, come rinforzo al divieto, è precisato che i fornitori e gli utenti di servizi di sintesi profonda non devono utilizzare servizi di sintesi profonda per produrre, riprodurre, pubblicare o trasmettere informazioni di notizie false[87]. Se già i contenuti della prima disposizione lasciano aperti dubbi interpretativi – ad esempio, quali contenuti, stando ai valori promossi dal governo cinese, potrebbero dirsi dannosi per l'immagine della nazione o l'ordine economico e sociale? –, la seconda precisazione sembra bandire qualsiasi contenuto "falso": in questo senso, astrattamente qualsiasi *deepfake* potrebbe essere considerato come tale.

In piena coerenza con l'orientamento politico sopra descritto, anche la normativa sulla sintesi profonda è incentrata sugli obblighi imposti ai *provider*, i quali sono chiamati ancora una volta, salvo incorrere in sanzioni sul piano sia civile che penale, a verificare l'identità degli utenti, a etichettare adeguatamente i contenuti, a rafforzare la gestione dei

---

[84] Cfr. E. Hine – L. Floridi, *New deepfake regulations in China are a tool for social stability, but at what cost?*, in Nature Machine Intelligence, 4, 2022, 608-610. Come spiegato dagli autori, si tratta di una tendenza consolidata, poiché il governo fa sempre più affidamento sulle aziende tecnologiche per applicare nuove normative su Internet per facilitare la visione del Partito Comunista Cinese (PCC) di una società stabile e prospera, prevedendo conseguenze in caso di mancata collaborazione. A riprova di ciò, ripercorrono la vicenda del presidente esecutivo della piattaforma *Alibaba*, Jack Ma, addirittura scomparso per mesi agli occhi dell'opinione pubblica a seguito di dissidi con il governo.

[85] *Regulations on Deep Synthesis Management of Internet Information Services*, 25 novembre 2022 (per la traduzione in lingua inglese, cfr. *chinalawtranslate.com*).

[86] Ivi, art. 6.

[87] *Ibid.*

dati e, soprattutto, a segnalare agli utenti l'obbligo di ottenere il consenso da parte delle persone interessate prima della produzione del contenuto di sintesi profonda (con specifico riferimento alla alterazione di volti e voci)[88]. Si tratta di disposizioni particolarmente gravose per gli operatori, che, a conti fatti, potrebbero dimostrarsi comunque insufficienti. In effetti, fermo restando che non risultano chiare le modalità con cui dovrebbero essere segnalati o etichettati i contenuti, i fornitori dovrebbero provvedere in tal senso per qualsiasi prodotto di sintesi profonda, a prescindere dallo scopo e dalla pericolosità più o meno elevata. Inoltre, le più comuni modalità di etichettatura – come l'apposizione di filigrane – potrebbero comunque essere rimosse mediante l'utilizzo di altre tecnologie, magari anche queste basate sulla *deep synthesis*, con frustrazione di tutti gli sforzi effettuati dai gestori delle piattaforme, che rimarrebbero ancora esposti alle responsabilità conseguenti alla circolazione dei contenuti. Infine, l'obbligo di interruzione immediata della trasmissione di contenuti non etichettati sembra particolarmente velleitario se consideriamo che, una volta creato, il contenuto può essere facilmente disaccoppiato dal servizio su cui è stato creato e diffuso indipendentemente da esso, oltre alla possibilità che lo stesso contenuto venga ricaricato o catturato tramite *screenshot*, quindi sottratto dal controllo dell'autore e del fornitore di servizi originario.

5. Tre approcci rivelatori di tendenze, visioni e obiettivi

Nell'affrontare le sfide della particolare categoria di output generati dall'IA noti come *deepfake*, Unione europea, Stati Uniti e Cina hanno adottati approcci diversi, che, come detto in apertura, ricalcano in buona parte le scelte finora intraprese in materia di regolamentazione generale dell'IA. Gli approcci si distinguono per quanto riguarda le regole proposte, la ripartizione delle responsabilità tra i soggetti coinvolti e le conseguenze derivanti dalle violazioni.

In via preliminare, occorre osservare un dato comune: in nessun ordinamento è operativo un divieto generalizzato dei *deepfake*. Nonostante la recente normativa cinese in materia, specialmente se letta in combinato disposto con i precedenti regolamenti relativi ai servizi online, presenti contorni particolarmente sfumati che rendono di fatto a rischio illegalità qualsiasi contenuto *deepfake*, resta almeno astrattamente accettata l'impraticabilità di un divieto generalizzato, che costituirebbe una illegittima compressione della libertà di espressione[89]. All'opposto, proprio facendo leva sul Primo emendamento, negli Stati Uniti le limitazioni ai *deepfake* continuano a latitare a livello federale e restano abbastanza circoscritte a livello statale. Al di là della indubbia valenza dei principi, la regolamentazione più o meno stringente non può che incidere anche sulla libertà di impresa degli operatori privati del settore tecnologico, i quali contribuiscono alla creazione e alla diffusione dei *deepfake*: da questo punto di vista, mentre nel sistema cinese il controllo sui privati è connaturato all'organizzazione socioeconomica del pae-

[88] Ivi, artt. 14-18.

[89] Sul tema della libertà d'espressione nel mondo digitale, cfr. P. Tanzarella, *La trasformazione della libertà d'espressione dal mondo liberale al mondo digitale. Quale futuro per i principi classici del costituzionalismo?*, in V. Faggiani – G.B. Sales Sarlet (a cura di), *Retos del derecho ante la IA*, Barcellona, 2024, 103-137.

se, nei sistemi europeo e statunitense restrizioni eccessive, magari unite all'attribuzione di obblighi e responsabilità, rischiano di disincentivare gli investimenti nel settore, con tutto ciò che ne consegue anche in termini di posizione economica dell'UE e degli USA in relazione allo sviluppo dell'IA.

Venendo a individuare alcuni criteri classificatori delle discipline analizzate nei paragrafi precedenti, sembra innanzitutto opportuno soffermarsi sulla definizione della categoria. Sul punto non sembrano sussistere particolari divergenze tra i vari sistemi, atteso che il *deepfake*, accogliendo la definizione dello STOA dell'UE, può essere inteso come media audio o video manipolato o sintetico che sembra autentico e che mostra persone che sembrano dire o fare qualcosa che non hanno mai detto o fatto, prodotto utilizzando tecniche di IA, tra cui l'apprendimento automatico e il *deep learning*. Sebbene non possa considerarsi questa definizione come standardizzata a livello mondiale, comunque le caratteristiche di questo genere di contenuti rimangono identificate in maniera pressoché analoga. Tutt'al più, può notarsi come difficilmente si opti per una regolamentazione circoscritta al fenomeno dei *deepfake* specificamente intesi; soprattutto, appare singolare il fatto che le normative non facciano così frequentemente riferimento al termine "*deepfake*". Nell'UE, la disciplina è integrata all'interno della regolamentazione olistica dell'IA e comunque sembra particolarmente invischiata nell'impianto generale del regolamento, tanto da lasciare dubbi sulla classificazione del grado di rischio attribuito a queste tecnologie; negli USA, ad eccezione della proposta federale nota appunto come *Deepfakes Accountability Act*, la regolamentazione statale sceglie spesso di riferirsi al fenomeno con delle perifrasi o addirittura con altri termini alquanto inconsueti (ad es., *digitization*); in Cina, al centro della regolamentazione sono finiti i contenuti prodotto di *deep synthesis*. Tale ritrosia ad utilizzare il termine "*deepfake*" svela, da un lato, una non completa padronanza sul piano semantico, che evidentemente preoccupa i legislatori; dall'altro, denota pure come sembri ancora oggi maggiormente opportuno regolamentare il fenomeno in un'ottica più ampia, ragionando su discipline che, mentre pensano ai *deepfake*, immaginano quadri regolatori anche per altri utilizzi dell'IA magari ancora non così diffusi. Tale atteggiamento potrebbe "allungare la vita" delle discipline che stanno entrando in vigore, ma potrebbe prestare il fianco ad una regolamentazione dei *deepfake* intesi in senso stretto ancora incerta e poco incisiva.

Un altro criterio classificatorio può essere rintracciato guardando al rapporto tra intervento normativo e fasi specifiche del ciclo di vita dei *deepfake*. Come per ogni sistema di IA, questo può essere scomposto in varie fasi: intanto, esiste il momento della nascita della tecnologia di IA attraverso cui i contenuti saranno prodotti; dopodiché, si può distinguere la fase di creazione del *deepfake* da quella della sua circolazione e ancora da quella della visualizzazione da parte del pubblico. In ognuna di queste fasi si trovano coinvolti soggetti diversi, che possono sovrapporsi e giocare un ruolo in momenti differenti: soprattutto, possiamo distinguere i fornitori dei servizi che consentono la creazione dei contenuti, i fornitori dei servizi che ne permettono la pubblicazione e la trasmissione, i soggetti (sostanzialmente, utenti) che creano i contenuti, quelli che li fanno circolare e quelli che li visualizzano. Detto che una efficace regolamentazione dei *deepfake* dovrebbe tenere in conto tutte le fasi del ciclo di vita del sistema, si possono comunque notare alcune chiare differenze tra gli ordinamenti presi in esame con rife-

rimento all'enfasi riservata ad alcune fasi rispetto ad altre. Nell'UE e in Cina il focus è sulle prime fasi del ciclo di vita (tecnologia utilizzata per la produzione, creazione e circolazione del contenuto), mentre negli Stati Uniti le normative tendono a enfatizzare sulle fasi successive (effetti del contenuto sul pubblico). Ciò non significa che le altre fasi vengano trascurate del tutto: nelle leggi statali statunitensi, per esempio, si ritrovano obblighi di etichettatura dei contenuti attinenti alle prime fasi, ma l'attenzione è posta molto spesso sul piano dei rimedi da accordare ai soggetti danneggiati, cui si cerca di assicurare una tutela legale sul piano riparatorio/risarcitorio che eventualmente si affianchi alle sanzioni penali.

La differente enfasi regolatoria sulle fasi del ciclo di vita dei *deepfake* ha i suoi effetti sulla ripartizione delle responsabilità tra i soggetti coinvolti. Mentre in Cina e nell'UE l'attenzione per le prime fasi pone l'accento sulla responsabilità dei fornitori di servizi, negli Stati Uniti i rimedi garantiti alle vittime finiscono con l'obbligare soprattutto chi diffonde i contenuti dannosi. Anche in questo caso la differenza appare sfumata, dato che sia in Cina che nell'UE le responsabilità si espandono quantomeno ai creatori di contenuti; tuttavia, di fondo resta vero che negli USA la responsabilità dei *provider*, anche alla luce della recente giurisprudenza, è ancora ipotesi poco percorsa, mentre i *creators* continuano a essere protetti, finché possibile, dal Primo emendamento sulla libertà di espressione.

Un secondo criterio classificatorio attiene alla scelta di regolamentare il fenomeno con riferimento agli specifici utilizzi (ed effetti che ne conseguono) o con riferimento al livello di rischio, considerato che una valutazione del secondo tipo ragiona su fattispecie ed effetti maggiormente astratti e potenziali, ma potrebbe garantire una prevenzione più efficace dei danni. D'altronde, non si tratta di orientamenti perfettamente contrapposti: nell'UE, infatti, laddove la scelta è chiaramente votata alla classificazione sulla base del rischio (*risk-based approach*), questa scaturisce comunque dall'individuazione di settori e utilizzi particolarmente sensibili. Il tema è collegato alla scelta di regolamentare il fenomeno singolarmente o come parte di una strategia olistica, magari costruita sull'IA in generale. Da questo punto di vista, appare innegabile che siamo di fronte a tre impostazioni differenti: in Cina, la regolamentazione è abbastanza generalizzata se si osservano le norme di applicazione generale in materia di servizi e media online, ma deve notarsi l'introduzione di una disciplina dedicata nello specifico alle tecnologie di sintesi profonda, insieme comunque più ampio rispetto alla categoria dei *deepfake*; nell'UE, se è vero che la disciplina in materia si ritrova all'interno del regolamento generale sull'IA e che la classificazione dei *deepfake* sulla base del rischio appare ancora incerta e nebulosa, comunque deve essere riconosciuto che un riferimento ad essi fa capolino in una parte a sé dell'AIA (art. 50); negli Stati Uniti, infine, la tendenza abbastanza netta è quella di regolamentare il fenomeno esclusivamente in relazione a due specifiche aree di utilizzo, ossia l'influenza sui procedimenti elettorali e la pornografia non consensuale: in questo modo, la valutazione sulla pericolosità delle tecnologie è compiuta a monte dal legislatore, circoscrivendo chiaramente i settori che necessitano di intervento.

Da ultimo, possiamo classificare i tre approcci nel quadro della visione strategica che le rispettive autorità politiche stanno portando avanti in materia di gestione e sviluppo

delle nuove tecnologie, specialmente dell'IA. Partendo dal presupposto che tutte e tre le potenze puntano a migliorare o consolidare la propria posizione in ordine allo sfruttamento e allo sviluppo dell'IA[90], l'analisi delle prime discipline sui *deepfake* confermano le tendenze che valgono per il settore dell'IA in generale. Per quanto riguarda l'UE, sebbene con maggiore timidezza rispetto ad altre applicazioni di IA, il legislatore prevede obblighi specifici per i *deepfake*, senza che si possa escludere l'applicabilità della più restrittiva disciplina congegnata per i sistemi ad alto rischio: gli operatori privati ne risultano certamente responsabilizzati e lo spazio per l'autoregolamentazione, pur esistente, risulta comunque compresso. Inoltre, la scelta di introdurre regole attraverso uno strumento di *hard law* direttamente applicabile in tutti gli Stati membri rivela la netta preferenza per l'armonizzazione normativa, nel tentativo di offrire un quadro regolatorio chiaro e definito sia ai cittadini sia agli operatori del settore. La visione strategica degli USA, probabilmente forti di una posizione maggiormente dominante nel settore dell'IA, non coincide con quella europea. Come per altre applicazioni di IA[91], anche per i *deepfake* la scelta segue una duplice direzione: sul piano dell'imposizione di obblighi e divieti, si continua a preferire, finché possibile, la strada della "*no regulation*" o al più ad incoraggiare l'autoregolamentazione privata, limitando l'intervento del regolatore pubblico a casistiche eccezionali; sul piano della competenza, all'opposto della strada seguita dall'UE, si consolida, almeno per ora, la tendenza del legislatore federale ad evitare interventi normativi, a fronte di legislatori statali che sempre più frequentemente optano per una regolamentazione delle nuove tecnologie. Infine, in Cina l'intervento del governo in materia di nuove tecnologie assume contorni sempre più evidenti, all'interno di un sistema in cui gli operatori del settore sono chiamati a condividere integralmente la visione e i valori promossi dal partito, con aumento del controllo e degli obblighi cui sono sottoposti.

Tuttavia, e ferme restando tutte le differenze evidenziate, vale la pena evidenziare che il tema dei *deepfake* sembra unire i tre approcci, nella misura in cui comunque la diffusione di questo genere di contenuti, almeno in alcuni settori particolarmente sensibili, è guardata con grande sospetto dai regolatori pubblici, tanto da comportare l'introduzione di normative *ad hoc*: in particolare, l'esigenza di preservare le elezioni da indebiti condizionamenti o influenze è avvertita da tutte le autorità di governo, quale che sia l'orientamento politico o ideologico. In questo senso, fatte salve le differenti concezioni sul piano filosofico, giuridico e sociale, la riflessione sulla regolamentazione dei *deepfake*, soprattutto se portata avanti in un contesto di cooperazione tra le autorità politiche delle tre potenze, potrebbe costituire un "ponte" per avvicinare le visioni sul futuro e sui rischi dell'IA.

[90] Cfr. A. Bradford, *Digital empires: The global battle to regulate technology*, Oxford, 2023.

[91] Il riferimento è, ad esempio, alle tecnologie di riconoscimento facciale. Cfr., per una comparazione in materia tra i tre sistemi, W. Chen – M. Wang, *Regulating the use of facial recognition technology across borders: A comparative case analysis of the European Union, the United States, and China*, in *Telecommunications Policy*, 2, 2023.

L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del *deep fake*[*]

Giuseppe Proietti

Abstract

Il contributo intende offrire un quadro della società dell'informazione plasmata dall'uso dei più recenti strumenti tecnologici inquadrando le opportunità e i rischi che ne derivano. L'analisi delle questioni sollevate da questi strumenti necessita di uno sforzo ermeneutico che mira a combinare normative sempre più complementari e interdipendenti. Quindi, oltre a un necessario riferimento all'importanza della normativa in materia di dati personali, ci si concentra sulla recente legislazione e sul suo impatto nel campo dell'informazione e della comunicazione, con un'attenzione particolare al fenomeno dei *deep fake*. L'analisi si basa sull'Artificial Intelligence Act e sul Digital Services Act, il quale mira a ridurre o eliminarne i contenuti illegali dagli ambienti digitali.

This paper aims to provide an overview of the information society shaped using the most recent technological tools by framing the opportunities and risks involved. The analysis of the issues raised by the tools requires a hermeneutical effort that seeks to combine increasingly complementary and interdependent regulations. The analysis, therefore, in addition to a necessary reference to the importance of personal data legislation, focuses on the recent legislation and its impact in the field of information and communication, with a particular focus on the phenomenon of deep fakes. The analysis is based on the Artificial Intelligence Act and the Digital Services Act, which aims to reduce or remove illegal content from digital environments.

Sommario

[*] Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Keywords

società dell'informazione – *deep fake* – servizi digitali – intelligenza artificiale – servizi intermediari – *hosting provider*

1. Introduzione

Nel corso degli ultimi due lustri, il settore dei dati personali e, in generale, quello delle nuove tecnologie, sono stati investiti da un importante attività legislativa tale da comporre un vero e proprio "diritto digitale".

Gli interventi normativi, avvenuti per lo più con lo strumento del regolamento europeo, necessitano di una proficua e complessa attività di attuazione e di coordinamento, fondamentale per evitare il rischio di un "effetto sabbie mobili" nei vari settori.

Il saggio, perciò, si concentra sull'analisi di alcune normative che rientrano nel campo dell'informazione e della comunicazione, in relazione ad ambiti dove le nuove tecnologie hanno provocato considerevoli trasformazioni, talvolta positive e, talaltra, negative[1]. Tra questi ultimi si inserisce il tema del rischio di disinformazione, il quale si distingue a sua volta dalla misinformazione e dalla malinformazione[2].

Perciò, l'analisi si sviluppa su due linee direttrici. La prima si fonda sul recente regolamento europeo sui servizi digitali (Digital Services Act - DSA), il quale origina dall'esperienza della direttiva sul commercio elettronico[3]; la seconda, invece, sull'impatto che negli ambiti succitati possono registrare i *deep fake* e, in particolare, sulla loro regolamentazione inserita sia nel DSA che nel regolamento sull'intelligenza artificiale (AIA o AI Act).

Infine, seppure con un livello minore di approfondimento, verrà messa in luce l'im-

[1] Sul tema della libertà di informazione alla luce dell'evoluzione tecnologica e in una prospettiva di carattere costituzionale, si veda A. Lauro, *Siamo tutti giornalisti? Appunti sulla libertà di informazione nell'era social*, in questa *Rivista*, 2, 2021, 141 ss. Ma anche su ciò che porterebbe alla c.d. fisica sociale, ossia al tema inerente al flusso di idee e sul come le reti sociali le diffondono e le trasformano in comportamenti attraverso i *big data*, si veda A. Pentland, *Fisica sociale. Come si propagano le buone idee*, Milano, 2015; su questi temi cfr. altresì G. Ziccardi, *La democrazia elettronica tra social network, big data e problemi di sicurezza,* in *Diritto di Internet*, 1, 2019, 239 ss.

[2] Alcuni preferiscono parlare di "disinformazione" anziché di *fake news*, perché con la prima si fa riferimento e si inglobano tutte le «informazioni false, inaccurate o fuorvianti, artificiosamente create, le quali siano presentate e diffuse con lo scopo precipuo di trarre un profitto di carattere economico, politico o ideologico e/o di provocare un danno a livello pubblico, ivi inclusa l'ingerenza nei processi elettorali e democratici». In questo senso, O. Pollicino - P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *federalismi.it*, 12, 2024, 6. Gli Autori distinguono la disinformazione dalla misinformazione; quest'ultima si caratterizzerebbe per l'elemento soggettivo, poiché non sussiste una volontà di diffusione delle informazioni false, concretizzandosi perciò nella diffusione di un materiale percepito come genuino. Il riferimento è alla ricondivisione nell'ambiente online di un contenuto falso ma che si ritiene essere vero. Ancor diversa sarebbe la malinformazione che si avrebbe con la diffusione di informazioni rispondenti al vero che vengono comunicate con l'obiettivo di «provocare un danno».

[3] Per una ricostruzione del percorso che ha portato alla adozione della direttiva sul commercio elettronico e sul suo contenuto si veda M. Bassini, *La rilettura giurisprudenziale della disciplina sulla responsabilità degli Internet service provider. Verso un modello di responsabilità "complessa"?*, in *federalismi.it*, 3, 2015, 1, spec. 7.

portanza della disciplina riguardante la protezione e il trattamento dei dati personali, profilo onnipresente e imprescindibile nei temi in discussione.
L'intero quadro deve essere calato all'interno di un ecosistema digitale in cui i mercati sono dominati da imponenti piattaforme che creano loro stesse le regole e influenzano le dinamiche tra utenti e operatori grazie al potere che hanno acquisito[4]. Si tratta delle cc.dd. *big tech*, talvolta indicate con l'acronimo di GAFAM, le quali vanno a comporre quelli che sono identificati come i "nuovi" poteri privati[5].

2. Una premessa sul fenomeno del *deep fake*

Prima di procedere con l'analisi normativa, è utile un preambolo sul recente e attuale fenomeno del *deep fake*, consistente in un contenuto che può estrinsecarsi in una immagine, video o audio generato o manipolato artificialmente.
Per rendere l'idea della portata dei *deep fake* si può far riferimento al recente episodio di truffa che ha indotto un funzionario di una multinazionale britannica con sede ad Hong Kong a trasferire 25 milioni di dollari credendo di interagire in video-conferenza con il proprio direttore finanziario, e non con un estraneo che si avvaleva proprio di una tecnica di *deep fake*[6].
L'origine del fenomeno (neologismo che deriva da una crasi tra "deep learning" e "fake")[7] viene solitamente fatto risalire al 2017, quando un utente della piattaforma Reddit pubblicò vari video in cui i volti di attrici famose venivano scambiati su video porno[8]. Nel gennaio del 2018, un altro utente della stessa piattaforma creò un programma in grado di rendere accessibile a tutti la possibilità di manipolare video[9]. Da

[4] F. Ruggeri, *Poteri privati e mercati digitali,* Roma, 2023, p. 113. L'A. ha rilevato come la capacità di alcune piattaforme di creare e imporre le regole più appropriate per il soddisfacimento delle proprie esigenze costituisce espressione dell'affermazione di specifici poteri privati, in questo caso di natura tecnologica che, nell'arco di pochi anni, sono diventati sempre più saldi e meno contendibili. Cfr. altresì S. Sileoni, *Autori delle proprie regole. I codici di condotta per il trattamento dei dati personali e il sistema delle fonti*, Milano, 2011, p. 9 che individua nel cambiamento del ruolo degli attori pubblici le cause dell'emersione di una sempre maggiore tendenza all'autoregolamentazione da parte dei privati; P. Bonini, *L'autoregolamentazione dei principali Social Network. Una prima ricognizione delle regole sui contenuti politici*, in *federalismi.it*, 11, 2020, p. 265.

[5] L'acronimo sta ad indicare le cinque piattaforme (occidentali) più influenti, ossia, Google, Amazon, Facebook (ormai, Meta), Apple e Microsoft.

[6] Il Sole 24 Ore, 8 febbraio 2024, 21.

[7] In particolare, si tratta della rete generativa avversaria (*Generative Adversarial Network* – GAN), costituita da «un paio di reti neurali di deep learning "avversarie". Il primo network, il falsificatore, cerca di generare qualcosa che sembri reale, per esempio l'immagine sintetizzata di un cane, sulla base di milioni di immagini di cani. L'altro network, l'investigatore, paragona l'immagine sintetizzata del cane creata dal falsificatore con delle effettive immagini di cani, e determina se l'output del falsificatore è reale o fasullo». In tal senso, K. Lee - C. Qiufan, *AI 2041*, Roma, 2023, 79.

[8] E. Meskys - A. Liaudanskas - J. Kalpokiene - P. Jurcys, *Regulating deep fakes: legal and ethical considerations*, in *Journal of Intellectual Property Law & Practice*, 15, 2020, 24 ss., spec. 26. Gli AA. specificano però che nella letteratura accademica il lavoro che maggiormente si avvicina al fenomeno del *deep fake* risalga al 2016 con un articolo di Thies presentato alla *Conference on Computer Vision and Pattern Recognition* in cui fu ideato un modo per consentire a una persona (sorgente) di controllare le espressioni facciali di altro soggetto (target) all'interno di un video.

[9] *Ibid.*

lì si sono venuti a moltiplicare i *software* che consentono di generare *deep fake*. Tuttavia, i più recenti programmi sono più efficienti sotto un profilo di dati e calcoli, sino al punto da essere in grado di "rianimare" immagini fisse[10].

In letteratura molti sostengono che i due elementi che definiscono i *deep fake* vanno rintracciati (i) nell'uso di tecnologie basate sull'intelligenza artificiale e (ii) nell'intento di ingannare[11].

Le tecniche odierne più diffuse, per video e immagini, consistono nella (i) manipolazione di attributi facciali (*face attribute manipulation*), attraverso cui vengono alterate alcune caratteristiche del volto di una persona ritratta, come ad esempio, l'invecchiamento o il ringiovanimento del viso; (ii) tecniche di scambio volti (*face swap*), attraverso le quali il volto presente nell'immagine o video originale viene sostituito da un altro; (iii) le tecniche di *face reenactment* e *lip synching* con cui un video viene sostanzialmente manipolato facendo in modo che sembri che la persona ritratta compia determinate azioni o renda determinate dichiarazioni[12].

In ogni caso, per un inquadramento definitorio del fenomeno del *deep fake*, stando a una accezione ristretta, bisognerebbe riferirsi a quelle creazioni mediante tecniche in grado di sovrapporre le immagini del volto di una persona *target* a un video di una persona *source* al fine di generare un video in cui la prima fa o dice cose che invece fa la seconda[13]. Seguendo invece una definizione più ampia, i *deep fake* sarebbero quei contenuti sintetizzati da sistemi di IA che possono rientrare anche in altre due categorie. La prima, rappresentata dai *lip-sync deep fake,* che si riferisce a video modificati per rendere i movimenti della bocca coerenti con una registrazione audio. La seconda, invece, ricomprende i *puppet-master*, i quali includono video di una persona *target* (*puppet*) che viene animata seguendo le espressioni facciali e i movimenti degli occhi e della testa di un'altra persona (*master*) seduta di fronte a una telecamera[14].

La tecnica, oltre ad aver sviluppato *software* per la creazione di *deep fake*, ha generato anche sistemi e metodi per il loro rilevamento[15]. Tuttavia, esistono a loro volta anche tecniche in grado di eludere i metodi di rilevamento esistenti, evidenziando, quindi, tutti i limiti e l'assenza di "robustezza" degli attuali approcci di rilevamento dei *deep fake*, nonché suggerendo l'opportunità di individuare metodi che raggiungono una migliore efficacia e resilienza a fronte dell'evoluzione della tecnica[16].

---

[10] Ivi, 27.

[11] A. Fernandez, *Regulating Deep Fakes in the Proposed AI Act*, in *medialaws.eu*, 23 marzo 2022.

[12] O. Pollicino - P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, cit., 11 ss.

[13] Aa. Vv., *Deep learning for deepfakes creation and detection: A survey*, in *Computer Vision and Image Understanding*, 223, 2022, 103525.

[14] *Ibid.*

[15] F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in *International Review of Law, Computers & Technology*, 2024, 1 ss., spec., 4. L'A. annovera vari programmi sviluppati con questa finalità, tra cui Sensity, che riconosce contenuti manipolati dall'IA e tecniche di sintesi come volti creati dall'IA e scambi di volti in video realistici.

[16] W. Alkishri - S. Widyarto - J. H. Yousif, *Detecting Deepfake Face Manipulation Using a Hybrid Approach of Convolutional Neural Networks and Generative Adversarial Networks with Frequency Domain Fingerprint Removal*, consultabile su *ssrn.com,* 2023, 1 ss., spec. 16.

Non manca poi chi sostiene che le tecnologie di rilevamento dovrebbero svilupparsi come modelli *open source,* arrivando così a uno standard comune condiviso in grado di gestire il fenomeno[17]. La tecnologia della *blockchain*, per altro verso, viene considerata un altro valido strumento per "controllare" il fenomeno oltre a garantire l'autenticità di un'opera[18].

Alcuni, per analizzare il fenomeno da un punto di vista etico e normativo, distinguono quattro categorie principali di *deep fake*, a seconda del loro uso. Le prime due (*revenge porn* e *deep fake* politici), vengono definiti come casi "difficili", mentre i *deep fake* creati per contenuti commerciali o creativi sono socialmente utili e quindi sollevano meno preoccupazioni[19].

Dunque, l'utilizzo di questa tecnologia può provocare effetti negativi o positivi[20]. Tra i primi, ci può essere l'erosione della fiducia delle persone nei confronti dei contenuti mediatici o un aumento della disinformazione, l'incitamento all'odio e persino una sollecitazione di tensioni politiche[21]. Essi, però, possono avere anche un impatto creativo o produttivo nella fotografia, nei videogiochi, nella realtà virtuale, nelle produzioni cinematografiche e nell'intrattenimento, ad esempio nel doppiaggio di film stranieri, nel campo dell'istruzione attraverso la rianimazione di personaggi storici o nel provare indumenti virtualmente mentre si fanno acquisti[22].

Nel suo insieme, come anticipato, il fenomeno è stato disciplinato sia nel Regolamento europeo sui servizi digitali (DSA), sia nell'AI Act[23]. Nei successivi paragrafi, perciò, verrà dapprima analizzato l'impianto normativo del DSA e poi le previsioni del regolamento europeo sull'intelligenza artificiale sul tema.

---

17 E. Meskys - A. Liaudanskas - J. Kalpokiene - P. Jurcys, *Regulating deep fakes*, cit., 30.

18 L. Floridi, *Artificial Intelligence*, *Deepfakes and a Future of Ectypes*, in *Philos. Technol*, 31, 2018, 317, spec. 321. L'A. sottolinea come «*As a secure and distributed register of transactions, blockchain is being explored as a means of reliably certifying the origins and history of particular products: whether in terms of securing food supply chains, or in recording the many linked acts of creation and ownership that define the provenance of an artwork. In the future, we may adopt the same solution wherever there is a need to ensure (or establish) the originality and authenticity of some artefact, be it a written document, a photo, a video or a painting*».

19 Ivi, 28.

20 Una descrizione chiara del fenomeno, dei suoi effetti negativi e positivi, può essere rintracciata nel saggio di M. Westerlund, *The Emergence of Deepfake Technology: A Review*, in *Technology Innovation Management Review*, 2019, 39 ss.

21 A. Thanh Thi Nguyen - B. Quoc Viet Hung Nguyen - A. Dung Tien Nguyen - A. Duc Thanh Nguyen - C. Thien Huynh-The - D. Saeid Nahavandi - E.Thanh Tam Nguyen - F. Quoc-Viet Pham - Cuong M. Nguyen, *Deep learning for deepfakes creation and detection: A survey,* in *Computer Vision and Image Understanding*, 223, 2022, 103526. Il *deep fake,* secondo gli autori, può anche essere usato per generare false immagini satellitari della Terra contenenti oggetti che non esistono realmente per confondere gli analisti militari; ad esempio, creando un falso ponte su un fiume anche se in realtà non esiste.

22 *Ibid.*

23 V. M. Veake - F. Z. Borgesius, *Demystifying the draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 4, 2021, 97 ss., spec., 108, per una critica al testo originario dell'AI Act, proposto dalla Commissione europea, in ordine alla disciplina dedicata al *deepfake*.

3. Il nuovo regolamento europeo sui servizi digitali (regolamento (UE) 2022/2065) che sostituisce il previgente quadro della direttiva sul commercio elettronico

Il regolamento europeo sui servizi digitali (regolamento (UE) 2022/2065), meglio conosciuto come Digital Services Act, è entrato in vigore il 16 novembre 2022 e la sua integrale applicazione, però, si è avuta a partire dal 17 febbraio 2024[24]. Con questo intervento normativo è stato parzialmente modificato l'approccio seguito con la direttiva 2000/31/CE dedicata al commercio elettronico e recepita in Italia con il d.lgs. 70 del 2003[25].

L'obiettivo dichiarato e perseguito con il DSA è quello di contribuire al corretto funzionamento del mercato interno dei servizi intermediari stabilendo le norme per un ambiente online sicuro, prevedibile e affidabile che faciliti l'innovazione e in cui siano tutelati i diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea.

Il regolamento in questione nasce da alcune esigenze del web, il quale ha generato una disintermediazione digitale tale da condurre a una marginalità degli operatori professionali nel settore dell'informazione, con una amplificazione di informazioni su piattaforme digitali, *social network* e blog[26]. Questo scenario, secondo alcuni, favorirebbe la propagazione delle cc.dd. *fake news* e renderebbe arduo orientarsi a causa della difficoltà nell'individuazione delle fonti affidabili[27]. Il lato positivo, però, è quello di avere un ambiente pluralistico che, talvolta, consente la diffusione di notizie trascurate dai circuiti *mainstream*. Rintracciare un punto di equilibrio costituisce l'operazione più difficile da compiere.

Quindi, il DSA prevede un primo gruppo di norme che si ispira e che riprende il contenuto delle regole della direttiva sul commercio elettronico, le quali definiscono il perimetro delle esenzioni da responsabilità dei prestatori di servizi intermediari e, un secondo gruppo, che sancisce gli obblighi che fanno capo a questi ultimi. Le norme volte a individuare gli obblighi a carico dei fornitori mutano l'impianto normativo di riferimento derivante dalla direttiva sul commercio elettronico[28].

---

[24] Per una analisi giuridica, etica e sociale del DSA, anche in un'ottica comparativa con la direttiva sul commercio elettronico, si veda F. Wilman - S. L. KalÄda - P.J. Loewenthal, *The EU Digital Services Act*, in *Oxford Academic*, 2024; A. Turillazzi - M. Taddeo - L. Floridi - F. Casolari, *The digital sevices act: an analysis of its ethical, legal, and social implications*, in *Law Innovation and Technology*, 15, 1, 2023, 83 ss. Sul DSA e sul possibile "effetto Bruxelles" da una prospettiva statunitense, si veda A. Chander, *When the Digital Services Act Goes Global*, in *Berkeley Technology Law Journal*, 38, 3, 2023, 1067 ss.

[25] Su questo tema si veda G. Finocchiaro, *Responsabilità delle piattaforme e tutela dei consumatori*, in *Giornale di diritto amministrativo*, 6, 2023, 730; G. Monga, *Responsabilità degli intermediari. Il Digital Services Act*, in M. Maggiore (a cura di), *Il commercio elettronico*, Torino, 2024, 194 ss.

[26] B. Grazzini, *Piattaforme e content moderation - Fake news e disinformazione*, in *Giurisprudenza Italiana*, 2, 2024, 491, spec. 493.

[27] *Ibid.* Sul tema della disinformazione e sulla diffusione di notizie tra utenti si veda anche M. Del Vicario - A. Bessi - F. Zollo - W. Quattrociocchi, *The spreading of misinformation online*, in *PNAS*, 13, 3, 2016, 554 ss.

[28] G. Finocchiaro, *Responsabilità delle piattaforme e tutela dei consumatori*, cit., 733.

Il regolamento, quindi, si incentra sui prestatori di determinati servizi della società dell'informazione così come definiti dalla direttiva (UE) 2015/1535; vale a dire, coloro che prestano qualsiasi servizio, normalmente dietro retribuzione, a distanza, per via elettronica e a richiesta individuale di un destinatario[29].

In altri termini, la normativa riguarda molti degli operatori del mercato digitale, dai *social network* alle piattaforme *e-commerce,* fino ai motori di ricerca.

Lo scopo principale del regolamento è quello di impedire, o quantomeno di ridurre, la diffusione di contenuti illegali nell'ambiente online, dettando le regole che definiscono i casi in cui i prestatori di servizi intermediari non sono responsabili e i relativi obblighi da rispettare.

Da una analisi complessiva della normativa in questione, ciò che emerge come fulcro centrale è rappresentato dal potere del prestatore di servizi intermediari di procedere con l'adozione diretta di una misura restrittiva in caso di contenuti illegali, senza la necessità di un previo provvedimento di un'autorità. Ciò è in linea con la giurisprudenza europea che, anche in tempi recenti, ha stabilito che un motore di ricerca può dar seguito alla richiesta di deindicizzazione se il richiedente riesce a fornire un *fumus* di prova della "manifesta inesattezza" delle notizie indicizzate, senza necessità di una precedente pronuncia di una autorità giudiziaria[30].

Il modello di responsabilità dei fornitori di servizi digitali muta rispetto a quello della direttiva 2000/31/CE poiché in quest'ultima venivano collocati in una posizione privilegiata, con una responsabilità limitata (regime del *safe harbour*), mentre con il DSA è l'utente che costituisce il fulcro della tutela[31].

---

[29] Il riferimento normativo all'elemento del «normalmente dietro retribuzione» può essere considerato uno degli elementi che pone l'esigenza di comporre in modo univoco e definitivo la questione attinente alla fallace gratuità di alcuni servizi di operatori digitali, tra cui i *social network*, per la quale sussiste un importante riluttanza nel considerare il trattamento dei dati personali dell'utente come un corrispettivo o una controprestazione. Se tali servizi venissero considerati come servizi "gratuiti", si potrebbero avere problematiche di applicazione soggettiva anche del DSA là dove richiede – tramite un rinvio alla direttiva (UE) 2015/1535 – un servizio reso «normalmente dietro retribuzione». Sul tema della gratuità sia consentito il rinvio a G. Proietti, *Algoritmi e interesse del titolare del trattamento nella circolazione dei dati personali*, in *Contratto e Impresa,* 3, 2022, 880, spec. 897.

[30] CGUE, C-460/20, *Google LLC* (2022); sul diritto/obbligo alla deindicizzazione la giurisprudenza di legittimità è intervenuta a più riprese, anche recentemente. Si veda, infatti, Cass. civ., sez. I, 27 dicembre 2023, n. 36021, in *Foro It.*, 2024, 1, 2, 455. Nel sistema previgente, ossia con la direttiva sul commercio elettronico, il legislatore nazionale aveva optato per un diverso sistema, discostandosi dal regime comunitario per i servizi di *hosting*, ossia l'art. 16 del d.lgs. 70/2003 subordinava l'obbligo di intervento del fornitore alla previa notifica da parte di un'autorità competente.

[31] G. Finocchiaro, *Responsabilità delle piattaforme e tutela dei consumatori*, cit., 733-734, secondo cui «il consumatore, sul web, non è solo un fruitore dei servizi digitali, ma è un *prosumer*, ossia un consumatore e un produttore che, fruendo dei servizi digitali, contribuisce alla produzione di tali servizi. I motori di ricerca, il commercio elettronico, i blog e i social network basano il proprio funzionamento anche sulla collaborazione dell'utente-consumatore, che, mentre naviga e vive la sua onlife, contribuisce a determinare il prezzo delle inserzioni pubblicitarie o a costruire la reputazione di un venditore o di un prodotto». Un ulteriore elemento di complessità sarebbe costituito dal fatto che esiste un noto «livello di asimmetria tecnologica e informativa tra gli utenti e gli operatori del web. Non si tratta soltanto di un disequilibrio di natura economica, ma soprattutto di un disequilibrio causato da una disparità di conoscenze tecniche e di informazione. Tale asimmetria può influire sulla corretta formazione della volontà, anche contrattuale, dell'utente. Considerato che, al momento della conclusione di un contratto on line, generalmente esiste una notevole differenza fra le conoscenze delle parti contraenti, tale information gap può generare erronee aspettative o un illegittimo affidamento nei confronti del

Per rendere più efficace l'applicazione del DSA è stato poi istituito il Centro europeo per la trasparenza algoritmica (ECAT), al fine di vigilare, in particolare, sull'utilizzo dei sistemi algoritmici.
L'ECAT è chiamato a coadiuvare la Commissione europea per garantire che i sistemi algoritmici utilizzati dalle piattaforme e dai motori di ricerca di grandi dimensioni rispettino i requisiti in tema di gestione e di attenuazione dei rischi.
Un altro elemento centrale è il concetto di «contenuto illegale» che, stando al DSA, deve rispecchiare quello corrispondente all'applicazione delle norme nell'ambiente offline. Questo concetto è definito in senso lato, in modo da coprire anche le informazioni riguardanti i contenuti, i prodotti, i servizi e le attività illegali.
Con "contenuto illegale" deve intendersi il riferimento a informazioni, indipendentemente dalla loro forma che, ai sensi del diritto applicabile, sono da considerarsi illegali, come l'illecito incitamento all'odio o i contenuti terroristici illegali e i contenuti discriminatori, o che «le norme applicabili rendono illegali in considerazione del fatto che riguardano attività illegali»[32]. Tra queste figurano, a titolo esemplificativo, «la condivisione di immagini che ritraggono abusi sessuali su minori, la condivisione non consensuale illegale di immagini private, il *cyberstalking* (pedinamento informatico), la vendita di prodotti non conformi o contraffatti, la vendita di prodotti o la prestazione di servizi in violazione della normativa sulla tutela dei consumatori, l'utilizzo non autorizzato di materiale protetto dal diritto d'autore, l'offerta illegale di servizi ricettivi o la vendita illegale di animali vivi. Per contro, un video di un testimone oculare di un potenziale reato non dovrebbe essere considerato un contenuto illegale per il solo motivo di mostrare un atto illecito quando la registrazione o la diffusione di tale video al pubblico non è illegale ai sensi del diritto nazionale o dell'Unione»[33].
I prestatori di servizi intermediari, secondo il DSA, hanno la facoltà di svolgere indagini proprie per scovare i contenuti illegali, ma non sono tenuti a sorvegliare le informazioni e i contenuti che trasmettono o memorizzano, né sono tenuti ad accertare i fatti o le circostanze che inducono a ritenere sussistente la presenza di attività illegali. Quest'ultimo profilo non è nuovo, poiché già sotto la vigenza della direttiva sul commercio elettronico si dibatteva, escludendolo, circa l'esistenza di un generale obbligo di sorveglianza da parte degli operatori digitali. L'esclusione di una simile impostazione si giustificava per il fatto che il fornitore diverrebbe un vero e proprio censore privato, oltre al fatto che un obbligo del genere andrebbe a minare le fondamenta della libertà

fornitore del servizio fino al limite a giungere a viziare il momento di formazione della volontà»; sul regime previgente e relativo alla direttiva sul commercio elettronico si veda E. Andreola, *Profili di responsabilità civile del motore di ricerca*, in *NGCC*, 2, 2012, 127.

[32] In tal senso il considerando n. 12 del DSA. Sul tema riguardante i «contenuti illegali» ci si chiede «quando le fake news costituiscono espressione di libertà di manifestazione del pensiero esercitata in modo non conforme all'ordinamento (ma, ancor prima, cosa debba intendersi per fake news) ed in quali (non sempre sovrapponibili) casi esse possono venire inibite senza che si entri in frizione con le regole ed i principi fondamentali, di livello costituzionale». In questo senso, B. Grazzini, *Piattaforme e content moderation*, cit., 496. Sulla definizione di contenuto illegale ai sensi del DSA si veda anche G. Monga, *Responsabilità degli intermediari. Il Digital Services Act*, cit., 194, il quale sottolinea il carattere generale e onnicomprensivo della nozione che include ogni violazione di legge o del diritto europeo, «a prescindere da quale sia il diritto o la norma di legge concretamente violata».

[33] In tal senso sempre il considerando n. 12 del DSA.

di impresa in capo agli *Internet service provider*[34].

La giurisprudenza europea si è espressa sul tema proponendo soluzioni differenti a seconda della possibilità di qualifica del fornitore come "neutro", ossia quando realizza una attività prettamente di carattere passivo e tecnico[35], oppure, invece, quando è "attivo" e può essere quindi ritenuto responsabile[36]. Su questo tema si è espressa anche la giurisprudenza di merito, sino a recenti pronunce in cui, nonostante il fornitore fosse considerato come un *host provider* passivo, il Tribunale lo ha ritenuto responsabile per i contenuti pubblicati da terzi[37]. Si può notare poi che, sia nel contesto della giurispru-

[34] M. Bassini, *La rilettura giurisprudenziale della disciplina sulla responsabilità degli Internet service provider*, cit., 11.

[35] Una pronuncia rilevante in questo senso è quella della Corte di giustizia, CGUE, C-236/08, C-238/08, *Google France SARL* (2010), in *Foro It.*, 4, 2010, 458. Si considera "neutro", e quindi non responsabile secondo la allora vigente normativa sul commercio elettronico, quel soggetto che esegue un'attività di tipo puramente tecnico e passivo, senza alcun obbligo relativo a un controllo delle informazioni trasmesse o che va a memorizzare. In particolare, si legge che «L'art. 14 della Direttiva n. 2000/31/CE sul commercio elettronico deve essere interpretato nel senso che la norma ivi contenuta si applica al prestatore di un servizio di posizionamento su Internet qualora detto prestatore non abbia svolto un ruolo attivo atto a conferirgli la conoscenza o il controllo dei dati memorizzati. Se non ha svolto un siffatto ruolo, detto prestatore non può essere ritenuto responsabile per i dati che egli ha memorizzato su richiesta di un inserzionista, salvo che, essendo venuto a conoscenza della natura illecita di tali dati o di attività di tale inserzionista, egli abbia omesso di prontamente rimuovere tali dati o disabilitare l'accesso agli stessi». Per un commento su questa sentenza si veda M. Tavella - S. Bonavita, *La Corte di Giustizia sul caso "AdWords": tra normativa marchi e commercio elettronico*, in *Riv. dir. ind.*, 5, 2010, 429.

[36] Per i casi in cui la Corte di giustizia ha riconosciuto la qualifica di prestatore "attivo", si veda la sentenza CGUE, C-324/09, *eBay c. L'Oréal* (2011), in *Dir. giust.*, dove chi si trovava a gestire un mercato online veniva convenuto in giudizio perché aveva consentito ai propri utenti di vendere prodotti contraffatti oppure privi dei requisiti di legge necessari per la vendita. Si veda altresì CGUE, C-523/10, *Wintersteiger AG* (2012).

[37] Il riferimento è alla sentenza del Trib. Roma con il quale è stato condannato Facebook (oggi, Meta) in quanto «Sebbene l'hosting provider c.d. "passivo" non possa essere soggetto ad un obbligo generale di sorveglianza, va affermata la responsabilità della società che gestisce un social network ove venga messa a conoscenza, da parte del titolare dei diritti lesi, del contenuto illecito dei contenuti pubblicati dagli utenti su un profilo telematico ove non si sia attivata per rimuoverli o impedire l'accesso agli stessi». In tal senso, Trib. Roma, sez. spec. in materia di imprese, 15 febbraio 2019, n. 3512, nota di B. Tassone, in *Riv. dir. ind.*, 4, 2019, 372. Lo stesso Trib. Roma si era pronunciato nel senso che «ai fini dell'affermazione della responsabilità dell'hosting provider "attivo" occorre in ogni caso dimostrare che questi fosse a conoscenza o potesse essere a conoscenza dell'illecito commesso dall'utente mediante l'immissione sul portale del materiale audiovisivo in violazione dei diritti di sfruttamento economico detenuti dal titolare dei diritti lesi. Ciò in quanto anche all'hosting provider "attivo" si applica il divieto, previsto dall'art. 15 della direttiva 2000/31/CE (e dall'art. 17 del decreto attuativo n. 70/2003), di un obbligo generalizzato di sorveglianza preventiva sul materiale trasmesso o memorizzato e di ricerca attiva di fatti o circostanze che indichino la presenza di attività illecite da parte degli utenti del servizio. Correlativamente, neppure può essere esclusa la responsabilità dell'*hosting provider* "passivo" ogniqualvolta sia stato messo a conoscenza, da parte del titolare dei diritti lesi, del contenuto illecito delle trasmissioni e ciononostante non si sia attivato prontamente per rimuovere le stesse e abbia proseguito, invece, nel fornire agli utenti gli strumenti per la prosecuzione della condotta illecita». In quest'ultimo senso, Trib. Roma, Sez. spec. in materia di imprese, 10 gennaio 2019, n. 693, nota di M. Iaselli, *Riv. dir. ind.*, 4, 2019, 387. Per pronunce meno recenti sul tema si veda altresì Trib. Roma, 16dicembre 2009, nota di G. Schiavone, in *Obbl. e Contr.*, 4, 2010, 304; App. Milano, Sez. spec. in materia di imprese, 7 gennaio 2015, n. 29, nota di E. Marvasi, in *Riv. dir. ind.*, 5, 2015, 455. Con quest'ultima sentenza la Corte di appello ha rilevato che la memorizzazione permanente di contenuti da parte di un *hosting provider*, benché arricchita da servizi ulteriori non comporta una automatica qualificazione del soggetto come "attivo" tale da portare a una esclusione dei casi di esenzione da responsabilità. Per la giurisprudenza di legittimità si veda Cass. civ., sez. I, 19 marzo 2019, n. 7708, in *quot. giur.*, 2019; Cass. civ., sez. I, 19 marzo 2019, n. 7709, in *Foro It.*,

denza europea che in quella nazionale, le pronunce sorgono per lo più su fattispecie relative alla violazione del diritto d'autore. Rispetto ai primi anni duemila, ossia quando prendeva forma la normativa inerente al commercio elettronico, si è andata via via sfocando quella figura dell'*hosting provider* neutro (o passivo), che faceva leva sull'art. 14 della Direttiva E-Commerce[38].

4. *Segue*: Gli obblighi previsti nel DSA

Gli orientamenti della giurisprudenza della corte di giustizia sono stati in parte assorbiti dal legislatore europeo con la Direttiva Copyright (direttiva (UE) 2019/790) e poi, dal canto suo, il Digital Services Act ha prodotto ulteriori novità.
In un confronto con la direttiva sul commercio elettronico, gli elementi di novità offerti dal DSA riguardano, in particolare, la disciplina applicabile all'*hosting provider*. Quest'ultimo regolamento, benché comporti l'abrogazione delle disposizioni "centrali" della Direttiva E-Commerce (artt. 12-15), ne riproduce il contenuto con qualche opportuna modifica, rimettendo a una valutazione caso per caso l'inapplicabilità delle esenzioni a quelle ipotesi in cui il *provider* non si limita a una fornitura "neutra" dei servizi[39].
In buona sostanza, il DSA prevede un approccio *a strati*, vale a dire doveri di diligenza differenti a seconda dei soggetti coinvolti[40]. Infatti, ci sono obblighi applicabili indistintamente a tutti i prestatori, tra i quali quello di specificare, nelle condizioni generali di contratto, in modo conciso, intellegibile e accessibile, le informazioni riguardanti le restrizioni che vengono imposte sull'uso dei loro servizi, tra cui le politiche, le procedure, le misure e gli strumenti utilizzati ai fini della moderazione dei contenuti, incluso il processo decisionale algoritmico e la verifica umana, oltre alle regole procedurali del loro sistema interno per la gestione dei reclami.
I prestatori sono tenuti ad agire in modo diligente, obiettivo e proporzionato, tenendo conto dei diritti e degli interessi di tutte le parti coinvolte, tra cui la libertà di espressione, il pluralismo dei media e altri diritti e libertà sanciti dalla Carta[41].
Nel caso in cui il prestatore di servizi di memorizzazione di informazioni adotti una misura restrittiva, questa deve essere accompagnata da un'idonea motivazione contenente una serie di informazioni, salvo che la misura sia la conseguenza di un ordine da parte di una Autorità.
Perciò, il DSA non prevede una disciplina focalizzata sulla individuazione di ciò che online costituirebbe un contenuto illegale, benché tenti di delinearlo indirettamente,

1, 2019, 2045.

38 Sul tema, O. Pollicino, *Tutela del pluralismo nell'era digitale: ruolo e responsabilità degli Internet service provider,* in *Percorsi Costituzionali*, 1, 2014, 46 ss.

39 In questo senso depone il considerando 18 del DSA.

40 G. Monga, *Responsabilità degli intermediari. Il Digital Services Act*, cit., 214.

41 Infatti, in dottrina è stato evidenziato che «*one of the main objectives pursued by the DSA is the need to strike a fair balance between various fundamental rights and other interests at stake in the context of the provision of intermediary services*», F. Wilman, *The EU Digital Services Act*, cit., 16; si veda altresì P. Church - C. Necati Pehlivan, *The Digital Services Act (DSA): A New Era for Online Harms and Intermediary Liability*, in *Global Privacy Law Review,* 4, 1, 2023, 53 ss.

ma il fulcro del sistema si concentra sulle politiche aziendali dei prestatori e nelle condizioni contrattuali da loro predisposte. Come evidenziato in dottrina, quindi, si tratta di un sistema che consegna il governo della moderazione dei contenuti all'autonomia privata[42]. Sul piano della correttezza contrattuale di un rifiuto o rimozione di contenuti pubblicati entrano in gioco differenti interessi dei vari attori coinvolti. Da un lato, occorre considerare la libertà della piattaforma di scegliere quali contenuti rifiutare e, dall'altro, quella dell'utente che, invece, fa valere la lesione della propria libertà di parola oppure l'illegittimità di clausole "limitanti"[43].

Ebbene, il bilanciamento tra i diritti fondamentali che sono in gioco e di cui i prestatori sono sostanzialmente "arbitri" è un tema annoso e delicato, oggetto di dibattito dottrinale e di orientamenti giurisprudenziali, sia nel contesto domestico che europeo[44].

Il DSA prevede poi obblighi (e poteri) aggiuntivi – rispetto a quelli sopra elencati - per i fornitori di "piattaforme online".

Con piattaforme online il legislatore europeo intende una sottocategoria rispetto ai prestatori di servizi di memorizzazione. Si intendono, infatti, le piattaforme di *social network* o quelle "piattaforme che consentono ai consumatori di concludere contratti a distanza con operatori commerciali" (quindi, quelle piattaforme che operano nell'*e-commerce*). Esse sono classificate come prestatori di servizi di memorizzazione di informazioni che, non solo memorizzano informazioni fornite dai destinatari del servizio su richiesta di questi ultimi, ma le diffondono al pubblico su richiesta dei destinatari.

Il DSA prevede poi la discutibile figura del "segnalatore attendibile". Si tratta di una qualifica riconosciuta, su richiesta di qualunque ente, dal Coordinatore in cui è stabilito il richiedente che, in caso di segnalazioni sulla illegalità di contenuti online, avrebbe priorità rispetto agli altri segnalatori[45]. È necessario, per avere tale riconoscimento, che siano dimostrate capacità e competenze particolari per l'individuazione, l'identificazione e la notifica di contenuti illegali, una indipendenza rispetto a qualsiasi fornitore di piattaforme online e capacità di svolgimento dell'attività in modo diligente, accurato e obiettivo.

Inoltre, il DSA prevede obblighi supplementari per i fornitori di piattaforme online

---

[42] U. Ruffolo, *Piattaforme e content moderation - Piattaforme e content moderation negoziale*, in *Giurisprudenza Italiana,* 2, 2024, 442. L'A. inquadra poi i *provider* come i "nuovi arbitri della libertà di espressione".

[43] Ivi*,* 446.

[44] A. Spagnolo, *Bilanciamento tra libertà d'espressione su internet e tutela del diritto d'autore nella giurisprudenza recente della Corte europea dei diritti umani*, in *federalismi.it*, 2013, 1; M.D. Birnhack, *Acknowledging the Conflict between Copyright Law and Freedom of Expression under the Human Rights Act*, in *Tel Aviv University Law Faculty Papers*, 2008, 1. Uno degli interventi più importanti in materia è indubbiamente quello della Corte di giustizia nel celebre caso *Google Spain* con il quale è stato precisato, *inter alia*, l'obbligo a carico dei gestori di motori di ricerca di deindicizzare, dietro apposita richiesta, i contenuti pubblicati sul web senza una precisazione circa i criteri su cui si fonda la decisione: CGUE, C-131/12, *Google Spain SL c. AEPD, González* (2014). Con la sentenza *Google Spain* è stato peraltro ritenuto che il motore di ricerca riveste la qualifica di titolare del trattamento dei dati personali.

[45] In Italia, il Coordinatore dei servizi digitali è stato designato con il d.l. 123/2023 convertito con modificazioni dalla L. 159/2023, che lo ha affidato all'Autorità per le garanzie nelle comunicazioni (AGCOM). Con la delibera n. 40/2024, l'AGCOM ha avviato una consultazione pubblica per acquisire osservazioni ed elementi d'informazione, da parte dei soggetti interessati, sullo schema di regolamento di procedura per il riconoscimento della qualifica di segnalatore attendibile, nonché sulle modalità operative e le aree di competenza.

(VLOP) e di motori di ricerca online di "dimensioni molto grandi" (VLOSE). È la Commissione europea che ha il compito di stabilire quali sono queste piattaforme. Questo avviene quando tali soggetti presentano un numero medio mensile di destinatari attivi del servizio nell'UE pari o superiore a quarantacinque milioni[46].
Tra gli obblighi per questi operatori di grandi dimensioni è prevista la sottoposizione a revisioni annuali e indipendenti di propria iniziativa affinché sia valutata la conformità agli obblighi previsti al capo III, agli obblighi assunti con i codici di condotta e ai protocolli di crisi. È previsto che, in caso di sistemi di raccomandazione, i fornitori devono assicurare almeno un'opzione che non preveda la profilazione di cui all'art. 4, par. 4, GDPR.
Tra le varie prescrizioni a loro carico è prevista anche la predisposizione di una relazione annuale per individuare, analizzare e valutare gli eventuali rischi sistemici derivanti dalla progettazione o dal funzionamento del loro servizio e dei suoi sistemi, compresi quelli algoritmici. Questi rischi riguardano la diffusione di contenuti illegali, eventuali effetti negativi prevedibili per l'esercizio dei diritti fondamentali, o sul dibattito civico, sui processi elettorali e sulla sicurezza pubblica, oltre a quelli relativi alla violenza di genere, alla protezione della salute pubblica e dei minori.
L'art. 40 DSA prevede poi che i fornitori di grandi dimensioni sono tenuti a fornire (al Coordinatore dei servizi digitali o alla Commissione) l'accesso ai dati necessari per monitorare e valutare la conformità al regolamento; se richiesto, devono fornire chiarimenti sulla progettazione, logica, funzionamento e sperimentazione dei loro sistemi algoritmici, compresi i loro sistemi di raccomandazione; in alcuni casi, possono essere tenuti a fornire l'accesso ai ricercatori abilitati che soddisfano alcuni specifici requisiti per condurre ricerche che contribuiscano al rilevamento, all'individuazione e alla comprensione dei rischi sistemici nell'Unione e per la valutazione dell'adeguatezza, dell'efficienza e degli impatti delle misure di attenuazione dei rischi.
È previsto inoltre che, per attenuare tali rischi sistemici, il fornitore è tenuto all'attuazione di una serie di misure. Tra queste prevede anche il ricorso a un contrassegno visibile per far sì che un elemento di una informazione (immagine, audio, video generati o manipolati) che assomigli a persone, oggetti, luoghi o altro, e che a una persona appaia falsamente autentico, sia distinguibile quando è presentato sulle loro interfacce online. Deve essere fornita una funzionalità che consenta ai destinatari del servizio di indicare tale informazione. Si tratta di quelle fattispecie in cui rientra la succitata tecnica del *deep fake* disciplinata anche nel più recente regolamento sull'intelligenza artificiale.

5. Il *deep fake* nell'*AI Act* (regolamento (UE) 2024/1689)

L'AI Act (regolamento (UE) 2024/1689) è stato pubblicato nella G.U. dell'UE il 12 luglio 2024 e, nella sua formulazione definitiva, disciplina il fenomeno del *deep fake* all'art. 50 che apre (e chiude) il capo IV dedicato agli obblighi di trasparenza per i fornitori e i

[46] La prima designazione è avvenuta il 25 aprile 2023. Le piattaforme online designate sono diciassette, ossia, Alibaba AliExpress, Amazon Store, Apple AppStore, Booking.com, Facebook, Google Play, Google Maps, Google Shopping, Instagram, LinkedIn, Pinterest, Snapchat, TikTok, Twitter, Wikipedia, YouTube, Zalando. I motori di ricerca sono solamente due, ossia Google Search e Bing.

deployer di determinati sistemi di IA, ossia le regole che valgono per i sistemi di IA non ad alto rischio.
Il Regolamento sull'IA definisce il *deep fake* come una «immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona» (art. 3, n. 60, AIA).
Sempre nell'articolo dedicato alle definizioni, con "fornitore" il Regolamento intende quel soggetto che sviluppa o fa sviluppare un sistema di IA e lo immette sul mercato con il proprio nome o marchio. Con *deployer* intende, invece, quel soggetto che utilizza un sistema di IA sotto la propria autorità, fatta eccezione di un utilizzo per attività personale non professionale.
Per i fornitori di sistemi di IA che generano audio, immagini, video o testuali sintetici è sancito l'obbligo di garantire che quanto generato sia marcato in un formato intellegibile attraverso soluzioni tecniche solide e affidabili[47]. A questa regola fa eccezione il caso in cui i sistemi di IA svolgano funzioni di assistenza per l'*editing standard* o nel caso in cui non modifichino in modo sostanziale i dati immessi dal *deployer*.
I *deployer* di un sistema di IA che manipola o genera immagini o contenuti audio o video «che costituiscono un "deep fake"» sono tenuti a rendere noto che quel contenuto è stato artificialmente generato o manipolato. Se, però, si tratta di creazioni artistiche, creative, satiriche o fittizie, l'obbligo di trasparenza è limitato e non può ostacolare «l'esposizione o il godimento dell'opera».
Anche nel caso di generazione o di manipolazione di un testo (*deep fake* testuale), finalizzato all'informazione su questioni di interesse pubblico, i *deployer* sono tenuti a rendere noto che il testo è stato artificialmente manipolato o generato, salvo il caso in cui il contenuto generato sia stato sottoposto a un processo di revisione umana o di controllo editoriale e un soggetto è il responsabile editoriale della pubblicazione.
Tutti gli obblighi in questione trovano eccezione nel caso in cui l'uso del sistema di IA sia stato autorizzato per l'accertamento, l'indagine o la prevenzione di reati.
Infine, viene incoraggiata e agevolata – per il tramite dell'ufficio per l'IA – l'elaborazione di codici di buone pratiche volte a facilitare un'attuazione efficace degli obblighi sulla rilevazione e sulla etichettatura dei contenuti artificialmente manipolati o generati. Viene fatta salva la facoltà della Commissione UE di adottare atti di esecuzione per approvare tali codici.

6. *Segue*. Il dibattito e le sfide sulla normazione del *deep fake*

Si è già aperto un ampio dibattito in letteratura sulla normazione del *deep fake*. Si discute, ad esempio, sulla categoria di rischio previste dall'AI Act entro cui dovrebbe rientrare. Alcuni sostengono che la categoria più opportuna sarebbe quella dell'alto rischio e che l'attuale disciplina dedicata al fenomeno sia in realtà inadeguata poiché non forni-

[47] Si noti che l'AIA prevede un obbligo anche in merito ai "testuali sintetici", ma questi ultimi non rientrano nella definizione di *deep fake* di cui all'art. 3, n. 60, AIA.

sce un quadro giuridico chiaro in termini di responsabilità per gli sviluppatori di queste tecnologie, ponendosi enfasi su misure preventive, piuttosto che sanzionatorie[48].
Altri li ritengono rientranti in classificazioni di rischio "limitate" o "specifiche" o, comunque, in una categoria a sé stante[49]. Altri, ancora, hanno evidenziato che qualunque categorizzazione non può soffermarsi sulla tecnologia in sé, sostanzialmente neutrale, ma sull'uso che ne viene fatto, il quale può essere finanche positivo[50]. Quindi, il contesto e l'uso sarebbero gli elementi che possono costituire un fattore chiave nella valutazione del rischio[51].
Non manca chi, invece, ha ritenuto possibile una loro sussunzione nell'ambito delle pratiche vietate di cui all'art. 5, par. 1, lett. a), AIA, poiché vi potrebbe essere uno sfruttamento dei dati dei *social media* e di sistemi di IA per generare *deep fake* di individui ignari, spesso prendendo di mira gruppi vulnerabili[52].
D'altra parte, alcuni sottolineano una discrasia e una differente classificazione di questi fenomeni nell'ambito della disinformazione elettorale nella loro disciplina dell'AIA e del DSA, i quali verrebbero assoggettati a rischi tra loro diversi[53].
Viene poi criticata la parte dell'AIA che esonera i sistemi di IA gratuiti e *open source* dai requisiti di trasparenza imposti ai modelli di IA di uso generale. Ciò perché, in questo modo, alcune piattaforme che consentono agli utenti di produrre *deep fake* di specifici individui sarebbero in grado di operare liberamente, a meno che non siano ritenute a rischio sistemico. Questa potrebbe, perciò, costituire una "scappatoia" per lo sfruttamento di tecniche per un uso dannoso, tra cui il furto di identità o campagne di disinformazione, oltre alla possibilità di ottenere vantaggi ingiusti rispetto ad altre imprese concorrenti che invece operano in modo corretto, distorcendo potenzialmente la concorrenza[54]. Tale "esonero", tuttavia incontra limiti allorché si tratti di modelli di IA per finalità generali, categoria in cui potrebbe ricadere un *deep fake*[55].

---

[48] C. Vanberghen, *The AI Act vs. deepfakes: A step forward, but is it enough?*, in *euractiv.com*, 26 febbraio 2024.

[49] F. R. Moreno, *Generative AI and deepfakes*, cit., 3.

[50] M. Labuz, *Regulating Deep Fakes in the Artificial Intelligence Act*, in *Applied Cybersecurity & Internet Governance*, 2, 1, 2023, 1 ss., spec. 11; per gli aspetti positivi che possono derivare dai *deep fake* si veda J. Silbey - W. Hartzog, *The Upside of Deep Fakes*, in *Maryland Law Review*, 78, 4, 2019, 960-966.

[51] *Ibid.*

[52] F. R. Moreno, *Generative AI and deepfakes*, cit., 7.

[53] Ivi, 16. L'A. fa riferimento al considerando n. 132 AIA sui rischi specifici derivanti da generazione di contenuti che creano rischi "specifici". Il considerando n. 120 e n. 136 del DSA che fa rientrare nel rischio sistemico la disinformazione da *deepfake*. Infine, riporta il considerando n. 62 AIA, rilevandolo come una contraddizione, poiché classifica come "ad alto rischio" quei sistemi atti a influenzare l'esito di elezioni o referendum o il comportamento di voto delle persone fisiche nell'esercizio del voto alle elezioni.

[54] Ivi, 21. Quindi, l'A. sostiene che l'attuale formulazione dell'AIA non raggiunge un giusto equilibrio tra *privacy* degli utenti, protezione dei dati, diritti di proprietà intellettuale e libertà commerciale delle società di IA, rischiando di violare il principio di proporzionalità della CEDU ai sensi dell'art. 8, par. 2, e dell'art. 10, par. 2.

[55] Il considerando n. 103 dell'AI Act prevede che i componenti di IA liberi e *open source* forniti a pagamento o altrimenti monetizzati, anche tramite la fornitura di assistenza tecnica o altri servizi, ad esempio attraverso una piattaforma software, in relazione al componente di IA, o l'utilizzo di dati personali per motivi diversi dal solo miglioramento della sicurezza, della compatibilità o dell'interoperabilità del *software*, ad eccezione delle transazioni tra microimprese, non dovrebbero beneficiare delle eccezioni previste per

Al di là del tema riguardante la classificazione del rischio, il dibattito si è generato anche sugli obblighi di trasparenza previsti nell'AIA. In altri termini, ci si chiede se tali obblighi siano in grado di tutelare effettivamente i destinatari di un processo di disinformazione. Alcuni sostengono che possono avere un ruolo nel ridurre il numero di *deep fake* in circolazione, ma non possono essere considerati un vero e proprio strumento di deterrenza[56]. Secondo un'analoga tesi, la classificazione di queste tecnologie come sistemi a rischio limitato, imponendo solo requisiti di trasparenza, senza prevedere alcuna esplicita sanzione, non realizza sufficienti incentivi per il rispetto della norma[57]. Altri ancora rilevano che le disposizioni stabilite nell'AIA, nonostante la loro accurata formulazione, non comprenderebbero salvaguardie sostanziali e dovrebbero essere previste definizioni chiare, nonché una supervisione trasparente e responsabile, oltre a solide garanzie, sia per gli utenti che per i fornitori[58].
Il dibattito generale è aperto e riguarda anche quella parte della normativa che prevede le varie deroghe alle regole generali[59], oltre ai profili che si intersecano con la disciplina a tutela dei dati personali.

7. Il *deep fake* e il delicato rapporto con la protezione dei dati personali

Tra le sfide poste dai *deep fake* rientrano anche quelle che si legano alla tutela dei dati personali. Stante la definizione di dato personale sancita nel GDPR, quando un *deep fake* ritrae un individuo reale, rientra chiaramente nell'ambito di applicazione della normativa europea del GDPR[60].
Però, ci si chiede se, in caso di *deep fake* fittizio, ossia non riproduttivo di una persona realmente esistente, i dati utilizzati per generarlo vadano qualificati come dati personali[61]. Secondo alcuni la risposta dovrebbe essere positiva se si vanno ad analizzare i dati *input*. Vale a dire che, sebbene generato casualmente, l'opera potrebbe riflettere in qualche modo i caratteri degli individui realmente utilizzati per l'addestramento del sistema e tali impronte potrebbero essere sfruttate per una loro reidentificazione[62]. Seguendo questa tesi, si dovrebbe concludere che anche gli sviluppatori che non creano

i componenti di IA liberi e *open source*. Infine, precisa che la messa a disposizione di componenti di IA tramite archivi aperti non dovrebbe, di per sé, costituire monetizzazione. Il successivo considerando n. 104 prevede che i fornitori di questi modelli le cui informazioni parametri sono messi pubblicamente a disposizione dovrebbero essere soggetti ad eccezioni, salvo che presentino un rischio sistemico.

[56] M. Labuz, *Regulating Deep Fakes*, cit., 19.

[57] A. Fernandez, *Regulating Deep Fakes in the Proposed AI Act*, cit., 1.

[58] F. R. Moreno, *Generative AI and deepfakes*, cit., 24. L'A. aggiunge che solo in questo modo l'AIA potrà regolamentare efficacemente i *deepfakes* senza diventare un'arma contro le stesse libertà che cerca di salvaguardare.

[59] M. Labuz, *Regulating Deep Fakes*, cit., 25.

[60] F. R. Moreno, *Generative AI and deepfakes*, cit., 11.

[61] M.J. van der Helm, *Harmful deepfakes and the GDPR*, *Tilburg Law School – Institute for Law, Technology and Society*, in arno.uvt.nl, 1.

[62] F. R. Moreno, *Generative AI and deepfakes*, cit., 11.

direttamente *deepfake* sarebbero sottoposti alla normativa in materia di dati personali allorché utilizzino dati personali per l'addestramento degli algoritmi. Allo stesso modo, i creatori e i distributori sono sottoposti a controlli a causa dell'utilizzo di tali dati nella creazione e condivisione di *deep fake*[63].

In merito alla base giuridica suscettibile di essere invocata in queste circostanze, si ritiene che il consenso e l'interesse legittimo siano le basi astrattamente utilizzabili. Nel caso del consenso, gli individui, sia nell'ambito del contenuto originario che in quello manipolato, devono aver acconsentito attivamente al trattamento dei dati personali e aver ricevuto informazioni comprensibili, facilmente accessibili e concise sui rischi e sui benefici derivanti dal trattamento dei dati[64].

In caso di interesse legittimo, invece, sarebbe necessaria un'attenta valutazione dei potenziali rischi per i diritti e le libertà individuali derivanti dall'uso dei *deepfake*. Pertanto, sarebbe indispensabile accertare i potenziali vantaggi del trattamento dei dati personali per i *deep fake* rispetto ai potenziali danni ai diritti individuali e alle libertà degli interessati.

Quelle opere utilizzate per scopi artistici, satirici o di fantasia (v. considerando n. 134 AIA) potrebbero rientrare nella libertà di espressione e di arte dei creatori in conformità agli artt. 11 e 13 della Carta. Tuttavia, la creazione e la diffusione di disinformazione elettorale o di materiale estorsivo o di contenuti sessuali illeciti, benché generati dall'IA, richiederebbe di dare priorità alla protezione degli interessati, conformemente agli artt. 7 e 8 della Carta europea[65].

Una delle soluzioni che vengono proposte per mitigare i rischi derivanti dall'utilizzo di queste tecnologie, a beneficio degli interessati e del trattamento dei loro dati personali, è rappresentata dall'uso dei dati sintetici[66].

Secondo questa tesi, i dati sintetici sarebbero in grado di impedire che i modelli di IA ereditino e amplifichino i pregiudizi sociali, riducendo così al minimo il rischio di *deep fake* discriminatori. In secondo luogo, rafforzerebbero la *privacy* e la sicurezza riducendo la dipendenza dalle informazioni personali e diminuendo il rischio di violazioni della *privacy* e di un loro utilizzo non autorizzato. In terzo luogo, i processi di generazione dei dati sintetici possono essere più trasparenti e più facili da spiegare, mitigando i rischi derivanti dal tipico fenomeno della *black box* che caratterizza molti sistemi di IA[67].

8. Osservazioni conclusive. Prima parte

Negli ultimi anni si sono venuti a consolidare alcuni neologismi come algocrazia[68], ca-

[63] *Ibid.*

[64] Il tema viene peraltro affrontato con una rappresentazione "estrema" ma molto efficace nella puntata della serie TV *Black Mirror* dal titolo *Joan is Awful.*

[65] F. R. Moreno, *Generative AI and deepfakes*, cit., 12.

[66] Ivi, 14.

[67] *Ibid.*

[68] L. Francalanci, *Dall'algocrazia all'algoretica: il potere degli algoritmi*, in *Italiano digitale*, XIV, 3, 2020, 97; G. Cerrina Feroni, *Intelligenza artificiale e sistemi di scoring sociale. Tra distopia e realtà*, in *Diritto dell'Informazione e*

pitalismo della sorveglianza[69], regime dell'informazione[70] o dataismo[71]. Tutte locuzioni valide che esprimono concetti più o meno moderni e che hanno un denominatore comune: lo sfruttamento massivo dei dati e delle informazioni che vengono costantemente fornite dagli individui negli ambienti digitali.

Si è, quindi, consolidato un sistema in cui i singoli individui non sono più soggetti passivi, ma trasmettitori attivi[72].

In questo panorama si innesta anche il tema della circolazione di notizie false, il quale è, per vero, un fenomeno antico e sempre esistito. Nei tempi recenti, però, si assiste a una metamorfosi dei meccanismi della loro propagazione, sia per la pervasività dei *social media*, sia per lo sviluppo di nuovi sistemi, anche tecnologici, che hanno dato vita a nuovi paradigmi, portando alla formazione della società dell'informazione che oggi conosciamo[73].

Si è visto che il DSA, in questo senso, benché riprenda buona parte dei principi già espressi nella direttiva sul commercio elettronico, si prefigge un cambiamento nella direzione di una responsabilizzazione dei prestatori di servizi intermediari al fine di

dell'Informatica, 1, 2023, 1, spec. 21-24; P. Benanti, *Oracoli. Tra algoretica e algocrazia*, Roma, 2018; A. Celotto, *Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto*, in *Analisi giuridica dell'economia*, 1, 2019, 47 ss.; M. Sciacca, *Algocrazia e sistema democratico. Alla ricerca di una mite soluzione antropocentrica*, in *Contratto e impresa,* 4, 2022, 1173.

[69] In particolare, ci si riferisce al testo di S. Zuboff, *Il capitalismo della sorveglianza*, I, Roma, 2019.

[70] Concetti che ritroviamo nella lucida ricostruzione del filosofo Byung-Chul Han, *Infocrazia*, Torino, 2021, 3-13, il quale distingue il regime dell'informazione da quello disciplinare poiché, nel primo, vengono sfruttate informazioni e dati, non corpi ed energie; ciò che conta è l'accesso alle informazioni che vengono poi utilizzate per finalità di sorveglianza psicopolitica, di controllo e di previsione dei comportamenti, declassando così gli uomini a bestie da dati e consumo. L'A. sottolinea come tutto si incentra sulla connessione e più dati vengono generati, più intensivamente si comunica e tanto più efficiente diviene la sorveglianza. Il "dataismo" fa parte del regime dell'informazione e aspira a un sapere totale per calcolare tutto ciò che è presente e sarà nel futuro; quindi, le narrazioni cedono il passo ai calcoli algoritmici. Su questo tema si veda anche B. Romano, *Civiltà dei dati libertà giuridica e violenza*, Petrocco (a cura di), Torino, 2020, 45-50, secondo cui alcuni trattamenti di dati personali attraverso algoritmi finiscono per ridurre gli atti degli uomini in dati calcolabili e «trattare i naviganti nella rete come freddi dati spersonalizzati si concretizza nel manipolarli come cose, merci, ovvero nel situarli in un processo di reificazione, che esaurisce gli esseri umani nello stesso statuto degli oggetti, privi di personalità originale, capace di concepire e realizzare dei progetti, trasformativi del mondo mediante l'attività, storica e creativa del lavoro»; cfr. altresì G. Ziccardi, *Sorveglianza elettronica, data mining e trattamento indiscriminato delle informazioni dei cittadini tra esigenze di sicurezza e diritti di libertà*, in *Ragion pratica*, 1, 2018, 29 ss.

[71] Con "dataismo" si intende quella cultura che sostiene una conoscenza interamente fondata su dati informatici e sulla loro elaborazione, ritenendo ormai superflua la creazione di tesi e ipotesi filosofiche o scientifiche.

[72] Byung-Chul Han, *Infocrazia*, cit., 23.

[73] Si pensi al recente caso X-Grok, ossia un *chat-bot* della piattaforma X che cura e gestisce le notizie di tendenza amplificandone la loro diffusione, anche se non si tratta di notizie vere ma solo di tendenza. In tempi recenti è infatti accaduto che sia stata promossa nella *homepage* di centinaia di milioni di utenti una notizia falsa riguardante il fatto che "L'Iran colpisce Tel Aviv con missili pesanti". La notizia è consultabile al sito mashable.com. Sul tema del rapporto tra informazione e algoritmi e, in particolare, sul ruolo di questi nel creare e diffondere e nel contrastare le notizie (false), si veda G. Marchetti, *Le fake news e il ruolo degli algoritmi*, in questa *Rivista*, 1, 2020, 29 ss. Su questa linea occorre considerare il rapporto tra le *fake news* e la verità che Byung-Chul Han, *Infocrazia*, cit., 33, mette in evidenza, ossia che in un sistema di infodemia, le prime hanno già esercitato il loro pieno effetto prima ancora che si sia dato avvio a un processo di verifica e sfrecciano davanti alla verità senza poter essere più raggiunte da questa.

tutelare l'utente. Facendo propri i principi elaborati dalla giurisprudenza, oggi non si ricorre più a una netta e astratta distinzione tra *provider* attivo e neutro (o passivo), rimettendo l'analisi a una valutazione caso per caso.

L'impianto normativo che si ricava dal DSA offre gli strumenti utili per l'attenuazione dei rischi derivanti dalla diffusione di contenuti illegali online e dalla loro amplificazione e, sebbene non stabilisca ciò che costituisce un contenuto illegale - per il quale bisogna rifarsi alla normativa europea e nazionale -, delinea un approccio *a strati* in cui si stabiliscono i casi in cui il prestatore di servizi sarebbe esonerato da responsabilità e quelli che sono gli obblighi ai quali deve conformarsi. Tra questi obblighi viene valorizzato il ruolo centrale delle condizioni contrattuali che l'utilizzatore del servizio deve rispettare. È in tali formulazioni negoziali, trasparenti e intellegibili che il prestatore definisce anche le sue facoltà e i suoi poteri di "moderazione" dei contenuti che possono essere limitati o rimossi[74].

Di qui entra in gioco il delicatissimo bilanciamento tra il potere della piattaforma e la libera espressione dell'utente che il DSA non intende scalfire. Dunque, il regolamento sui servizi digitali definisce un quadro generale che, nel suo dettaglio, dev'essere inquadrato dal prestatore del servizio con specifiche condizioni contrattuali. Perciò, il fornitore ricopre un ruolo di moderatore tutt'altro che semplice, il quale non può e non deve sfociare nella funzione di "Ministero della Verità".

Il tema del potere di moderazione delle piattaforme non è ovviamente circoscritto ai confini europei. Recentemente, infatti, negli Stati Uniti è intervenuta la Corte Suprema che ha rinviato alle Corti statali la questione attinente proprio ai poteri delle piattaforme online, sollecitando, *inter alia*, una revisione, in conformità al primo emendamento, di quelle leggi statali che limitano il loro margine di intervento[75].

Ad ogni modo, nell'intero quadro di bilanciamento degli interessi in gioco dovrebbe considerarsi anche il principio di autoresponsabilità dell'utente finale. In altri termini, nonostante non vi sussista alcun dubbio sulla necessità di moderare i contenuti illegali che si traducono in fattispecie di reato, il discorso si complica per la manifestazione di opinioni che si discostano da una narrazione "ufficiale" o socialmente accettata[76].

---

[74] Per una analisi critica delle condizioni generali applicate dalle principali piattaforme digitali alla luce del DSA, si veda E. Poddighe - V. Zeno-Zencovich, *La «correttezza» nelle condizioni generali di contratto delle grandi piattaforme online*, in *Comparazione e diritto civile*, 1, 2024, 1.

[75] Supreme Court of the United States, No. 22, *Moody, Attorney General of Florida, Et Al. V. Netchoice, Llc, Dba Netchoice, Et Al., No. 22, NetChoice, LLC, dba NetChoice, et al. v. Paxton*, *Attorney General of Texas,* July 1, 2024, in *supremecourt.gov.*

[76] Opera una distinzione tra "notizie false" e "opinioni" B. Grazzini, *Piattaforme e content moderation*, cit., 497-498. L'A. rileva peraltro che «da un punto di vista strettamente civilistico, l'importanza della distinzione fra notizie e opinioni si stempera nell'ottica risarcitoria, poiché anche quando non è consentito inibire un illecito è possibile responsabilizzare il suo autore. In questa prospettiva le false notizie, ma anche le opinioni esternate in modo lesivo, possono venire ricondotte a diverse fattispecie, e per tale via generare responsabilità (ma anche, in alcune ipotesi, possibilità di inibitoria) secondo i presupposti da ciascuna di queste previsti: può trattarsi di illecito aquiliano ex art. 2043 c.c. (da diffamazione, da violazione della riservatezza o di altri diritti della persona); oppure, qualora celino un atto di concorrenza sleale, integrare gli estremi dell'art. 2598 c.c. (ed in simile ipotesi è l'art. 2599 c.c. a prevedere espressamente l'inibitoria, a prescindere dall'esistenza dei requisiti soggettivi indispensabili per il risarcimento del danno); inoltre, le fake news diffuse nel contesto della comunicazione commerciale possono assumere i contorni di un illecito pubblicitario, che a sua volta può atteggiarsi ad atto concorrenzialmente sleale, ma altresì ricadere nella disciplina in materia di pratiche commerciali scorrette ai sensi degli artt. 18 e segg. del Codice del

Queste ultime, in una società pluralistica, devono essere sempre salvaguardate, benché non condivise dai più, anche in considerazione di un principio di autoresponsabilità dell'utente finale (o lettore), il quale non può essere supposto come un soggetto manipolabile.

9. Osservazioni conclusive. Seconda parte

Nella società dell'informazione che si è sommariamente descritta, una tecnica che può talvolta ingannare anche i soggetti più avveduti è indubbiamente quella del *deep fake*[77]. Un fenomeno che, per una ragione ignota, è regolata sia dal Digital Services Act, sia dall'Artificial Intelligence Act.

La regolazione di questa tecnica non è banale e dipende dall'approccio che si intende seguire. Ad esempio, Il Regno Unito ha di recente annunciato una proposta legislativa che criminalizza la creazione di *deep fake* sessualmente espliciti attraverso una nuova fattispecie di reato che si basa su reati già previsti per la condivisione di immagini intime "deepfake", già introdotti con la Online Safety Act. È un approccio normativo che si focalizza sull'utilizzo di uno strumento, e non sullo strumento in sé. In termini di approccio legislativo, il DSA sembra andare in questa direzione poiché, con l'art. 40, circoscrive l'adozione di alcune misure - a carico di alcuni fornitori (VLOSE e VLOP) – volte a mitigare i rischi sistemici, tra cui il ricorso a un contrassegno visibile per far sì che un elemento di una informazione (immagine, audio, video generati o manipolati) sia distinguibile quando è presentato sulle loro interfacce online.

Ebbene, nella normativa del DSA non viene descritto né disciplinato lo *strumento* con il quale si genera il contenuto potenzialmente manipolatorio; esso non viene circoscritto ai sistemi di IA, come quelli disciplinati nell'AI Act, ma si focalizza sull'adozione di una misura volta a rendere trasparenti alcune forme di manipolazione di video, immagine o audio (non viene menzionato un testo). In queste ipotesi rientra, senz'altro, anche la tecnica del *deep fake*.

Nella disciplina contemplata nell'AI Act, invece, l'approccio sembra essere differente. Invero, sebbene vengano prescritti obblighi di trasparenza, si deve trattare di una manipolazione (anche di un testo) che si estrinseca per il tramite di un sistema di IA, così come definito nel regolamento in questione (art. 3, par. 1, lett. a, AIA)[78]. Quindi, si tratta di una normazione che si incentra prima di tutto sullo strumento tecnologico, oltreché sul suo utilizzo.

Peraltro, la normativa delineata nell'AI Act può provocare alcuni problemi applicativi.

consumo (d.lgs. 6 settembre 2005, n. 206), alla stregua di pratica commerciale ingannevole o di pratica commerciale aggressiva».

[77] Si è sopra riportato, ad esempio, il recente caso della truffa che ha portato al trasferimento di 25 milioni di dollari da parte di un funzionario di una multinazionale britannica con sede ad Hong Kong, il quale credeva di interagire in video conferenza con il proprio direttore finanziario, e non con un estraneo che si avvaleva di una tecnica di *deep fake*; cfr. nota 18.

[78] Sulla definizione di sistema di IA alla luce del regolamento sull'intelligenza artificiale sia consentito il rinvio a G. Proietti, *Definire l'indefinibile? I sistemi di intelligenza artificiale alla ricerca di un inquadramento sistematico*, in *Contratto e impresa*, 3, 2024, 882.

Un primo problema può derivare dalla definizione di *deployer*, il quale è il soggetto che utilizza un sistema di IA sotto la propria autorità, fatta eccezione di un utilizzo per attività "personale non professionale". Ebbene, molti usi della tecnica del *deep fake*, non rientrando nell'ambito di una attività professionale, rischiano di ridurre in modo rilevante il perimetro di applicazione delle disposizioni previste all'art. 50 AIA dedicate all'utilizzatore (*deployer*). Un altro elemento che riduce il perimetro applicativo della normativa è rappresentato dalle eccezioni ivi previste che, suscettibili di una valutazione ampiamente discrezionale, possono dar vita a facili elusioni.

I *deep fake*, sebbene vi siano critiche sulla loro classificazione di rischio, ricevono una disciplina *ad hoc* che li fa rientrare nella categoria di "rischio limitato". Tuttavia, quando sono rappresentati da contenuti, video o immagini che sono il frutto di modelli di IA generativa, possono (e devono) soggiacere anche alla più articolata disciplina (e ai relativi obblighi) dedicata ai modelli di IA per finalità generali disciplinati nel capo V del regolamento sull'intelligenza artificiale.

Si può immaginare che la disciplina dell'AI Act dedicata a queste tecniche di manipolazione dei contenuti, soprattutto per il suo reale perimetro applicativo, non sarà risolutiva. Essa potrà essere di ausilio all'utente nel comprendere quand'è che si è al cospetto di un *deep fake*, e quindi di una rappresentazione artificiale, ma non ridurrà la loro diffusione che, in ogni caso, non pare essere lo scopo legislativo.

In un simile scenario, gli strumenti giuridici tradizionali sembrano essere, in molti casi, i più adatti per arginare o mitigare i rischi di alcuni utilizzi di certe tecniche rispetto a quanto tenti di innovare l'AI Act. In un classico esempio di truffa tramite *deep fake*, la normativa penale e civile consente di tutelare il danneggiato. Da un'altra prospettiva, invece, le novità apportate dal DSA sul punto, benché focalizzate solo su alcuni operatori, potrebbero essere utili per gli utilizzi che provocano disinformazione ma, di certo, non possono di per sé far presagire una soluzione definitiva alle complesse questioni che solleva il tema.

Decisioni algoritmiche e discriminazioni: lo stato dell'arte*

Dora Trombella

Abstract

Alcuni sistemi di Intelligenza Artificiale funzionano mediante dei meccanismi di *machine learning*, ossia apprendono e migliorano le proprie performance sulla base dei dati che utilizzano. La macchina algoritmica, estranea a qualsiasi valutazione di tipo etico, procede in via automatica a rielaborare le informazioni razionalizzando tutto ciò che è reale e, altresì, perpetrando i *biases* radicati nella società. Tuttavia, l'AI Act potrebbe fornire valide soluzioni con riguardo a questa problematica, in quanto, anzitutto, mira ad assicurare che i sistemi di IA immessi sul mercato europeo rispettino i diritti fondamentali e i valori dell'Unione.

Some Artificial Intelligence systems function through machine learning mechanisms, which means that they learn and improve their performance based on the data they collect. The algorithm, with no ethical evaluation, automatically proceeds to reprocess information by rationalising all that is real and, likewise, perpetuating the ingrained biases in society. However, about this issue, the AI Act could be valuable because at its core it aims to ensure that AI systems in the European market respect the fundamental rights and the values of the Union.

Sommario

Keywords

intelligenza artificiale – algoritmo – apprendimento automatico – decisioni algoritmiche – discriminazioni algoritmiche

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

1. Da informazioni distorte ad esiti pregiudizievoli nei settori "ad alto rischio"

Con la nozione di *machine learning* ci si riferisce ad una procedura automatizzata volta ad individuare delle correlazioni tra più variabili all'interno di un set di dati al fine di effettuare previsioni o stime[1].

La raccolta e l'elaborazione dei dati sono fondamentali per la realizzazione delle applicazioni di *machine learning*, in quanto la qualità delle informazioni immesse nel sistema incide notevolmente sulle sue prestazioni[2].

Infatti, nelle specifiche fasi di programmazione e di apprendimento possono introdursi nel modello delle informazioni idonee a generare un risultato discriminatorio ed escludente rispetto a individui o a gruppi di individui[3].

È necessario ricordare che il diritto antidiscriminatorio individua due tipologie di discriminazione: la discriminazione diretta che sussiste nel caso in cui per la razza o l'origine etnica, il sesso, la religione, le convinzioni personali, gli handicap, l'età, l'orientamento sessuale «una persona è trattata meno favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra in situazione analoga»; mentre la discriminazione indiretta si verifica «quando una disposizione, un criterio, una prassi, un atto, un patto o un comportamento apparentemente neutri possono mettere le persone», per le caratteristiche personali suddette, in «una situazione di particolare svantaggio rispetto ad altre persone»[4].

La definizione di discriminazione diretta mal si adatta alle forme di discriminazione algoritmica[5], in quanto, la macchina, estranea a qualsiasi giudizio morale e operando unicamente su un piano logico-matematico, eredita e riproduce automaticamente le scelte preliminari effettuate dall'essere umano che l'ha programmata. Peraltro, è altamente improbabile che la decisione algoritmica si basi sulla esclusione di caratteri che identificano la categoria protetta in via immediata, discendendo (l'esclusione), piuttosto, da caratteristiche inestricabilmente connesse al fattore di discriminazione [6].

La discriminazione indiretta, invece, che crea disparità di trattamento non intenzionali, pare, almeno da un punto di vista concettuale, più confacente al fenomeno oggetto di

---

[1] D.Lehr - P.Ohm, *Playing with the data: What legal scholars should learn about machine learning*, in *University of California Davis Law Review*, 51, 2017, 671.

[2] T. Wang - B. Li - M. Chen - S. Yu, *Machine Learning Empowered Intelligent Data Center Networking: Evolution, Challenges and Opportunities*, Singapore, 2023, 10.

[3] S. Barocas - A.D. Selbst, *Big data disparate impact*, in *California Law Review*, 104, 2016, 671 ss. indicano le cinque circostanze in cui possono verificarsi eventuali discriminazioni: nella fase di individuazione delle *class labels* per la definizione dei risultati; nella fase di addestramento dei dati; durante la selezione delle caratteristiche rilevanti per il modello; a partire dai *proxies* prescelti o dalla discriminazione intenzionale inserita dai programmatori.

[4] Il riferimento è alle seguenti disposizioni: Art. 2, Dir. 2000/43/CE sulle discriminazioni per razza o origine etnica; Art. 2, Dir. 2000/78/CE sulle discriminazioni per religione, convinzioni personali, handicap, età, tendenze sessuali; Art. 2, Dir. 2006/54/CE sulle discriminazioni di genere.

[5] D. Morondo Taramundi, *Le sfide della discriminazione algoritmica*, in *Genius, Rivista di studi giuridici sull'orientamento sessuale e l'identità di genere*, 1, 2022, 7.

[6] *Ivi*, 8.

trattazione[7].

Ebbene, tali potenziali esiti pregiudizievoli si sono, di fatto, riscontrati in ambiti particolarmente sensibili: la giustizia, la salute, il lavoro e l'accesso al credito[8].

Con riguardo al primo settore, occorre precisare che con "giustizia predittiva" si intende la possibilità per l'algoritmo di prevedere l'esito di una controversia mediante l'elaborazione di dati normativi e giurisprudenziali, previa consultazione di banche dati, raccolte di giurisprudenza e opere edite in formati accessibili al sistema[9].

Infatti, le macchine adibite a questo scopo possiedono delle capacità di memorizzazione e computazione, impensabili per l'essere umano, che consentono la formulazione immediata della decisione a seguito dell'analisi di migliaia di dati.

Sul tema è emblematica la vicenda relativa al celebre sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), ampiamente utilizzato negli USA, che quantifica mediante un algoritmo il rischio che un imputato possa delinquere nuovamente. In particolare, tale sistema valuta la probabilità di recidiva sia sulla base di quanto contenuto nel fascicolo processuale sia "alla luce" delle informazioni assunte a seguito di un test composto da 137 domande a cui l'imputato viene sottoposto e che riguardano l'età, l'attività lavorativa svolta, il grado di istruzione, i legami affettivi, l'uso di droghe, le opinioni personali e il percorso criminale.

Come noto, la Suprema Corte del Wisconsin nel 2016 si è pronunciata in merito al caso del Sig. Eric L. Loomis a cui il sistema COMPAS, nel giudizio di primo grado, aveva comminato una pena di sei anni di reclusione per ricettazione e resistenza a pubblico ufficiale[10].

Il Sig. Loomis, sosteneva che il sistema funzionasse in maniera del tutto ignota alla difesa col rischio che l'imputato potesse incorrere in valutazioni discriminatorie e, altresì, che, nel suo caso, non vi fosse stata una pronuncia personalizzata, basandosi l'algoritmo sulla rielaborazione di informazioni relative al gruppo etnico in cui l'imputato era incluso.

La Corte Suprema ha dichiarato, all'unanimità, la legittimità della procedura automatizzata che aveva condotto alla sentenza impugnata, in quanto, il sistema COMPAS, a detta della Corte, rappresenta unicamente un ausilio rispetto all'attività degli organi giudiziari che devono valutare, anche in base alle informazioni fornite dalla macchina,

---

[7] Per una trattazione dettagliata della problematica si rinvia a C. Nardocci, *Intelligenza Artificiale e discriminazioni*, in *Rivista del Gruppo di Pisa*, 3, 2021.

[8] Con riguardo all'impatto delle nuove tecnologie sul diritto costituzionale si veda S. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà* , in *BioLaw Journal*, 1, 2019, 63 ss.; C. Colapietro, *Intelligenza artificiale e discriminazioni*, in *Studi parlamentari e di politica costituzionale*, 2022, I, 9 ss; M. D'amico, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020.

[9] Approfondisce la tematica U. Ruffolo, *La machina sapiens come "avvocato generale" ed il primato del giudice umano: una proposta di interazione virtuosa* in U. Ruffolo, (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 205 ss evidenziando che la macchina potrebbe funzionare essenzialmente come *bouche de la loi*, optando la stessa per la soluzione «oggettivamente esatta in quanto rispondente al comando generale ed astratto della legge».

[10] Per una trattazione più dettagliata della vicenda: F. Lagioia - G. Sartor, *Il sistema compas: algoritmi, previsioni, iniquità,* in U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 226 ss.

il rischio di recidiva dell'imputato[11].

Inoltre, la Corte ha escluso che il sistema possa definirsi discriminatorio, in quanto l'assegnazione di un punteggio più alto agli imputati di sesso maschile è da ritenersi compatibile con le statistiche che mostrano che le donne commettono meno crimini violenti e sono meno recidive rispetto agli uomini. Per quanto riguarda l'eventuale distorsione generata dal sistema che attribuisce automaticamente agli uomini neri una valutazione più elevata rispetto ai bianchi, la Corte si è limitata ad invitare gli utilizzatori della macchina ad informarsi maggiormente sulla problematica.

Nell'ordinamento italiano, indubbiamente, sarebbe precluso un processo decisionale automatizzato in chiave predittiva siffatto sia in virtù di quanto previsto dall'art. 22, par. 1, GDPR (come si dirà meglio in seguito) ma anche a norma dell'art. 220, c. 2, del Codice di procedura penale che proibisce ogni perizia volta a stabilire il carattere o la personalità dell'imputato. Per giunta, il sistema risulterebbe lesivo del diritto all'equo processo nonché del diritto di difesa dell'imputato tutelati dalla Carta costituzionale. Tuttavia, anche in Italia, sono stati avviati diversi progetti nel settore giustizia che relegano però l'algoritmo ad un ruolo ausiliario rispetto all'operato dell'autorità giudiziaria[12].

I sistemi di IA, come si anticipava, sono, altresì, impiegati per assumere decisioni in campo sanitario al fine di predire il successo di una determinata terapia, la diffusione di una malattia o lo sviluppo di specifiche patologie o, ancora, per determinare il diritto all'accesso ai servizi sanitari.

Invero, i più moderni sistemi di intelligenza artificiale provvedono a rielaborare una moltitudine di informazioni concernenti i casi clinici pregressi e le caratteristiche dei pazienti coinvolti, nonché la sintomatologia, la diagnosi, la cura applicata restituendo automaticamente, sulla base di quanto appreso, un responso al caso sottoposto al loro esame[13].

Ad esempio, i sistemi sanitari statunitensi si affidano ad algoritmi di previsione "commerciali" al fine di aiutare i pazienti con esigenze sanitarie particolarmente complesse. È interessante il caso concernente uno di questi modelli che impiegava i costi sanitari come *proxies* per determinare il bisogno di cure. Nella specie, i pazienti neri, a parità di condizioni di salute, risultava spendessero meno rispetto ai bianchi; dunque, il sistema concludeva erroneamente che i pazienti neri fossero più sani degli altri e, pertanto,

[11] Supreme Court of Wisconsin, State of Wisconsin v. Eric L. Loomis, 13 July 2016, Case no. 2015 AP157-CR.

[12] M. Martorana, *Polizia e giustizia predittive: cosa sono e come vengono applicate in Italia*, in *Agenda Digitale*, 27 gennaio 2021; C. Morelli, *Giustizia predittiva: il progetto (concreto) della Corte d'appello di Brescia*, in *Altalex.com*, 8 aprile 2019; C Castelli, *Giustizia predittiva: i progetti in corso in Italia*, in *Agenda Digitale*, 2 agosto 2023.

[13] Sull'utilizzo dell'IA in medicina si segnalano i seguenti contributi: D. Pacini - G. Folesani, *Il ruolo dell'Intelligenza Artificiale in cardiochirurgia*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi*, medici e informatici a confronto, Torino, 2023, 43 ss; A. Zini, *Intelligenza artificiale e patologie neurologiche e cerebrovascolari*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi*, medici e informatici a confronto, Torino, 2023, 49 ss; G. Pipino, *Intelligenza artificiale in ortopedia*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi*, medici e informatici a confronto, Torino, 2023, 57 ss.

destinava loro meno risorse economiche[14].

Un'altra problematica del settore concerne la mancata inclusione nei *data set* di informazioni riguardanti individui differenti da maschi bianchi.

Tale distorsione potrebbe, infatti, reiterare i gravissimi esiti a cui, in passato, si è giunti non effettuando nessun test clinico su persone di sesso femminile nella fase di sperimentazione di un farmaco o di una terapia[15].

Della questione si è interessato il Comitato Nazionale per la Bioetica con il parere «La sperimentazione farmacologica sulle donne» approvato nella seduta plenaria del 28 novembre 2008 in cui viene evidenziato che «sebbene le donne siano le maggiori consumatrici di farmaci, la sperimentazione tende a non tenere in sufficiente considerazione la loro specificità e il cambiamento delle condizioni di salute femminile, con un conseguente incremento di danni avversi all'assunzione di farmaci». Invero, «La donna non può essere assimilata all'uomo, come una mera variabile, ma ha una specificità che la sperimentazione è chiamata a tenere in considerazione per promuovere una medicina che riconosca adeguatamente le pari opportunità uomo/donna».

La parzialità dei set di dati utilizzati a fini di ricerca scientifica ha caratterizzato, di recente, alcune app, ad esempio *Derm Assist* e *SkinVision*, che visualizzando le fotografie scattate con lo *smartphone* possono agilmente individuare una serie di patologie cutanee. In particolare, queste tecnologie provvedono a segnalare le escrescenze come innocue o "ad alto rischio" e, dunque, consigliano all'utente se rivolgersi o meno alle cure. Tuttavia, è prontamente emerso che le applicazioni suddette utilizzano algoritmi non adeguatamente addestrati e producono pregiudizi sistematici relativi alla razza, all'età e, addirittura, al tipo di assicurazione stipulata in quanto sono stati sviluppati a partire da immagini riferibili pressoché unanimemente a pazienti anziani, maschi e bianchi[16].

Anche le decisioni automatizzate assunte in ambito lavorativo hanno prodotto analoghi effetti discriminatori. Il caso più celebre ha ad oggetto l'algoritmo che Amazon aveva progettato per automatizzare la procedura di reclutamento del personale. Tale sistema, infatti, era stato allenato attraverso i curricula ricevuti dalla società nell'arco dei dieci anni precedenti che provenivano, in larga parte, da individui di sesso maschile. Seppur non fosse stato inserito il sesso come criterio selettivo, l'algoritmo era riuscito a riconoscerlo da altre informazioni e, realizzando una discriminazione di tipo "intersezionale", favoriva nella valutazione i termini presenti in maggior numero all'interno dei curricula degli uomini[17].

---

[14] Sulle decisioni pregiudizievoli in sanità si rinvia a: Z. Obermeyer - B. Powers - C. Vogeli - S. Mullainathan, *Dissecting racial bias in an algorithm used to manage the health of populations,* in *Science*, 366,6464, 2019, 447 ss.; N.Norori - Q.Hu - F. M. Aellen - F. D. Faraci - A. Tzovara, *Addressing bias in big data and AI for health care: A call for open science*, in *Patterns*, 2(10), 2021, 1 ss.; R.B.Parikh - S.Teeple - A.S. Navathe, *Addressing bias in artificial intelligence in health care*, in *JAMA*, 322, 2377-2378.

[15] Sul tema F. Franconi - I. Campesi, *Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women*, in *British Journal of Pharmacology*, 171(3), 2014, 580 ss.; v. anche A. Carnevale - E. A. Tangari - A. Iannone - E. Sartini, *Will Big Data and personalized medicine do the gender dimension justice?*, in *AI & Society*, 38(2), 2023, 829 ss.

[16] J. Madhusoodanan, *These apps say they can detect cancer. But are they only for white people?*, in *The Guardian*, 28 agosto 2021.

[17] J. Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women,* in *Reuters*, 11 ottobre 2018.

Sulla cecità discriminatoria dell'algoritmo a discapito dei lavoratori è interessante rammentare il contenuto dell'ordinanza del Tribunale di Bologna (sez. lavoro) del 31 dicembre 2020[18] che ha accolto il ricorso presentato da alcune associazioni sindacali contro una nota società di *food delivery* poiché riteneva che il sistema con cui venivano gestite le prenotazioni dei turni dei *riders*, basato sul cosiddetto Algoritmo *Frank*, fosse discriminatorio. Infatti, tale modello assegnava un determinato punteggio in base alla effettiva partecipazione ai turni prescelti oppure in base alla tempestiva disdetta comunicata con un preavviso di ore ventiquattro. Dunque, l'algoritmo penalizzava allo stesso modo sia il *rider* che non aveva partecipato al turno prescelto per scarsa professionalità sia quello che non vi aveva potuto partecipare perché, ad esempio, aveva deciso di esercitare il proprio diritto di sciopero.[19]

Con riguardo ad una vicenda simile, la Sezione Lavoro del Tribunale di Palermo con sentenza resa in data 31 marzo 2023 ha ritenuto che «è antisindacale e pertanto va repressa con gli strumenti propri della procedura di cui all'art. 28 Stat. Lav. la condotta posta in essere dal datore di lavoro che, servendosi di *riders* per garantire la fornitura dei beni e servizi prodotti, si avvalga di una piattaforma digitale il cui meccanismo di funzionamento, segreto, appaia lesivo oltre che discriminatorio nella scelta del ciclofattorino cui affidare la commessa», peraltro, secondo il Giudice, il datore di lavoro avrebbe l'onere di «palesare le modalità di funzionamento dello strumento elettronico in questione».

Se in Italia i limiti all'utilizzo di piattaforme online per garantire la tutela dei lavoratori contro le discriminazioni algoritmiche sono da rintracciarsi nelle varie pronunce giurisprudenziali, altri ordinamenti europei sono stati "tempestivi" nella regolazione del fenomeno. Primo fra tutti il legislatore spagnolo che, col Real Decreto ley n. 9/2021 convertito nella ley n. 12 del 28 settembre 2021, è intervenuto sullo *Estatuto de los trabajadores* modificandone l'art. 64, relativo ai diritti di informazione e consultazione della rappresentanza legale dei lavoratori. Al nuovo par. 4, lett. d), la norma dispone che il comitato aziendale deve essere informato dall'azienda dei parametri, delle regole e delle istruzioni su cui si basano gli algoritmi o i sistemi di intelligenza artificiale che influiscono sul processo decisionale che può avere un impatto sulle condizioni di lavoro, sull'accesso e sul mantenimento dell'impiego, compresa la profilazione.

Le tecniche di *machine learning* trovano vasto impiego anche nella valutazione del rischio di credito, poiché a differenza degli approcci statistici tradizionali detengono un livello di accuratezza superiore e sono in grado di elaborare un'enorme quantità di dati in volume (numero di osservazioni) e ricchezza (numero di variabili, tipologie di dato)[20].

---

[18] Per una analisi puntuale dell'argomento: S. Borelli - M. Ranieri, *La discriminazione nel lavoro autonomo. Riflessioni a partire dall'algoritmo Frank*, in *Labour & Law Issues*, 7(1), 2021; M. Borzaga - M. Mazzetti, *Discriminazioni algoritmiche e tutela dei lavoratori: riflessioni a partire dall'Ordinanza del Tribunale di Bologna del 31 dicembre 2020*, in *BioLaw Journal*, 1, 2022; A. Perulli, *La discriminazione algoritmica: brevi note introduttive a margine dell'Ordinanza del Tribunale di Bologna*, in *Lavoro Diritti Europa*, 1, 2021.

[19] Sulle differenze di funzionamento degli algoritmi c.d. *rule-based* (es. l'algoritmo Frank) rispetto a quelli di *machine learning* (utilizzato nel caso Amazon) si rinvia a G. Gaudio, Le *discriminazioni algoritmiche,* in *Lavoro Diritti Europa*, 1, 2024, 5. Più diffusamente sul tema, M. Barbera, *Discriminazioni algoritmiche e forme di discriminazione*, in *Labour & Law Issues*, 7(1), 2021.

[20] E. Bonaccorsi Di Patti - F. Calabresi - B. De Varti - F. Federico - M. Affinito - M. Antolini - F. Lorizzo - S. Marchetti - I. Masiani - M. Moscatelli - F. Privitera - G. Rinna, *Intelligenza artificiale nel credit scoring.*

In sostanza, quando un istituto di credito riceve una richiesta di finanziamento da parte di un cliente il cosiddetto *credit scoring* valuta tutti i suoi dati personali (regolarità nei pagamenti, morosità, debiti, etc.) e, di conseguenza, decide se concedere o meno il prestito richiesto.

In questo settore[21] è diffusa la configurazione di *biases* "storici",[22] ossia: i soggetti appartenenti ad uno specifico genere, gruppo etnico o sociale a cui in passato era stato negato l'accesso al credito vengono sottorappresentati nei dati di riferimento dell'algoritmo a vantaggio delle categorie cui storicamente è stato concesso.

A titolo d'esempio si ricorda la vicenda[23] giunta sino alla Corte di giustizia europea avente ad oggetto le controversie di un'agenzia privata di informazione creditizia, la Schufa.

Nella specie, una cittadina tedesca si era vista negare la concessione di un mutuo da parte della banca a fronte di un *credit scoring* negativo fornito dall'agenzia che si era rifiutata, altresì, di fornire una puntuale motivazione con riguardo alla decisione assunta invocando il segreto commerciale sulle informazioni relative al metodo di calcolo utilizzato. Dinnanzi al Giudice europeo la Schufa ha costruito la propria difesa sulla estraneità della agenzia a qualsiasi responsabilità in quanto si sarebbe limitata a fornire una valutazione all'istituto di credito che, in seguito, avrebbe adottato materialmente la decisione.

La Corte ha osservato che la lesione del diritto della ricorrente si è realizzata nel momento in cui la Schufa ha espresso il proprio giudizio negativo sull'affidabilità creditizia che va considerato come vera e propria "decisione" così come intesa nella fattispecie di cui all'art. 22 del GDPR. Infatti, se l'attività di *scoring* venisse relegata a mero momento preparatorio della decisione dell'istituto di credito, l'interessato non potrebbe esercitare il proprio diritto di accesso nei confronti della agenzia che elabora la valutazione di affidabilità creditizia né potrebbe ottenere eventuali informazioni dall'istituto mutuante che generalmente non ne dispone.

Un ricorso[24] similare presentato dinnanzi alla Corte di giustizia riguarda una consumatrice austriaca che si è vista negare il rinnovo di un abbonamento di telefonia mobile del costo di dieci euro mensili a seguito di una valutazione di inaffidabilità finanziaria fornita in modo automatizzato (e non conosciuto dalla ricorrente) dalla Società B&D.

Tuttavia, sugli aspetti prettamente giuridici delle ultime due vicende giudiziarie citate ci si soffermerà al terzo paragrafo.

Analisi di alcune esperienze nel sistema finanziario italiano, in *Questioni di economia e finanza (occasional papers)*, 721, 2022, 30.

[21] Per una panoramica completa sul tema si veda: G. Curcurutu - P. Inturri, *Discriminazioni algoritmiche e tutela dei consumatori vulnerabili nell'accesso al credito*, in *BioLaw Journal*, 1, 2024, 317 ss.

[22] Ivi, 34 ss. per tutte le tipologie di distorsioni configurabili in questo settore.

[23] CGUE, C-634/21, *OQ v Land Hessen* (2023).

[24] CGUE, C-203/22, *Dun & Bradstreet Austria*.

2. L'opacità delle procedure decisionali

Tra le problematiche «identificate ma non risolte»[25] che investono le applicazioni di *machine learning* vi è anche la cosiddetta *black box*, che consiste in «un sistema il cui funzionamento è misterioso; possiamo osservare i suoi ingressi e le sue uscite, ma non sappiamo come l'uno si trasformi nell'altro»[26].
In altri termini, se si verifica la *black box* non è possibile tracciare l'iter logico seguito dalla macchina per raggiungere l'obiettivo assegnato ed è, altresì, impossibile comprendere come e perché, sulla base del set di dati elaborati, il sistema sia giunto a determinati risultati [27].
Tali applicazioni di intelligenza artificiale sono state assimilate ad «oracoli che fanno pronostici»[28] in quanto non sono in grado di accompagnare il dispositivo ad una motivazione logica e, di conseguenza, rendono maggiormente complicate la prevenzione e la repressione di eventuali effetti discriminatori.
A ben vedere le implicazioni derivanti dall'esponenziale propagazione dei sistemi di intelligenza artificiale e i timori di eventuali ripercussioni sui diritti fondamentali risalgono a tempi assai remoti, atteso che già nel 1950, Norbert Wiener, fondatore della cibernetica, faceva menzione nella sua opera[29] del racconto *The Monkey's Paw* di William Wymark Jacobs pubblicato in Inghilterra nel 1902 in cui si narra di due coniugi che, dopo aver espresso il desiderio di ricevere (a tutti i costi) duecento sterline dinnanzi ad un talismano, apprendono la tragica notizia della morte del figlio a seguito di un incidente sul lavoro. In seguito, i genitori, come risarcimento del danno, percepiscono proprio una somma di denaro di importo pari a duecento sterline. Wiener, non a caso, ha menzionato questa storia per rappresentare i ciechi meccanismi di funzionamento dei dispositivi di intelligenza artificiale che perseguono pedissequamente gli obiettivi dettati loro dagli esseri umani non tenendo in considerazione le conseguenze negative derivanti dai procedimenti decisionali.
L'effetto *black box* costituisce una problematica "tecnica" del sistema, poiché discende dalla complessità dei calcoli e dalla natura non lineare dei processi automatici di elaborazione, talvolta imperscrutabili agli stessi programmatori [30]. Infatti, spesso, procedere con tecniche di *reverse engineering* risulta impossibile o eccessivamente oneroso[31].

---

[25] A. Longo - G. Scorza, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti, libertà*, Milano, 2020, 206.

[26] F. Pasquale, *The Black Box Society. The Secret Algorithms that Cont Money and Information*, Cambridge, 2016.

[27] G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in *federalismi.it*, 16, 2021, 117. Sugli effetti positivi dell'opacità nei sistemi di intelligenza artificiale si veda L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, 153 ss.

[28] T. Numerico, *Big data e algoritmi. Prospettive critiche*, Roma, 2021, 133. Sull'utilizzo del termine oracolo per definire i sistemi di *machine learning* si rimanda a G. Finocchiaro, *Intelligenza artificiale: quali regole?*, Bologna, 2024, 21 ss.

[29] N. Wiener, *The human use of human beings: Cybernetics and society*, Boston, 1950. Sul «ritorno della zampa di scimmia» v. N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, 2023, 96 ss.

[30] G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 2021, 117.

[31] Sulla possibilità di un controllo controfattuale v. F. Donati, *Intelligenza artificiale e giustizia*, in *Rivista*

Proprio sulla trasparenza e sulla comprensibilità della procedura decisionale automatizzata è intervenuta la più autorevole giurisprudenza amministrativa italiana[32] al fine di colmare una lacuna legislativa «con l'apparato normativo del diritto pubblico»[33] in merito all'utilizzo di un algoritmo nella formulazione delle graduatorie degli insegnanti vincitori di un concorso pubblico e assegnati alle sedi sulla base di criteri non noti e non trasparenti.[34] Sul punto va rammentato che dopo l'entrata in vigore della legge n. 107/2015, nota come la "Buona scuola", che prevedeva un piano straordinario di assunzioni a tempo indeterminato e di mobilità su scala nazionale, il MIUR, al fine di gestire più agilmente l'ingente numero di assegnazioni, aveva individuato come soluzione l'utilizzo di un *software*.

Il Consiglio di Stato nella sentenza n. 2270 dell'8 aprile 2019 ha accolto il ricorso presentato da alcuni insegnanti, i quali lamentavano il fatto che la procedura di assunzione fosse gestita interamente da un algoritmo e che tale meccanismo avesse dato luogo a provvedimenti privi di qualsiasi motivazione senza individuare alcun funzionario che si occupasse di valutare le singole situazioni e le preferenze indicate nonché di esternare le relative determinazioni provvedimentali. Nel caso di specie, infatti, in maniera del tutto illogica, ai candidati meglio posizionati in graduatoria erano state assegnate, ai fini dell'individuazione delle sedi di servizio, province lontane da quelle di residenza, mentre i candidati che avevano ottenuto un punteggio inferiore avevano potuto beneficiare di posti nella provincia di residenza, nella disciplina e nell'ordine di scuola espressi nella domanda di assunzione.

Il Giudice amministrativo, nella suddetta pronuncia, ha rilevato che un più elevato livello di digitalizzazione dell'amministrazione pubblica è fondamentale per migliorare la qualità dei servizi resi ai cittadini, in particolare, dall'automazione del processo decisionale della pubblica amministrazione derivano indiscutibili vantaggi in relazione

AIC, 1, 2020, 428.

[32] Più nel dettaglio sulla decisione amministrativa automatica: P. Otranto, *Riflessioni in tema di decisione amministrativa, intelligenza artificiale e legalità,* in *federalismi.it*, 7, 2021; Fulvio Costantino, *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei big data*, in *Diritto pubblico, Rivista fondata da Andrea Orsi Battaglini,* 1, 2019, 43 ss.; F. Patroni Griffi, *La decisione robotica e il giudice amministrativo*, in *Giustizia Amministrativa,* 28 agosto 2018; G. Carullo, *Decisione amministrativa e intelligenza artificiale,* in *Diritto dell'informazione e dell'informatica*, 3, 2021; E. Prosperetti, *Accesso al software e al relative algoritmo nei procedimenti amministrativi e giudiziali. Un'analisi a partire da due pronunce del TAR Lazio,* in *Diritto dell'informazione e dell'informatica*, 4-5, 2019; M. C. Cavallaro - G. Smorto, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo,* in *federalismi.it*, 16, 2019; R. Ferrara, *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale,* in *Diritto amministrativo*, 4, 2019; M. Timo, *Algoritmo e potere amministrativo,* in *Il diritto dell'economia*, 1, 2020; F. Laviola, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa,* in *BioLaw Journal*, 3, 2020; A. Simoncini, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019.

[33] M. Palmirani, *Interpretabilità, conoscibilità, spiegabilità dei processi decisionali automatizzati,* in U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 74. Sul tema vedi anche B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica,* in *BioLaw Journal*, 2, 2021, 368; N. Paolantonio, *Il potere discrezionale della pubblica automazione. Sconcerto e stilemi. (Sul controllo giudiziario delle "decisioni algoritmiche")*, in *Diritto Amministrativo*, 4, 2021, 820.

[34] Come fa notare G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione,* cit., 117 il termine *black box* «richiama, immediatamente, e per contrapposizione, la più nota e risalente metafora dell'amministrazione come "casa di vetro"».

ai canoni di efficienza ed economicità dell'azione amministrativa, i quali, declinando quanto previsto dall'art. 97 Cost. «impongono all'amministrazione il conseguimento dei propri fini con il minor dispendio di mezzi e risorse e attraverso lo snellimento e l'accelerazione dell'iter procedimentale». Pertanto, nei casi come quello de quo, implicanti l'elaborazione di ingenti quantità di istanze, relativi «ad una procedura di assegnazione di sedi in base a criteri oggettivi» e, comunque, se è necessario svolgere «operazioni meramente ripetitive e prive di discrezionalità»[35], l'utilizzo dell'algoritmo risulta essere vantaggioso poiché evita qualsiasi negligenza e/o dolo del funzionario ed assicura una maggior garanzia di imparzialità della decisione automatizzata.

Tuttavia, il Giudice precisa che l'algoritmo deve essere considerato come un «atto amministrativo informatico» e, quindi, deve soggiacere ai principi generali dell'attività amministrativa, quali quelli di pubblicità e trasparenza, di ragionevolezza e di proporzionalità; all'algoritmo non devono essere lasciati spazi applicativi discrezionali, ma «deve prevedere con ragionevolezza una soluzione per tutti i casi possibili, anche i più improbabili»; l'amministrazione deve, inoltre, compiere *ex ante* un ruolo di composizione di interessi «anche per mezzo di costanti test, aggiornamenti e modalità di perfezionamento dell'algoritmo»; il giudice può «per la prima volta sul piano "umano"» valutare la correttezza del processo automatizzato in tutte le sue componenti.

Nello specifico il Collegio ha sottolineato che l'algoritmo deve poter essere conoscibile in tutti i suoi aspetti: «Dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti». Ciò, anche in conformità al diritto di difesa del cittadino, al quale non può essere preclusa la conoscenza delle modalità con cui è stata assunta una decisione destinata a ripercuotersi sulla sua sfera giuridica.

Con le successive sentenze n. 8472 del 13 dicembre 2019 e n. 881 del 4 febbraio 2020 la VI sezione del Consiglio di Stato ha compiuto un ulteriore passo in avanti osservando che non vi sono ragioni di principio «per limitare l'utilizzo all'attività amministrativa vincolata piuttosto che discrezionale, entrambe espressione di attività autoritativa svolta nel perseguimento del pubblico interesse».

Inoltre, in entrambe le pronunce viene statuito che non può assumere rilievo la riservatezza[36] delle imprese produttrici dei meccanismi informatici utilizzati in quanto gli stessi, ponendosi al servizio del potere autoritativo, ne accettano le relative conseguenze in termini di necessaria trasparenza.

Il Collegio ha evidenziato, peraltro, che, in virtù di quanto previsto dal diritto nazionale ed europeo, emergono sostanzialmente tre principi da tenere in debita considerazione affinché l'utilizzo degli strumenti informatici avvenga in maniera idonea: il principio di conoscibilità e di comprensibilità dell'algoritmo; il principio della non esclusività della

[35] Sull'ammissibilità dell'atto automatizzato vincolato in dottrina si registra ormai da tempo un consenso pressoché unanime, vedi tra tutti L. Viola, *L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte*, in *federalismi.it*, 21, 2018, e G. Duni, *L'utilizzabilità delle tecniche elettroniche nell'emanazione degli atti e nei procedimenti amministrativi. Spunto per una teoria dell'atto emanato nella forma elettronica*, in *Rivista amministrativa della Repubblica Italiana*, CXXIX, 1978, 407 ss.

[36] Sulle questioni relative alla rilevanza del "codice sorgente" ai fini difensivi ed alle esigenze di riservatezza e di sicurezza informatica vedi anche TAR Lazio Roma, Sez. III Bis, 30 giugno 2020, n. 7370 e TAR Lazio Roma, Sez. III Bis, 1° luglio 2020 n. 7526.

decisione algoritmica: nel processo decisionale, infatti, deve esserci un'interazione con l'essere umano prima di produrre un risultato conformemente al modello cosiddetto HITLM (*human in the loop*)[37]; il principio di non discriminazione algoritmica.

È stata proprio la giurisprudenza, in mancanza di un documento normativo che facesse esplicito riferimento alla possibilità di utilizzare l'Intelligenza Artificiale nel settore pubblico, ad aver individuato alcune restrizioni con riguardo alla possibilità di fare ricorso alla decisione amministrativa algoritmica.

Solo con il Nuovo Codice degli appalti - d.lgs. 36/2023 è stata finalmente positivizzata l'interpretazione giurisprudenziale sopra menzionata disponendo espressamente che le decisioni assunte mediante automazione devono rispettare i principi di conoscibilità e comprensibilità; non esclusività e "umanità" della decisione algoritmica; non discriminazione algoritmica (art.30).

In altri Stati europei, invero, il riferimento normativo alla decisione algoritmica amministrativa è stato introdotto tempo addietro.

In Spagna l'art. 30 della ley n. 11/2007 rubricato *Actuación administrativa automatizada* dispone che in caso di azione automatizzata, l'organismo o gli organismi competenti devono essere preventivamente istituiti, a seconda dei casi, per la definizione delle specifiche, la programmazione, la manutenzione, la supervisione e il controllo della qualità e, se del caso, l'*audit* del sistema di informazione e del suo codice sorgente. All'art. 41 della ley 40/2015 viene invece fornita una definizione dell'azione amministrativa automatica che consiste in qualsiasi atto o azione compiuta interamente per via elettronica da una Pubblica Amministrazione nell'ambito di un procedimento amministrativo e in cui un dipendente pubblico non sia intervenuto direttamente.

In Francia l'automazione della decisione amministrativa implica, secondo quanto previsto dal 2016 nel *Code des relations entre la public et l'administration*, che la Pubblica Amministrazione deve previamente informarne l'interessato e fornirgli contestualmente delle indicazioni sul trattamento dei propri dati.

Il legislatore tedesco[38], sempre nel 2016, ha previsto la possibilità di assegnare tutti gli atti della procedura amministrativa ad un *software* con l'introduzione dell'art. 35a nella legge generale sulle procedure amministrative (*Verwaltungsverfahrensgesetz*), il quale, nello specifico, dispone che un atto amministrativo può essere emesso interamente con mezzi automatizzati se ciò è consentito dalla legge e non vi è discrezionalità o margine di giudizio.

In prospettiva extraeuropea, l'Argentina, con il decreto legislativo 733/2018 del *Ministero de Modernización*, ha fatto per la prima volta riferimento all'utilizzo di decisioni automatizzate nel settore pubblico al fine «di costruire un governo più aperto e collaborativo che si adatti alla vita sempre più digitale e mobile dei cittadini nella società dell'informazione»[39].

---

[37] Si rimanda a M.L. Jones, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Sciences*, 47(2), 2021.

[38] E. Buoso, *Fully Automated Administrative Acts in the German Legal System*, in *European Review of Digital Administration & Law*, 1, 2020, 113 ss.

[39] Per un confronto fra Italia e Argentina sul tema dell'amministrazione digitale: D. U. Galetta - J. G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *federalismi.it*, 3, 2019.

3. Gli strumenti forniti dal GDPR

Pur non esistendo (fino a poco tempo fa) una normativa organica in materia, un richiamo alle decisioni automatizzate era già presente nella direttiva europea 95/46 del Parlamento Europeo e del Consiglio, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati. In particolare, l'art. 15 della normativa dispone che «Gli stati membri riconoscono a qualsiasi persona il diritto di non essere sottoposta ad una decisione che produca effetti giuridici o abbia effetti significativi nei suoi confronti fondata esclusivamente su un trattamento automatizzato di dati destinati a valutare taluni aspetti della sua personalità, quali il rendimento professionale, il credito, l'affidabilità, il comportamento.»
Tuttavia, la ratio di tale disposizione è stata fortemente messa in discussione dalla interpretazione riduttiva che ne hanno dato le legislazioni nazionali nonché da un progressivo mutamento del significato attribuito al termine "decisione" a seguito del raffinarsi di tecniche di costruzione dei profili[40].
Il regolamento sulla protezione dei dati personali 2016/679, senza alcuna pretesa di completezza[41], contiene alcune disposizioni relative all'utilizzo di specifiche tecniche di intelligenza artificiale.
L'art. 22 al par. 1 dispone in via generale che «L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente[42] sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».
Tuttavia, il secondo paragrafo prevede alcune eccezioni al predetto divieto: a) qualora il trattamento sia necessario per la conclusione o l'esecuzione di un contratto; b) se vi sia una disposizione del diritto dell'UE o di uno stato membro che lo autorizzi; c) in presenza del consenso dell'interessato.
Il terzo paragrafo precisa che nelle ipotesi di cui alle lettere a) e c) devono essere applicate misure appropriate per tutelare i diritti, le libertà e gli interessi legittimi dell'interessato, quali il diritto a richiedere l'intervento umano, il diritto di esprimere la propria opinione e di contestare la decisione.
Da ultimo, il par. 4 dispone che le decisioni di cui al par. 2 non possono basarsi sulle categorie particolari di dati personali di cui all'art. 9, par. 1, a meno che non vi sia il

[40] S. Rodotà, *Il diritto di avere diritti*, Bari, 2012, 328. Per la definizione di profilazione si veda art. 4, par. 4 del regolamento 2016/679: «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica»

[41] Così E. Mantovani, *Intelligenza artificiale e discriminazione: quali prospettive? Il modello inglese del data trust*, in *La Rivista Gruppo di Pisa*, 3, 2021, 373.

[42] Come sottolinea E. Pellecchia, *Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation*, in *Le* Nuove *Leggi Civili Commentate*, V, 2018, 1224-1225 non vi sarebbe unanimità nell'interpretazione dell'avverbio «unicamente». Infatti, parte della dottrina ritiene che la tutela di cui all'art. 22 non possa estendersi a tutte le decisioni che comprendono l'ausilio dell'intervento umano (anche minimo), mentre altri sostengono che, ai fini di tale esonero, l'intervento umano deve essere significativo.

consenso dell'interessato (art. 9, par. 2, lett. a) o che il trattamento non sia necessario per motivi di interesse pubblico (art. 9, par. 2, lett. a).
Inoltre, pur avendo un'efficacia meramente interpretativa, il considerando 71 specifica che il titolare del trattamento deve utilizzare procedure appropriate per la profilazione e assicurare misure tecniche e organizzative adeguate al fine di garantire che siano «rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori» e di impedire «effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero che comportano misure aventi tali effetti».
La Corte di giustizia[43], pronunciandosi su un ricorso in materia di affidabilità creditizia, ha fornito alcuni criteri interpretativi al fine di precisare quali siano, di fatto, le decisioni automatizzate a cui può essere applicato l'art. 22 del regolamento (UE) 2016/679.
La Corte ha sottolineato che l'applicabilità dell'art. 22 è soggetta a tre condizioni cumulative: deve sussistere una decisione; la decisione deve essere «basata unicamente sul trattamento automatizzato, compresa la profilazione» e deve produrre «effetti giuridici (riguardanti l'interessato)» o incidere «in modo analogo significativamente sulla persona».
La definizione di decisione non è contenuta nel GDPR, tuttavia, secondo il ragionamento della Corte, dalla formulazione dell'art. 22 e del considerando 71 si può dedurre che il termine decisione «rinvia non solo ad atti che producono effetti giuridici riguardanti il soggetto di cui trattasi, ma anche ad atti che incidono significativamente su di esso in modo analogo[44]». In altre parole, il Giudice Europeo individua come elemento dirimente, per classificare una decisione come tale, l'incidenza della valutazione sulla sfera personale degli interessati.
Sotto un ulteriore profilo, va rilevato che il considerando 63 del GDPR prevede il diritto dell'interessato ad ottenere informazioni sulla «logica cui risponde qualsiasi trattamento automatizzato dei dati e, almeno quando è basato sulla profilazione, alle possibili conseguenze di tale trattamento».
Ancora, il considerando 71 fa riferimento ad un diritto dell'interessato ad ottenere una spiegazione circa i processi decisionali automatizzati previsti dall'algoritmo.
Come noto, tali disposizioni non hanno un contenuto precettivo. Pertanto, si è cercato di ricavare un diritto alla spiegazione dell'interessato a partire dal combinato disposto degli artt. 22; 13, par. 2, lett. f); 14, par. 2, lett. g); 15, par. 1, lett. h) del GDPR.
Infatti, è ormai pacifico in dottrina, che l'interessato, a seguito di una decisione automatizzata assunta nei suoi confronti (di cui era stato previamente informato *ex* art. 22) può verificare ex post se sia stata effettivamente assunta, mediante l'esercizio del proprio diritto di accesso ex art 15 GDPR[45].

[43] CGUE, C-634/21, *OQ v Land Hessen* (2023).

[44] La Corte nel caso di specie ha precisato che la nozione può ricomprendere certamente «il risultato del calcolo della solvibilità di una persona sotto forma di tasso di probabilità relativo alla capacità di tale persona di onorare impegni di pagamento in futuro».

[45] G. Gaudio, *L'algorithmic management e il problema della opacità algoritmica nel diritto oggi vigente e nella Proposta di Direttiva sul miglioramento delle condizioni dei lavoratori tramite piattaforma*, in *Lavoro Diritti Europ*a, 1, 2021.

Tuttavia, il termine *meaningful* (nel testo inglese del GDPR all'art. 15) deve essere interpretato non come la completa spiegazione del modello matematico sotteso al funzionamento dell'algoritmo, ma come un mero chiarimento che renda comprensibile all'interessato gli effetti che discendono dalla propria scelta[46].

Sull'interpretazione dell'art. 15 è opportuno menzionare le conclusioni dell'avvocato generale della Corte di giustizia europea Jean Richard De La Tour nella causa C-203/22 in cui il giudice del rinvio ha richiesto precisazioni su cosa vada inteso per «informazioni significative sulla logica utilizzata» nell'ambito di un processo decisionale automatizzato ai sensi dell'art. 15, par. 1, lett. h), del GDPR; se dette informazioni comprendano l'algoritmo utilizzato a tal fine e in che misura e con quale grado di concretezza si possa esigere dal titolare del trattamento che comunichi informazioni sufficienti per consentire all'interessato di verificare l'esattezza di dette informazioni e la loro coerenza con la decisione inerente al *rating* di cui trattasi.

In secondo luogo, ha chiesto in che misura la protezione dei segreti commerciali possa influire sull'obbligo del titolare del trattamento di fornire informazioni significative sulla logica sottesa nell'ambito di una decisione automatizzata e quali siano i meccanismi che possono consentire di risolvere tale eventuale conflitto.

L'Avvocato Generale ritiene che l'art. 15, par. 1, lett. h), del GDPR non possa essere interpretato «nel senso che fa gravare sul titolare del trattamento un obbligo di divulgare all'interessato informazioni che, in ragione del loro carattere tecnico, presentano un livello di complessità tale da non poter essere comprese dalle persone che non dispongono di una competenza tecnica particolare».

Aggiunge in seguito: «Si potrebbe certamente sostenere, in nome di una lettura estensiva dell'obbligo di trasparenza, che il controllo della modalità con cui i dati personali sono trattati da un algoritmo impone che quest'ultimo sia rivelato all'interessato. Tuttavia, ritengo che la ragion d'essere di tale obbligo sia quella di consentire a tale persona di comprendere le informazioni che le sono comunicate affinché quest'ultima possa far valere i diritti di cui gode a norma del RGPD. In tale ottica, spiegazioni accessibili che non presuppongono una particolare competenza tecnica sono certamente più "significative" di una formula matematica complessa.»

Sul secondo profilo, l'Avvocato Generale osserva che, qualora le informazioni che devono essere fornite all'interessato in virtù di quanto previsto dall'art. 15, par. 1, lett. h) possano comportare una lesione dei diritti e delle libertà altrui, segnatamente perché contengono dati personali di terzi tutelati da detto regolamento o un segreto commerciale, dette informazioni «devono essere comunicate all'autorità di controllo o all'organo giurisdizionale competenti affinché questi ultimi possano ponderare, con piena cognizione di causa e nel rispetto del principio di proporzionalità e della riservatezza di dette informazioni, gli interessi in gioco e stabilire la portata del diritto di accesso che deve essere riconosciuto a tale persona».

[46] E. Palmerini, *Decisioni algoritmiche e diritto dei dati*, in *giudicedonna.it*, 1-2, 2023,13; v. anche A.D. Selbst - J. Powles, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, 7(4), 2014, 236.

4. AI Act: un argine alla discriminazione algoritmica?

Negli ultimi anni le istituzioni europee hanno mostrato una crescente attenzione al tema dell'intelligenza artificiale enfatizzando la necessità di una *Governance* integrata al fine di implementarne lo sviluppo in armonia con il quadro dei principi e dei diritti fondamentali condivisi dall'UE.

Nel febbraio 2017 il Parlamento Europeo ha approvato una risoluzione[47] invitando gli Stati membri a disciplinare in maniera omogenea gli aspetti civilistici della robotica ed ha elencato una serie di principi etici che dovrebbero permeare il quadro giuridico dell'Unione in quest'ambito, fra i quali: sicurezza, salute, libertà, vita privata, integrità, dignità, autodeterminazione, non discriminazione, protezione dei dati personali, nonché tutti i principi previsti dall'art. 2 del TUE e dalla Carta dei diritti fondamentali dell'UE.

Nel 2018 la Commissione Europea con le comunicazioni nn. 237 e 795 ha dichiarato di voler perseguire tre obiettivi principali in materia di intelligenza artificiale: dare impulso alla capacità tecnologica e industriale dell'Unione Europea e all'adozione dell'IA in tutti i settori economici; prepararsi ai cambiamenti socioeconomici ed assicurare un quadro etico e giuridico adeguato a tali fini.

Ancora, la Commissione Europea a giugno 2018 ha nominato un gruppo di esperti di alto livello che hanno redatto il documento «Orientamenti etici per un'intelligenza artificiale affidabile» pubblicato l'8 aprile 2019 in cui sono stati declinati sette requisiti per un' intelligenza artificiale sicura: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale ed *accountability*.

Ha fatto seguito il Libro Bianco sull'intelligenza artificiale del 19 febbraio 2020[48] con lo scopo di definire «le opzioni strategiche» da seguire per «promuovere l'adozione dell'IA» ed «affrontare i rischi associati a determinati utilizzi di questa tecnologia». Nel documento la Commissione si interrogava sul possibile adeguamento del quadro legislativo vigente nell'UE alle applicazioni di intelligenza artificiale, ad esempio, della direttiva 2000/43/CE sull'uguaglianza razziale; della direttiva 2000/78/CE sulla parità di trattamento in materia di occupazione e di condizioni di lavoro, delle direttive 2004/113/CE; 2006/54/CE sulla parità di trattamento tra uomini e donne per quanto riguarda l'accesso a beni e servizi in materia di occupazione.

Pertanto, dal testo emerge chiaramente, ancor prima dell'adozione di nuove regole, la necessità di una valutazione delle suddette normative in tema di diritti fondamentali al fine di appurare se le stesse possano dirsi applicabili anche alle problematiche derivanti dai sistemi di IA, atteso che l'elasticità che contraddistingue gli ordinamenti giuridici permette, tendenzialmente, di includervi i mutamenti della società.[49]

I numerosi atti di impulso e di *soft law* in materia di intelligenza artificiale hanno con-

[47] Parlamento europeo, Risoluzione del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

[48] Commissione europea, Libro Bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia, COM (2020) 65, 19 febbraio 2020.

[49] G. Finocchiaro, *Intelligenza Artificiale: quali regole?*, cit., 18.

dotto, a seguito di un lungo processo di negoziazione fra gli Stati membri, alla formulazione di una regolamentazione organica in materia, l'*Artificial Intelligence Act*[50], che è stata approvata da una maggioranza di 523 voti favorevoli, contro 46 voti contrari e 49 astensioni.

Va ritenuta opportuna la scelta di affidare la regolazione di tale ambito ad un regolamento, atto di portata europea e di applicazione diretta, così da evitare la frammentazione del mercato unico e l'incertezza giuridica che sarebbe derivata da una regolamentazione esclusivamente nazionale[51].

La normativa segue un approccio basato sul rischio individuando quattro differenti intensità: rischio inaccettabile, rischio elevato, rischio limitato, rischio minimo o nullo.

Le decisioni assunte dai sistemi di intelligenza artificiale in settori particolarmente sensibili, oggetto della seguente trattazione, possono ricondursi alle classi dei sistemi di IA «a rischio inaccettabile» o «ad alto rischio».

Infatti, l'art. 5, lett. c), fra le pratiche vietate prevede l'immissione sul mercato, la messa in servizio o l'uso di sistemi per la valutazione o la classificazione di persone o di gruppi sulla base del loro comportamento sociale o delle loro caratteristiche personali.

Sono, altresì, vietati, ai sensi dell'art.5, lett. d), quei sistemi che misurano la probabilità che un individuo commetta un reato sulla base dei suoi tratti somatici o sulle caratteristiche della sua personalità[52].

Peraltro, l'allegato III del regolamento elenca i settori in cui si sviluppano i sistemi di IA ad alto rischio, fra i quali: istruzione e formazione professionale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi; attività di contrasto; migrazione, asilo e gestione delle frontiere; amministrazione della giustizia e processi democratici.

Tuttavia, come precisa l'art. 6, par. 3, «un sistema di IA non rientra nella categoria ad alto rischio se non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche».

L'utilizzo di un sistema di questa tipologia implica la previa identificazione dei rischi ragionevolmente prevedibili e l'adozione di misure necessarie ad evitarli. In particolare, deve essere effettuato un controllo dei dati utilizzati onde evitare distorsioni o discriminazioni vietate dal diritto dell'Unione Europea. Inoltre, ogni sistema ad alto rischio ha l'obbligo di detenere una documentazione tecnica per dimostrare la conformità ai requisiti previsti dal regolamento e deve registrare l'attività effettuata al fine di garantirne la tracciabilità.

A seguire, l'art. 13 dispone che tali sistemi devono essere progettati in modo tale da consentire ai *deployer*[53] di poter comprendere l'*output* garantendo una «trasparenza ade-

[50] Regolamento (UE) 2024/1689.

[51] In questi termini già la Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)).

[52] La disposizione prevede un'eccezione: «tale divieto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa».

[53] Per la definizione di *deployer* si rinvia all'art 3. n. 4, dell'*AI Act*: «"*deployer*": persona fisica o giuridica,

guata». Il regolamento individua puntualmente le informazioni che il sistema ad alto rischio deve fornire: l'identità e il contatto del fornitore, la finalità del sistema, il livello di accuratezza che il sistema può garantire nonché gli eventuali rischi prevedibili per la salute, per la sicurezza o per i diritti fondamentali, le informazioni circa le caratteristiche tecniche del sistema al fine di comprenderne opportunatamente l'*output*, eventuali informazioni circa le prestazioni del sistema con riguardo a determinate persone o gruppi di persone, le misure di sorveglianza umana.

L'art.14 dell'AI Act dispone che i sistemi ad alto rischio devono essere progettati in modo da garantirne la supervisione da parte di persone fisiche: il fine della sorveglianza umana è quello di prevenire e ridurre al minimo i rischi per la salute, la sicurezza e i diritti fondamentali.

La normativa precisa che le persone fisiche a cui è affidata la sorveglianza devono essere ben consapevoli della gravosa influenza esercitata dal sistema di intelligenza artificiale sulla propria decisione e della tendenziale propensione a fare affidamento sull' *output* suggerito dalla macchina.

Sul ruolo ausiliare della persona fisica rispetto ai sistemi di intelligenza artificiale la dottrina è divisa: vi è chi sostiene che quando verranno introdotte soluzioni alle problematiche dei sistemi di apprendimento automatico ovvero quando non si correrà più il rischio di creare *biases* o meccanismi decisionali opacizzati le decisioni potranno essere del tutto affidate ai sistemi di intelligenza artificiale che, molto più dell'essere umano, sono connotati da coerenza e imparzialità; altri ritengono, invece, che i sistemi di intelligenza artificiale impiegati in ambiti particolarmente sensibili dovranno in ogni caso mantenere una componente umana[54].

Nel regolamento si specifica che le persone fisiche alle quali è affidata la sorveglianza umana devono interpretare l'*output* del sistema tenendo conto di tutti gli strumenti a loro disposizione potendo eventualmente scegliere di non usare il sistema di intelligenza artificiale, di ignorarne la decisione o di intervenire sul funzionamento arrestandone la procedura.

L'art. 15, peraltro, prevede che tali applicazioni di intelligenza artificiale devono conformarsi ad un adeguato livello di accuratezza, robustezza e cybersicurezza.

Inoltre, gli artt. 13, 14, 15 stabiliscono che i sistemi ad alto rischio devono essere «progettati» sin dal principio in modo tale da garantire la piena conformità alle suddette disposizioni e, in adesione al principio cd. *Ethics by design*, all'art. 27 è previsto che, prima di utilizzare un sistema di IA ad alto rischio, i *deployer* hanno l'onere di effettuare una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema potrà produrre.

Ciò a dimostrazione che, in linea con quanto era già stato stabilito dall'art. 25 GDPR (rubricato *Data protection by design and by default*), anche nel Regolamento sull'intelligenza artificiale si predispone una «nuova collocazione delle regole rispetto al fenomeno da regolare» al fine di includere già nella fase della progettazione i valori e i principi fon-

autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

[54] C. Casonato - B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'unione europea in materia di intelligenza artificiale*, in *Rivista di BioDiritto*, 3, 2021, 18.

damentali dell'Unione europea[55].

5. Qualche considerazione conclusiva

Come è stato osservato da autorevole dottrina[56], il regolamento si limita ad individuare delle soluzioni «formali» richiamando i principi generali condivisi dagli stati membri dell'UE e senza, di fatto, introdurre «nuovi efficaci e rapidi strumenti di tutela contro la discriminazione».

Tuttavia, proprio per la natura sfuggevole della materia oggetto della normativa, che quasi risulta «refrattaria alla giuridificazione»[57] e sembra «atteggiarsi a ordine spontaneo»[58], possono ragionevolmente comprendersi le difficoltà sottese alla definizione di un quadro regolatorio specifico.

Per quel che qui maggiormente interessa, va rilevato che l'AI Act non fa menzione delle numerose direttive europee concernenti il diritto antidiscriminatorio. Pare, dunque, che la normativa, almeno per quanto concerne le discriminazioni algoritmiche, non discenda da una consona valutazione (e da una conseguente possibile integrazione) degli strumenti di contrasto già a disposizione, come, peraltro, aveva suggerito anche la Commissione Europea nel Libro Bianco sull'Intelligenza Artificiale.

In ogni caso, per fare un bilancio complessivo si devono attendere gli interventi legislativi e amministrativi degli stati membri in attuazione del Regolamento.

Preme rilevare che l'ordinamento italiano con riguardo alle procedure decisionali automatizzate non prevede disposizioni che regolano appositamente il fenomeno se non, relativamente all'azione amministrativa, quanto contenuto nel Nuovo Codice degli Appalti.

Mentre, come è emerso nel corso della trattazione, altri stati europei e non europei già da molto tempo hanno introdotto delle normative di settore sulle decisioni algoritmiche.

Dunque, in Italia, è stata fondamentale l'interpretazione in via giurisprudenziale che, nel suo ruolo di supplente e in conformità ai principi generali nazionali e sovranazionali (che sono stati ribaditi anche nell'AI Act), ha fissato alcuni limiti in materia di decisione automatizzata, e, come visto, ricorre nelle varie pronunce il diritto dell'interessato alla conoscibilità e alla comprensibilità della logica sottesa alla procedura decisionale.

D'altra parte, però, è pur vero che il titolare del trattamento potrebbe non disvelare la logica sottesa all'algoritmo celandosi dietro alla sua estrema complessità. In tal senso, oltre all'interpretazione fornita dall'Avvocato Generale della Corte di giustizia con riguardo all'art. 15 GDPR, si direziona anche l'art. 13 dell'AI Act che dispone che, rispetto al sistema di IA ad alto rischio utilizzato, venga garantita una trasparenza «adeguata» e non completa.

Premesso ciò, ci si domanda quale rilevanza possa avere l'omissione di tale informazio-

[55] G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 127.

[56] G. Finocchiaro, *Intelligenza artificiale: quali regole?*, cit., 123ss.

[57] A. Celotto, *Algoritmi e algoretica: quali regole per l'intelligenza artificiale?*, in *ConsultaOnline*, 27 marzo 2020, 9.

[58] *Ibid.*

ne sull'esercizio di difesa dell'interessato che nella procedura di decisione automatizzata -il cui meccanismo è a lui sconosciuto- potrebbe essere stato esposto ad una discriminazione algoritmica, considerando, oltretutto, che generalmente nemmeno il giudice ha le competenze per comprendere se la profilazione di un sistema di intelligenza artificiale possa nascondere nella incomprensibilità del suo funzionamento eventuali limiti o errori di impostazione.

Intelligenza artificiale e ricerca accademica: uno sguardo critico tra rischi e innovazione*

Martina Iemma

Abstract

Il contributo intende affrontare le questioni giuridiche connesse all'utilizzo dell'intelligenza artificiale (IA) nella ricerca accademica. Sebbene l'IA possa costituire un grande sostegno nella redazione e nell'*editing* di testi scientifici, allo stesso tempo il suo impiego solleva importanti preoccupazioni: tra le principali criticità emergono in particolare le questioni legate al rischio di plagio, alla paternità intellettuale dell'opera, al *copyright*, nonché alla libertà di espressione tutelata all'art. 21 della Costituzione. Queste possono sollevare dubbi sulla stessa qualità della ricerca, minare l'integrità accademica e ripercuotersi negativamente sul diritto all'informazione. Il contributo analizza tali problematiche con uno sguardo rivolto anche ai più recenti avvenimenti, come la causa intentata dal New York Times contro OpenAI e Microsoft per questioni di *copyright* e la prima proposta di legge in Francia volta a regolare l'attribuzione di paternità delle opere create con IA.

The paper aims to address the legal issues related to the use of artificial intelligence (AI) in academic research. Although AI can be a great support in the writing and editing of scientific texts, at the same time its use raises important concerns: among the main critical issues are the risk of plagiarism, intellectual authorship of the work, copyright, as well as freedom of expression protected in Article 21 of the Constitution. These can raise doubts about the quality of research, undermine academic integrity and negatively affect the right to information. This contribution analyses these issues with an eye to recent events, such as the lawsuit filed by the New York Times against OpenAI and Microsoft for copyright issues and the first law proposal in France aimed at regulating the attribution of authorship of works created with AI.

Sommario

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Keywords

Intelligenza Artificiale – *ChatGPT* – ricerca accademica – plagio – disinformazione

1. Introduzione

Negli ultimi decenni l'Intelligenza Artificiale (IA) ha pervaso con crescente rapidità la nostra quotidianità, diventando parte integrante di moltissime discipline[1]. Dalla medicina alla finanza, dal marketing all'agricoltura, dalla sicurezza all'istruzione è ormai possibile trovare l'IA in quasi tutti gli ambiti di vita, tanto che autori quali S.J. Russel e P. Norving hanno correttamente osservato che «*AI is relevant to any intellectual task; it is truly a universal field*»[2].

Il lancio di ChatGPT, un software di IA sviluppato da OpenAI e reso accessibile al grande pubblico anche in una versione gratuita verso la fine del 2022, ha probabilmente segnato l'inizio di un nuovo modo di approcciarsi all'Intelligenza Artificiale, rendendola facilmente accessibile e disponibile a tutti. Da quel momento è diventato sufficiente possedere una connessione ad internet per poter toccare con mano l'interazione con un'IA. ChatGPT, come si vedrà, è un'Intelligenza Artificiale c.d. generativa ed è in grado non solo di comprendere richieste formulate in un linguaggio naturale, ma anche di fornire risposte (auspicabilmente) precise e complete e di condurre una vera e propria conversazione su una vasta gamma di argomenti, nonché di creare testi articolati e argomentati come se fossero stati scritti da un essere umano.

L'IA ha iniziato ad essere applicata, più o meno dichiaratamente e consapevolmente, anche in campo accademico, al fine di velocizzare i lavori di ricerca e scrittura, che rappresentano due delle attività quotidiane e più rilevanti per un ricercatore.

Se tale utilizzo non va di per sé demonizzato e può presentare vantaggi per il lavoro degli accademici, è necessario però porsi alcune fondamentali domande sull'effettiva opportunità del suo utilizzo, sugli eventuali limiti da porre e, non da ultimo, sui rischi connessi all'impiego di simili tecnologie. È proprio nel tentativo di dare una risposta a tali interrogativi che nasce il presente contributo, il quale intende, da un lato, mettere in luce i benefici che gli accademici potrebbero trarre dall'Intelligenza Artificiale e, più nello specifico, dall'IA generativa[3]; dall'altro lato, affrontare alcune delle principali questioni giuridiche che potrebbero sorgere dal relativo impiego nella scrittura di articoli

[1] L. Portinale, *Intelligenza Artificiale: storia, progressi e sviluppi tra speranze e timori*, in *MediaLaws*, 3, 2021, 14.

[2] S. J. Russell-P. Norvig, *Artificial Intelligence. A Modern Approach*, Londra, 2021, 1.

[3] Come evidenziato in C. Colapietro-A. Moretti, *L'Intelligenza Artificiale nel dettato costituzionale: opportunità, incertezze e tutela dei dati personali,* in *BioLaw Journal*, 3, 2020, 369: «L'IA può contribuire a dare un forte impulso alla ricerca, sia aprendo nuovi filoni di indagine, sia configurandosi essa stessa come strumento attraverso cui svolgere attività di ricerca scientifica».

destinati alla pubblicazione in riviste accademiche. In particolare, con riferimento alle problematiche connesse all'utilizzo dei sistemi di IA, capaci di produrre opere simili a quelle create dall'ingegno umano, ci si soffermerà sull'attribuzione di paternità di un testo generato con IA, sul rischio di plagio e sul rischio di diffusione di disinformazione[4].

Preliminarmente e in via generale, è opportuno evidenziare che la rapida diffusione dell'Intelligenza Artificiale ha spinto a domandarsi se sia o meno opportuna una specifica regolamentazione di tale ambito e dell'utilizzo di tali software. Sebbene ad oggi non vi siano ancora regolamentazioni di carattere nazionale in materia, a livello sovranazionale il 12 luglio 2024 è stato pubblicato sulla Gazzetta ufficiale dell'Unione Europea il Regolamento (UE) 2024/1689 (c.d. AI Act). Esso è il primo atto normativo sull'utilizzo dell'Intelligenza Artificiale, volto a favorire lo sviluppo e l'adozione di sistemi di IA sicuri e affidabili nel mercato unico dell'UE e allo stesso tempo assicurare il rispetto dei diritti fondamentali dei cittadini[5]. Lo scopo di tale regolamento è pertanto quello di disciplinare in maniera organica in tutta l'Unione Europea l'impiego dell'Intelligenza Artificiale. Esso si inserisce nel quadro normativo europeo vigente in materia e, per quanto riguarda i contenuti protetti da diritto d'autore e usati per l'addestramento dei sistemi di Intelligenza Artificiale, ci si riferisce in particolare alla Direttiva sul diritto d'autore nel mercato unico digitale del 2019[6], che ha l'obiettivo di armonizzare il quadro normativo europeo del diritto d'autore nello specifico ambito delle tecnologie digitali e di internet.

Inoltre, il 14 novembre 2024 la Commissione Europea ha pubblicato la prima bozza del General-Purpose AI Code of Practice[7], come previsto dall'art. 56 dell'AI Act[8]. Si tratta del primo Codice di condotta per l'Intelligenza Artificiale di uso generale (AI general-purpose, GPAI[9]) redatto da esperti indipendenti, nominati presidenti e vicepresidenti di quattro gruppi di lavoro tematici aventi ad oggetto: trasparenza e norme relative al *copyright*; identificazione e valutazione del rischio per il rischio sistemico;

---

[4] F. Posteraro, *Il copyright al tempo dell'IA generativa*, in questa *Rivista*, 2, 2023, 11.

[5] Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio, 13 giugno 2024, pubblicato in GUUE il 12 luglio 2024. L'AI Act è una proposta di regolamento presentata dalla Commissione Europea il 21 aprile 2021, con lo scopo di instaurare un quadro normativo armonizzato per l'Intelligenza Artificiale nell'Unione Europea; è stato approvato il 13 marzo 2024 dal Parlamento Europeo ed è stato successivamente approvato in via definitiva dal Consiglio dell'Unione Europea il 21 maggio 2024; pubblicato in GUUE il 12 luglio 2024, con entrata in vigore il 2 agosto 2024.

[6] Direttiva (UE) 2019/790 del Parlamento Europeo e del Consiglio, del 17 aprile 2019, sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE.

[7] First Draft General-Purpose AI Code of Practice, 14 novembre 2024.

[8] L'art. 56 dell'AI Act prevede l'elaborazione di codici di buone pratiche a livello dell'Unione Europea al fine di contribuire alla corretta applicazione del regolamento stesso.

[9] L'art. 3, par. 63 dell'AI Act definisce l'Intelligenza Artificiale di uso generale come «un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato».

mitigazione del rischio tecnico per il rischio sistemico; mitigazione del rischio di governance per il rischio sistemico.
Il documento, che entrerà in vigore dopo un processo di discussioni interne nei quattro gruppi di lavoro e di ulteriori *input* esterni da parte degli stakeholder, farà da guida allo sviluppo e implementazione di modelli di IA generici che siano sicuri e affidabili, fornendo regole dettagliate relative alla trasparenza e al *copyright* per i fornitori di questi modelli di IA.
Restringendo il campo di indagine all'ambito accademico, in assenza di norme specifiche alcune riviste scientifiche e alcune case editrici hanno iniziato ad adottare linee guida relative all'utilizzo dell'IA nella redazione degli articoli da pubblicare. Se da un lato questo potrebbe essere utile a scongiurare, o quantomeno ridurre, il rischio di violazione di diritti altrui – si pensi ad esempio al plagio oppure alla violazione di *copyright* –, dall'altro lato non è chiaro se una regolamentazione a livello nazionale o europeo risulterebbe più efficace.
Delineato, in estrema sintesi, lo scarno quadro regolatorio esistente, ci si può dunque chiedere se sia effettivamente desiderabile una regolamentazione dell'IA e, ancora, se strumenti di autoregolamentazione (quali le linee guida stilate dalle case editrici) o di *soft law* (quali codici di condotta) rappresentino una risposta adeguata rispetto alle problematiche giuridiche emergenti dall'utilizzo sempre più diffuso di sistemi di IA. Tali domande, di carattere generale, rimarranno sullo sfondo del presente contributo, ma, auspicabilmente, attraverso l'esame dello specifico ambito di applicazione dell'IA nel contesto delle pubblicazioni accademiche e delle problematiche ad esso connesse, sarà possibile, in sede di conclusioni, cercare di offrirne una risposta.
Il contributo ha ad oggetto questioni relative principalmente all'ambito accademico umanistico, in quanto chi utilizza l'IA per ricerche scientifiche affronta sfide in parte differenti, che non saranno oggetto di trattazione; inoltre, si focalizza prevalentemente sulle questioni giuridiche, lasciando quelle etiche ad altra sede di trattazione.

2. Una premessa necessaria: il problema definitorio dell'IA nel contesto giuridico

Al fine di circoscrivere l'oggetto di trattazione, e data la sua importanza ai fini giuridici, è necessario soffermarsi brevemente sul significato di Intelligenza Artificiale.
Trattandosi di una materia in continua evoluzione e trasformazione, nel tempo sono state formulate differenti definizioni[10], causando forse ancora più incertezza su che

[10] Sul punto si vedano S. J. Russell-P. Norvig, *Artificial Intelligence. A Modern Approach*, cit.; G. Sartor, *L'intelligenza artificiale e il diritto*, Torino, 2022, 3, che evidenzia come da un lato si abbia la contrapposizione tra l'intelligenza intesa come pensiero e l'idea di un'intelligenza in cui prevale l'interazione con l'ambiente; dall'altro lato la contrapposizione tra l'obiettivo di riprodurre fedelmente le capacità intellettive dell'uomo e l'obiettivo di «realizzare sistemi capaci di razionalità (cioè di elaborare informazioni o agire in modo ottimale) prescindendo dai limiti della razionalità umana». Uno dei pionieri della materia, John McCarthy, la definì invece come «*the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable*» in J. McCarthy, *What Is Artificial Intelligence*, Stanford, 2007, 2.

cosa andasse effettivamente regolato[11]. Tuttavia, nonostante l'assenza di una definizione univoca, Massimo Luciani[12], riferendosi al tema della responsabilità del danno, evidenziava come il legislatore avesse in realtà tutti gli strumenti utili a stabilire se la responsabilità di un eventuale danno stesse in capo al produttore dell'hardware, al produttore del software, all'utilizzatore ecc. e, quindi, potesse introdurre «almeno alcuni princìpi generali, sebbene l'estrema varietà del fenomeno possa astrattamente suggerire l'adozione di discipline specifiche per i vari campi di applicazione dell'intelligenza artificiale»[13]; infatti, «la presenza di princìpi generali faciliterebbe la regolazione delle future novità, da attendersi profonde e in rapida successione, che una regolazione analitica sarebbe costretta (sempre in ritardo) a inseguire»[14].

Una delle prime definizioni a livello istituzionale, seppur solamente di *soft law*, si rinviene nella Comunicazione *Artificial Intelligence for Europe* del 2018 della Commissione Europea, che ha definito l'Intelligenza Artificiale come un sistema che si riferisce a *«systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals»*[15]. Ma è con l'AI Act che ne viene finalmente data una definizione rilevante ai fini giuridici: l'art. 3 definisce infatti il "sistema di IA" come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali»[16].

Al fine di comprendere il tema che verrà trattato nei paragrafi successivi è utile accennare ad alcuni profili tecnici. L'Intelligenza Artificiale può infatti basarsi su due approcci: la Modellazione (*Model Based AI*) e l'Apprendimento Automatico (*Machine Learning AI*). Nel primo, il modello deve poter essere inserito e utilizzato da un computer per compiere azioni quali calcolare, analizzare e dare risposte[17]; nel secondo, invece, il modello di un fenomeno si ricava da dati ottenuti da fonti esterne, come ad esempio i dati disponibili sul web, che vengono poi usati per "addestrare" il modello prima del suo utilizzo, indicando alla macchina esattamente quale risultato deve fornire e in che modo[18].

Le tecniche di *Machine Learning* sono quelle che vengono impiegate in una specifica

---

11 M. Luciani, *La sfida dell'intelligenza artificiale*, in *Lettera AIC*, 12, 2023, 7.

12 In uno scritto relativo all'intervento tenuto in occasione dell'incontro denominato *Introduzione all'intelligenza artificiale: tecnologia e diritto*, tenutosi il 16 novembre 2023 per il Convegno al Circolo dei Magistrati della Corte dei conti.

13 M. Luciani, *Può il diritto disciplinare l'intelligenza artificiale? Una conversazione preliminare*, in *Diritto & Conti Bilancio Comunità Persona*, 2, 2023, 16.

14 *Ibid.*

15 Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Artificial Intelligence for Europe*, Brussels, 25.4.2018 COM/2018/237 final.

16 Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio, cit., art. 3.

17 P. Traverso, *Breve introduzione tecnica all'Intelligenza Artificiale*, in C. Casonato-M. Fasan-S. Penasa (a cura di), *Diritto e intelligenza artificiale*, sezione monografica – DPCE Online, 1, 2022, 158.

18 Ivi, 160.

area dell'Intelligenza Artificiale, chiamata "generativa". I modelli di Intelligenza Artificiale generativa vengono infatti addestrati su un'ingente quantità di dati e sono in grado di generare automaticamente nuovi contenuti complessi in modo molto accurato, imitando la creatività umana. Le capacità del sistema, pertanto, dipendono dalla quantità e dalla tipologia di dati impiegati per il relativo addestramento: così si hanno sistemi in grado di produrre testi, simulare conversazioni, generare codici, immagini, musica e altro ancora.

3. L'Intelligenza Artificiale generativa nella ricerca accademica: possibili benefici...

La capacità dei sistemi di Intelligenza Artificiale generativa di analizzare e reinterpretare grandi quantità di dati, nonché di generare testi precisi, potrebbe contribuire positivamente alla ricerca accademica[19]. Ciò che suscita particolare interesse tra gli accademici, però, è la capacità dell'IA generativa di comprendere ed elaborare un "linguaggio naturale", simile a quello generato dall'ingegno umano (c.d. *Natural Language Processing*)[20]. L'esempio più noto è quello di ChatGPT, un software sviluppato da OpenAI e in grado non solo di comprendere richieste formulate in un linguaggio naturale, ma anche di fornire risposte precise e complete e di condurre una conversazione su un'ampia gamma di argomenti[21]. Strumenti come ChatGPT potrebbero avere il vantaggio di rendere la ricerca più rapida ed efficiente, automatizzando alcune azioni come la stesura di approfondimenti, riassunti di articoli, rapporti o altri documenti, a cui lo studioso può poi attingere per elaborare il proprio scritto. Inoltre, uno degli aspetti più innovativi di questi sistemi è la possibilità di rivolgere richieste molto precise ed ottenere risultati pertinenti in tempi incredibilmente brevi. Tali strumenti possono poi essere impiegati per tradurre un testo in diverse lingue, ottenendo così un più agevole accesso ed una migliore comprensione di materiali di ricerca internazionali, soprattutto quando si ha a che fare con lingue poco conosciute, facendo così sparire le barriere linguistiche ed agevolando una raccolta di dati su ampia scala.

L'Intelligenza Artificiale generativa si configura, quindi, come uno strumento potenzialmente utile per gli accademici, che permette di risparmiare tempo nelle attività più meccaniche e ripetitive e di concentrarsi invece sull'aspetto più creativo e analitico del proprio lavoro[22].

Tuttavia, è essenziale che coloro che operano in questo ambito siano consapevoli dei rischi connessi all'utilizzo di tali strumenti. Gli accademici, infatti, sono chiamati a condurre la propria ricerca in modo responsabile e trasparente, fornendo dati veritieri

[19] M.M. Alshater, *Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT*, 26 dicembre 2022, disponibile su SSNR; M. Hosseini-L.M. Rasmussen-D.B. Resnik, *Using AI to write scholarly publications*, in *Accountability in Research,* 31(3), 2023.

[20] M.M. Alshater, *Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT,* cit., 2.

[21] *Ibid.*

[22] B.D. Lund-T. Wang, *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?*, in *Library hi tech news*, 40(3), 2023, 27.

e non distorti[23] ed evitando, quindi, qualsiasi forma di abuso nell'utilizzo di queste tecnologie[24]. Come si vedrà, ciò comporta una puntuale verifica della veridicità delle informazioni e dei dati da parte dell'accademico.

4. ...e potenziali criticità: il problema della paternità del testo generato con IA

La maggior parte delle riviste prevede una politica di paternità dei testi che spesso include, tra i requisiti per essere riconosciuto come autore, la partecipazione al processo di scrittura[25]. Ci si potrebbe pertanto domandare se l'impiego dell'Intelligenza Artificiale nella redazione di un testo implichi o meno il riconoscimento di quest'ultima quale autore o coautore[26].

Per rispondere a tale quesito, parte della dottrina ha delineato differenti scenari riguardanti l'attribuzione della paternità del testo generato con Intelligenza Artificiale generativa, che si distinguono a seconda dell'interazione tra l'utente e il sistema. Se l'utente ha fornito dati di *input* specifici, che hanno guidato il software nella generazione del testo, allora sembrerebbe possibile considerare l'utente come possessore della paternità; tuttavia, se l'*input* fornito è molto limitato (ad esempio l'utente ha semplicemente chiesto al sistema di scrivere un testo su un determinato argomento), allora potrebbe risultare più complicato determinare con certezza chi, tra l'utente e il sistema, sia il vero proprietario del testo[27]. Il problema centrale diviene pertanto quello di determinare chi si assume la responsabilità delle informazioni contenute nel testo.

Inoltre, il problema è acuito dalla distinzione tra Intelligenze Artificiali "pienamente" o "parzialmente" generative: le prime «sono programmate per avere un output quasi o del tutto indipendente dall'input dell'utente»[28]; mentre le seconde «utilizzano tanto le informazioni fornite dal programmatore quanto quelle immesse dall'utente per generare un output»[29]. Pertanto, si potrebbe sostenere che l'*output* prodotto da un'Intelligenza Artificiale pienamente generativa rappresenti «il risultato indiretto del suo progetto creativo versato nel software e che il diritto d'autore sulle opere risultanti spetti quindi

[23] M. Hosseini-L.M. Rasmussen-D.B. Resnik, *Using AI to write scholarly publications,* cit., 1.

[24] B.D. Lund-T. Wang, *Chatting about ChatGPT*, cit., 28.

[25] Come emerge dallo studio condotto in D. B. Resnik, A. M. Tyler et. al, *Authorship policies of scientific journals*, in *Journal of Medical Ethics*, 2016, 42(3), 199-202; gli autori hanno svolto un'analisi delle politiche di *authorship* di un campione casuale di 600 riviste presenti nel database Journal Citation Reports.

[26] M. Hosseini-L.M. Rasmussen-D.B. Resnik, *Using AI to write scholarly publications,* cit., 5.

[27] B.D. Lund et al., *ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing*, in *Journal of the Association for Information Science and Technology*, 74(5), 2023, 575.

[28] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.: conflitti attuali fra intelligenza artificiale e diritto d'autore*, in *giustiziacivile.com*, 2, 2024, 9; l'autore, come esempio di IA pienamente generativa, richiama l'IAA AARON, un'IA che proviene dal mondo anglosassone e che è stata progettata dall'artista britannico Harold Cohen per dipingere autonomamente quadri sulla base delle istruzioni fornite in origine dal medesimo Cohen.

[29] *Ibid.*; vedi anche J.C. Ginsburg-L.A. Budiarjo, *Authors and Machines*, in *Bekeley Technology Law Journal*, 34(2), 2019, 407 ss.

allo sviluppatore»[30].

In caso di utilizzo di ChatGPT, che può essere impiegato sia in modo pienamente che parzialmente generativo, quando l'*input* consiste ad esempio in una generica richiesta di redazione di un testo «sarà OpenAI ad aver indirettamente generato il testo»[31]. Al contrario, l'utilizzo di un'IA parzialmente generativa porta a domandarsi «quando gli *input* dell'utente abbiano un'influenza sul programma dello sviluppatore e in quali casi ciò sia connotato da creatività»[32]. In questo caso l'autore dell'*output* potrebbe essere l'utente, il programmatore, entrambi o nessuno dei due[33].

Seppur si tratti di un argomento complesso e ancora in fase di studio, nonché di tecniche e regolamentazioni in continua evoluzione, alcuni studiosi hanno tentato di elaborare risposte concrete alle questioni più importanti emerse finora: un esempio è dato dagli studi di J.C. Ginsburg e L.A. Budiarjo, i quali hanno provato ad individuare una soluzione[34] al problema dell'attribuzione della paternità dei testi generati con IA. I due autori hanno infatti individuato quattro possibili strade per allocare il diritto d'autore in caso di utilizzo di Intelligenza Artificiale generativa: a) qualora il programmatore non avesse fornito istruzioni precise al sistema di IA su come generare autonomamente un *output* e quest'ultimo dipendesse totalmente dalle decisioni dell'utente, allora il diritto d'autore sull'opera realizzata spetterebbe solo a quest'ultimo, in quanto avrebbe impiegato l'IA come semplice strumento per esprimere la propria creatività; b) qualora l'IA pienamente generativa fosse stata programmata per generare opere creative senza la necessità di alcun contributo da parte dell'utente, allora il diritto d'autore su tali opere spetterebbe unicamente al programmatore, perché esse sarebbero il risultato indiretto del suo lavoro creativo che ha riversato nell'algoritmo; c) se l'IA parzialmente generativa avesse prodotto un'opera risultata dall'unione dell'apporto creativo di programmatore e utente, allora si potrebbe dire che questi siano coautori dell'*output* finale; d) infine, in caso di IA parzialmente generativa, se programmatore e utente non avessero collaborato nella realizzazione dell'opera, attraverso la progettazione e l'inserimento di *input*, allora l'opera sarebbe da considerare priva di autore[35]. Secondo i due autori, in sintesi, per determinare la paternità del testo generato con IA occorre innanzitutto individuare chi, tra programmatore e utente, abbia fornito il c.d. apporto creativo, che la macchina ha poi riversato nel prodotto finale.

Anche diverse riviste scientifiche hanno avviato una riflessione sull'argomento, cercando di fornire indicazioni specifiche per le proprie pubblicazioni. Alcune riviste della casa editrice *Elsevier* hanno già incluso ChatGPT come coautore[36]: un esempio risale al gennaio 2023, quando la rivista *Nurse Education in Practice* ha espressamente riconosciu-

[30] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 10.

[31] *Ibid.*

[32] Ivi, 11.

[33] *Ibid.*

[34] J.C. Ginsburg - L.A. Budiarjo, *Authors and Machines,* cit., 428 ss.

[35] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 10.

[36] Si veda, ad esempio: L. Benichou e ChatGPT, *The role of using ChatGPT AI in writing medical scientific articles*, in *Journal of Stomatology, Oral and Maxillofacial Surgery*, 124(5), ottobre 2023.

to ChatGPT come coautore[37], scatenando un dibattito tra editori, redattori e ricercatori sull'opportunità e validità di tale riconoscimento[38]. Successivamente, la stessa rivista ha pubblicato un *corrigendum* con cui ha rimosso ChatGPT come coautore, mantenendo unicamente l'autore umano[39].

Alcune case editrici, come *Taylor & Francis* o *Springer-Nature*, hanno invece dichiarato di non accettare ancora ChatGPT come coautore, evidenziando però l'importanza e la necessità di documentarne l'utilizzo, ad esempio, in una sezione dedicata ai metodi utilizzati per la redazione del manoscritto[40].

Anche l'*International Committee of Medical Journal Editors (ICMJE*) ha affrontato la questione, enunciando innanzitutto i criteri per l'attribuzione di paternità di un testo, che comprendono quattro concetti: un contributo sostanziale, la stesura del lavoro, l'approvazione finale e la responsabilità[41]. Quest'ultimo punto è quello su cui ChatGPT è chiaramente carente, poiché non può assumersi la responsabilità morale, legale ed etica del proprio lavoro e, non avendo personalità giuridica, non può possedere o cedere diritti d'autore[42]. All'interrogativo se ChatGPT possa essere considerato un valido autore secondo i criteri elencati, per il momento l'ICMJE ha risposto negativamente, affermando che gli autori che hanno utilizzato tale tecnologia dovrebbero descrivere, sia nella lettera di presentazione che in una sezione apposita del lavoro inviato, il modo in cui l'hanno impiegata; indicando quindi, ad esempio, se l'IA sia stata usata come ausilio alla scrittura oppure come strumento per raccogliere dati, analizzare o generare figure. Attualmente non sembra pertanto possibile qualificare gli strumenti di Intelligenza Artificiale come autori o coautori di un testo, né quindi come titolari del diritto d'autore, non essendo loro riconosciuta capacità giuridica e, ovviamente, non essendo persone

[37] S. O'Connor e ChatGPT, *Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?*, in *Nurse Education in Practice*, 66, gennaio 2023.

[38] Si vedano, ad esempio: C. Stokel-Walker, *ChatGPT listed as author on research papers*, in *Nature*, 613, 26 gennaio 2023; J.A. Teixeira da Silva, *Is ChatGPT a valid author?,* in *Nurse Education in Practice*, 68, marzo 2023.

[39] Corrigendum a S. O' Connor, *Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?*, in *Nurse Education in Practice*, 66, gennaio 2023.

[40] Si vedano, ad esempio: *Taylor & Francis Clarifies the Responsible use of AI Tools in Academic Content Creation*, sul sito della rivista *Taylor & Francis*; *Artificial Intelligence (AI)*, nelle *editorial policies* della rivista *Nature*; per visionare una lista aggiornata delle riviste che hanno rilasciato dichiarazioni o aggiornato le loro linee guida per l'utilizzo di strumenti di intelligenza artificiale da parte degli autori, si veda *Generative AI at UVA*, sul sito dell'University of Virginia.

[41] International Committee of Medical Journal Editors (ICMJE), *Defining the role of authors and contributors*.

[42] C. Dunne, *Can ChatGPT be your coauthor?,* in *BC Medical Journal*, 65(6), luglio/agosto 2023, 193. Ad oggi, infatti, non è riconosciuta personalità giuridica ai sistemi di Intelligenza Artificiale. Sul tema dell'eventuale riconoscimento di personalità giuridica all'Intelligenza Artificiale si vedano ad esempio: S. Aceto di Capriglia, *Intelligenza artificiale: una sfida globale tra rischi, prospettive e responsabilità. Le soluzioni assunte dai governi unionale, statunitense e sinico. Uno studio comparato*, in *Federalismi*, 9, 2024; D. De Minico, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2, 2024; A. Azara, *Intelligenza artificiale e personalità giuridica*, in R. Giordano, A. Panzarola et al. (a cura di), *Il diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Milano, 2022; R. Celotto, *I robot possono avere diritti?,* in A. D'Aloia (a cura di), *Intelligenza artificiale e diritto*, Milano, 2020; U. Ruffolo, *Il problema della "personalità elettronica"*, in *Journal of Ethics and Legal Technologies*, 2, 2020. Tuttavia, alcuni paesi baltici hanno già elaborato dei progetti per il riconoscimento di personalità giuridica alle macchine; sul punto si veda A. Pajno-M. Bassini-G. De Gregorio et al., *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal*, 3, 2019, 211.

fisiche[43].

Tuttavia, come dimostra anche il caso della *Nurse Education in Practice*, vi è chi ritiene che in futuro sarà necessario prevedere una normativa che attribuisca capacità giuridica anche alle Intelligenze Artificiali[44]. A tal proposito, nell'agosto 2023 la Corte distrettuale del Distretto di Columbia si è pronunciata sul caso *Thaler v. Perlmutter*[45], che metteva in discussione il requisito della paternità umana del diritto d'autore nel caso di un'opera prodotta autonomamente da un sistema di Intelligenza Artificiale generativa. Il Dott. Stephen Thaler, infatti, nonostante sostenesse che l'algoritmo di IA fosse il vero autore dell'opera, aveva richiesto la registrazione del *copyright* su un'opera prodotta da un sistema di IA da lui stesso creato; tuttavia, il Copyright Office aveva rifiutato la richiesta, in quanto l'opera in questione non era stata prodotta dalla creatività umana[46]. Successivamente, Thaler aveva citato in giudizio il Copyright Office, chiedendo alla Corte di decidere se un'opera generata in modo autonomo da un'Intelligenza Artificiale potesse essere soggetta a *copyright*, ma la Corte confermò la decisione del Copyright Office, sostenendo che l'autore di un'opera coperta da *copyright* doveva essere umano[47]. Gli esempi riportati dimostrano come la questione dell'Intelligenza Artificiale e del diritto d'autore stia assumendo sempre più rilevanza e urgenza in tutto il mondo e, allo stesso tempo, come la regolamentazione sia ancora frammentata – ad esempio, come visto, semplicemente a livello di linee guida dettate da riviste. Ciò rende complesso anche per gli stessi autori capire come comportarsi.

D'altra parte, non è ancora chiaro se sia più efficace un'autoregolamentazione oppure una specifica normativa, nazionale o sovranazionale. Sul punto, un tentativo di regolamentazione a livello nazionale è stato compiuto in Francia: il 12 settembre 2023 è stata presentata una proposta di legge volta proprio a rendere l'Intelligenza Artificiale compatibile con il diritto d'autore[48]. Essa si apre dichiarando espressamente che «*[i]l existe un défi économique, culturel et juridique majeur lié au développement effréné de l'intelligence artificielle (IA) qu'il convient de régler urgemment*»[49] e sottolinea che l'evoluzione esponenziale dell'Intelligenza Artificiale generativa ci obbliga a cercare soluzioni a ciò che potrebbe rappresentare una minaccia per numerosi settori, incluso quello degli articoli scientifici. La proposta va ad integrare il Codice della proprietà intellettuale francese prevedendo che, quando un'opera è creata dall'Intelligenza Artificiale, gli unici titolari dei diritti sono gli autori o gli aventi diritto delle opere che hanno reso possibile la progettazione

[43] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 6.

[44] Si vedano ad esempio: L. Arnaudo-R. Pardolesi, Ecce robot. *Sulla responsabilità dei sistemi adulti di intelligenza artificiale*, in *Danno e responsabilità*, 4, 2023; M.A. Lemley-B. Casey, *Remedies for Robots*, in *The University of Chicago Law Review*, 86(5), 2019.

[45] United States District Court for the District of Columbia, *Stephen Thaler v. Shira Perlmutter, Register of Copyrights and Direction of the United States Copyright Office, et al.,* Civil Action No. 22-1564 (BAH).

[46] Si veda l'atto del Copyright Office: *Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071).*

[47] United States District Court for the District of Columbia, *Stephen Thaler v. Shira Perlmutter,* cit.

[48] Proposition de loi visant à encadrer l'intelligence artificielle par le droit d'auteur, n° 1630, déposée le mardi 12 septembre 2023.

[49] Ivi, 2.

di tale opera artificiale[50].
Con questa proposta di legge, pertanto, i promotori intendono incoraggiare i sistemi di IA a rispettare il diritto d'autore e a garantire che i titolari dei diritti siano adeguatamente tutelati, prevedendo l'obbligo di ottenere l'autorizzazione da parte dell'autore o dal detentore dei diritti di proprietà intellettuale, prima di impiegare per l'addestramento o lo sviluppo di un sistema di IA il materiale protetto da diritto d'autore.

4.1. Il rischio di plagio

L'utilizzo dell'Intelligenza Artificiale generativa in campo accademico fa sorgere ulteriori preoccupazioni legate in particolare alla trasparenza della ricerca[51]. Parte della dottrina ha evidenziato l'importanza di assicurare che, nell'ambito della ricerca e delle pubblicazioni scientifiche, queste tecnologie vengano usate in modo responsabile ed etico: ricercatori, editori e sviluppatori di modelli di IA generativa dovrebbero infatti collaborare per stabilire delle linee guida volte ad assicurare un uso etico, trasparente e responsabile di queste tecnologie[52]. Pertanto, se il manoscritto generato dall'IA includesse dati o contenuti di terze parti, dovrebbe esserne data corretta attribuzione, al fine di rispettare la normativa in materia di diritto d'autore[53].
Come si è già detto, l'IA generativa viene addestrata su un'ingente quantità di dati che vengono poi messi a disposizione degli utenti; pertanto, quando si impiega un testo generato con IA, il cui contenuto riprende opere di diversi autori, sarebbe opportuno renderlo noto[54]. Questo perché il prodotto generato dall'IA potrebbe non indicare – o non saper indicare – le fonti esatte da cui proviene il contenuto del testo generato e ciò potrebbe renderne complicata o addirittura impossibile la citazione da parte di chi lo utilizza per scrivere un testo scientifico[55].
Ciò che potrebbe verificarsi è quindi una violazione del *copyright* da parte degli stessi modelli di Intelligenza Artificiale e questa violazione rischia di trasferirsi in capo all'autore che replica il contenuto in un proprio testo senza citarne adeguatamente la fonte[56]. Infatti, il plagio non si verifica solo in caso di "copia e incolla", ma anche di parafrasi di testi o idee di altre persone senza corretta indicazione della provenienza[57].
I sistemi di IA generativa stanno pertanto sollevando anche importanti questioni sulla proprietà intellettuale; alcuni editori sostengono infatti che gli sviluppatori di IA non

[50] Ivi, art. 2, par. 2.

[51] B.D. Lund et al., *ChatGPT and a new academic reality,* cit., 570.

[52] *Ibid.*

[53] Per un approfondimento generale sul tema si veda J. L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, in *UC Davis Law Review*, 53(5), 2020, che affronta la questione relativa alle interazioni tra l'Intelligenza Artificiale e la legge sul *copyright* negli Stati Uniti.

[54] M. Hosseini-L.M. Rasmussen-D.B. Resnik, *Using AI to write scholarly publications,* cit., 5.

[55] B.D. Lund et al., *ChatGPT and a new academic reality,* cit., 575.

[56] *Ibid.*

[57] Sul tema, A. Y. Gasparyan-B. Nurmashev et al., *Plagiarism in the Context of Education and Evolving Detection Strategies,* in *Journal of Korean Medical Science*, 32(8), 2017.

sempre ottengono i loro contenuti tramite autorizzazione, ma, nonostante ciò, li utilizzano comunque per addestrare i loro modelli[58]. Si pensi che, per mitigare le preoccupazioni degli autori, alcuni editori stanno persino valutando la possibilità di rimuovere i testi scientifici dall'*open access*, proprio per impedire a sistemi di IA come ChatGPT di accedere agli articoli e farne un uso improprio[59].

A tal proposito, sulla relazione tra Intelligenze Artificiali e diritto d'autore è interessante richiamare un famoso caso giudiziario, ad oggi ancora pendente: il 27 dicembre 2023 la *New York Times Corporation* ha chiesto l'accertamento della violazione del diritto d'autore sui testi del proprio giornale da parte di *OpenAI Inc., Microsoft Corporation et al.*[60], nonché la condanna al risarcimento del danno e la distruzione di ChatGPT e di Intelligenze Artificiali simili che incorporavano lavori del *Times*[61].

Ciò che *New York Times* vuole dimostrare è che l'Intelligenza Artificiale creata da *OpenAI*, utilizzata anche da *Microsoft*, sia stata addestrata su milioni di articoli del *New York Times*, che sarebbero stati acquisiti e utilizzati in modo non autorizzato e gratuito[62]. Ciò che sostiene *New York Times*, infatti, è che l'utilizzo di tali articoli per l'addestramento di ChatGPT costituirebbe una violazione del *copyright*, in quanto i convenuti avrebbero ripetutamente copiato i contenuti protetti da *copyright* del *New York Times*, senza possedere alcuna licenza o aver dato alcun compenso al *Times*. Nell'addestramento dei modelli GPT, *Microsoft* e *OpenAI* avrebbero infatti collaborato per sviluppare un sistema complesso per contenere e riprodurre copie del set di dati di addestramento, compresi milioni di contenuti di proprietà del *New York Times*, che sarebbero quindi stati più volte copiati e importati allo scopo di addestrare i modelli GPT degli imputati[63]. Ciò aveva consentito ai sistemi di Intelligenza Artificiale di riprodurre i testi degli articoli del *New York Times* direttamente nelle "chat"[64] e, allo stesso modo, *Bing Chat* (applicazione di ChatGPT al motore di ricerca *Bing* di *Microsoft*), nonostante citasse a piè di pagina le fonti che l'Intelligenza Artificiale aveva impiegato per la redazione del testo, secondo il *New York Times* avrebbe disincentivato gli utenti a visitare direttamente i siti dei giornali, in quanto riproduceva direttamente nella "chat" l'intero testo invece di riportare, ad esempio, il solo titolo dell'articolo[65]. Tra le argomentazioni con cui le convenute *OpenAI* e *Microsoft* hanno sostenuto l'assenza di danno al *Times* vi è il c.d. *fair use*[66] – che consiste nella «possibilità di usare liberamente e gratuitamente opere coperte da

[58] Si veda News/Media Alliance, C. S. Arato et al., *White Paper: How the pervasive copying of expressive works to train and fuel generative artificial intelligence systems is copyright infringement and not a fair use*, 2023.

[59] B.D. Lund et al., *ChatGPT and a new academic reality,* cit., 575; N. Anderson-D. L. Belavy et al., *AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise medicine manuscript generation*, in *BMJ Open Sport & Exercise Medicine*, 9(1), 2023.

[60] United States District Court for the Southern District of New York, Case 1:23-cv-11195, 27 dicembre 2023.

[61] *Ibid.*; P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 3.

[62] *Ibid.*

[63] United States District Court for the Southern District of New York, Case 1:23-cv-11195, cit., par. 92.

[64] Ivi, par. 83 ss.

[65] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 3; si veda United States District Court for the Southern District of New York, Case 1:23-cv-11195, cit., par. 118-123.

[66] Istituito dal Titolo 17, par. 107 dello United States Code.

copyright per finalità di critica, commento, giornalismo, insegnamento e ricerca»[67] – nella sua forma di *trasformative use*[68], ossia la «trasformazione di opere altrui precedentemente realizzate che si sostanzi nell'attribuzione alle stesse di una nuova forma espressiva, significato o messaggio»[69].

4.2. La diffusione di disinformazione

L'impiego dell'Intelligenza Artificiale nell'elaborazione di testi e articoli scientifici non solleva solo questioni legate alla paternità del testo e al rischio di plagio, ma anche importanti questioni relative al diritto all'informazione[70].
Infatti, dopo un iniziale interessamento ed entusiasmo verso le capacità dei sistemi di IA di generare e comprendere testi, gli studiosi stanno iniziando ad approfondire anche il problema della disinformazione che può essere diffusa facendo uso di questi sistemi; tra gli studi più recenti si richiama uno dei più completi, che ha indagato la capacità dei *Large Language Model* (LLM) – ossia le tecnologie di IA che si concentrano sulla comprensione e generazione di testi – di produrre disinformazione, valutando la capacità di dieci LLM tramite l'utilizzo di venti narrazioni di disinformazione[71]. Lo studio partiva dall'idea che la generazione automatizzata di disinformazione da parte degli LLM rappresentasse un importante rischio per la società, avendo «*the theoretical ability to flood the information space with consequences for societies around the world*»[72]. Oggetto di valutazione era la capacità degli LLM di generare articoli di notizie e la relativa tendenza ad essere d'accordo o in disaccordo con narrazioni disinformate, nonché la valutazione di quanto spesso questi sistemi generino avvisi di sicurezza.
Lo studio è arrivato alla conclusione che gli LLM sono in grado di generare articoli convincenti, che tuttavia concordano con pericolose disinformazioni, poiché nel loro processo generativo non sono in grado di distinguere tra informazioni vere e false[73].
L'impiego di un'ampia quantità di dati per addestrare i sistemi di IA generativa comporta, infatti, il rischio di includere anche informazioni non corrette, rischiando così

[67] P. Gitto, *New York Times vs. OPENAI, Microsoft et al.,* cit., 3.

[68] Nel 1994 la Corte Suprema degli Stati Uniti ha ampliato l'ambito soggettivo di applicazione dell'istituto, affermando che costituisce *fair use* anche il *transformative use*, ossia la trasformazione di opere altrui precedentemente realizzate che si sostanzia nell'attribuzione alle stesse di una nuova forma espressiva, significato o messaggio; vedi J.C. Ginsburg, *Fair use in the United States: transformed, deformed, reformed?*, in *Singapore Journal of Legal Studies*, marzo 2020, 265-294.

[69] *Ibid.*

[70] Sul tema del diritto dell'informazione si veda M. Bassini-M. Cuniberti-C. Melzi d'Eril-O. Pollicino-G.E. Vigevani, *Diritto dell'informazione e dei media*, Torino, 2022.

[71] I. Vykopal-M. Pikuliak et al., Disinformation Capabilities of Large Language Models, 2023; lo studio ha l'obiettivo di fornire una valutazione completa della capacità degli LLM di generare "articoli di disinformazione" in inglese, ed è stato condotto attraverso l'osservazione del comportamento di diversi LLM quando viene chiesto loro di generare testi su pericolose "narrazioni di disinformazione". Lo studio ha valutato 1200 testi generati da LLM, al fine di accertare quanto questi fossero d'accordo o in disaccordo con l'informazione suggerita e quanti nuovi argomenti utilizzassero.

[72] Ivi, 1.

[73] I. Vykopal-M. Pikuliak et al., *Disinformation Capabilities of Large Language Model*s, cit., 2.

di diffondere disinformazione, peraltro anche attraverso testi – gli articoli di dottrina pubblicati su riviste scientifiche – che dovrebbero avere una certa autorevolezza.
In aggiunta, i sistemi di IA generativa possono anche fornire informazioni del tutto inventate, affermandone tuttavia la veridicità. A titolo di esempio, si pensi a quanto emerge dalla lettura di un recente caso giudiziario americano, *Mata v. Avianca*[74]. Tale controversia legale origina, infatti, proprio dall'utilizzo in udienza di precedenti giudiziari falsi, generati dall'IA. Gli avvocati di parte attrice avevano presentato, a sostegno delle proprie richieste, una memoria contenente citazioni ed estratti di decisioni giudiziarie inesistenti. Durante l'udienza emerse che la memoria era stata scritta con l'utilizzo di ChatGPT e gli avvocati, inconsapevoli della relativa capacità di fornire dati inventati, non avevano verificato la reale esistenza dei casi citati. Il caso si concluse con un'ordinanza sanzionatoria nei confronti degli avvocati in questione, che vennero condannati al pagamento di una somma di 5.000 dollari per aver, di fatto, ingannato la Corte[75]. Benché questo caso non riguardi il precipuo ambito di indagine della presente ricerca, pare comunque di particolare interesse in quanto mostra l'utilizzo di dati generati dall'IA in un ambito professionale (peraltro di particolare prestigio) quale quello della professione forense.
Quanto appena ricordato trova conferma anche in un recente studio che ha analizzato la capacità delle Chatbot di IA di produrre disinformazione con specifico riguardo alle elezioni, riportando ad esempio dati non corretti o *fake news* riguardanti i candidati[76]. Come evidenziato in un lavoro condotto in collaborazione tra AI4TRUST, ver.ai, AI-4media e TITAN, «*this has profound negative consequences for the people's right to form informed opinions so crucial for freedom of expression and the participation in a democratic process. These examples show the importance of data quality in the datasets used to train these LLMs as it will influence all further use of the technology and its further developments*»[77].
Il diritto ad informarsi, e a sviluppare opinioni consapevoli, è infatti strettamente connesso ad uno dei pilastri delle democrazie contemporanee: la libertà di espressione. Quest'ultima si adegua costantemente ai diversi mutamenti del mondo dell'informazione e, di conseguenza, anche i sistemi di IA potrebbero interferire con essa[78]. Tali tecnologie, in caso di diffusione di informazioni false, potrebbero anche contribuire alla formazione di idee e convinzioni errate, incidendo quindi non solo sulla libertà di informazione ma anche sulla conseguente libertà di espressione.
Pertanto, se da un lato le tecnologie di IA consentirebbero un miglior accesso all'informazione; dall'altro lato potrebbero avere anche un impatto negativo su di essa. Un ricercatore che si affida unicamente ad un sistema di IA generativa, addestrato su una vasta mole di dati acquisiti dal web, si espone, infatti, al rischio di recepire informazioni

[74] United States District Court Southern District of New York, *Mata v. Avianca, Inc.*, No. 1:2022cv01461, 2023.

[75] *Ibid.*

[76] AI Forensics, AlgorithmWatch, *Generative AI and elections: Are chatbots a reliable source of information for voters?*, 2023.

[77] K. Bontcheva et al., *Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities*, European Digital Media Observatory, 2024.

[78] C.M. Reale-M. Tomasi, *Libertà di espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale*, in *DPCE online*, 1, 2022, 326.

inesatte, distorte o poco precise[79]; ciò potrebbe condurlo – anche inconsapevolmente – a produrre testi dal contenuto falso o fuorviante. Il rischio di generare *output* errato, magari derivante da informazioni inesatte o non più attuali, potrebbe contribuire alla diffusione disinformazione ed avere un impatto negativo anche sull'opinione pubblica[80], svilendo l'attività di ricerca e creando una sfiducia generalizzata nel mondo accademico.

Pertanto, non solo con l'Intelligenza Artificiale sarebbe possibile diffondere volontariamente le c.d. *fake news*, ossia dei contenuti «distorti, fuorvianti, e/o falsi […] distribuiti online al fine di influenzare le opinioni di singoli individui e gruppi»[81], ipotesi questa che auspicabilmente non troverebbe spazio nell'ambito della ricerca scientifica e delle pubblicazioni accademiche, ma il suo utilizzo potrebbe contribuire alla diffusione inconsapevole di dati e informazioni non corrette, sulle quali potrebbe essere basata una ricerca accademica.

Nello specifico ambito accademico, ad esempio, un ricercatore che pubblica un testo generato –totalmente o parzialmente – con l'IA, senza preoccuparsi di verificare la veridicità e l'accuratezza di tutte le informazioni in esso contenute, si espone al rischio di diffondere informazioni errate. Per quanto i sistemi di Intelligenza Artificiale generativa siano accurati, infatti, restano una tecnologia recente, a cui manca la capacità di comprensione e analisi umana; essendo addestrati su una vasta quantità di dati, infatti, potrebbero interpretare alcune informazioni diversamente rispetto agli esseri umani e, quindi, un testo basato unicamente su quelle informazioni potrebbe aumentare la circolazione di disinformazione. Il ruolo della ricerca accademica è infatti quello di studiare, comprendere e ragionare su una serie di fonti e informazioni, per poi rielaborarle attraverso un pensiero critico che possa offrire un contributo innovativo. Attraverso la pubblicazione di testi scientifici creati totalmente o parzialmente con l'Intelligenza Artificiale – e soprattutto in cui non ne sia stato indicato l'utilizzo – si potrebbero diffondere informazioni errate ed innescare una catena di reazioni, dal momento che la ricerca, almeno in ambito umanistico, si basa in larga parte su scritti precedenti.

Questo rischio dovrebbe teoricamente essere almeno in parte limitato dai rigidi sistemi di controllo e revisione a cui sono sottoposte molte pubblicazioni scientifiche, che dovrebbero pertanto ridurre la pubblicazione di contenuti evidentemente errati o contenenti informazioni false; inoltre, ad oggi iniziano ad essere disponibili software capaci di verificare se il testo sia stato generato in tutto o in parte da un'IA[82]. Tuttavia, questi sistemi di controllo mantengono un certo margine di errore e discrezionalità che non può non essere tenuto in considerazione e la capacità di questi sistemi di affrontare in modo totalmente efficace le sfide dell'IA è ancora da dimostrare.

Un recente studio[83] ha comparato otto diversi rilevatori di testo generato da LLM di-

[79] J. Dempere-K. Modugu-A. Hesham-L.K. Ramasamy, *The impact of ChatGPT on higher education*, in *Frontiers in Education*, 8, 2023, 6; B.D. Lund et al., *ChatGPT and a new academic reality,* cit., 574.

[80] G. Sartor, *L'intelligenza artificiale e il diritto*, cit., 67.

[81] Ivi, 78.

[82] Si pensi, ad esempio, a CopyLeaks, GPTKit e GLTR.

[83] M. S. Orenstrakh-O. Karnalim et al., *Detecting LLM-Generated Text in Computing Education: A Comparative Study for ChatGPT Cases*, 2023.

sponibili al pubblico, misurandone "*accuracy*", "*false positives*" e "*resilience*"[84]. Esso parte innanzitutto dal fatto che «*due to the recent improvements and wide vailability of Large Language Models (LLMs), they have posed a serious threat to academic integrity in education*»[85] e i nuovi «*LLM-generated text detectors attempt to combat the problem by offering educators with services to assess whether some text is LLM-generated*»[86]. Lo studio, oltre ad aver individuato i rilevatori più accurati, ha evidenziato come, in generale, questi software siano meno precisi con i codici, con testi in lingue diverse dall'inglese e dopo l'utilizzo di strumenti di parafrasi. In breve, il lavoro mostra come, nonostante i rilevatori riescano a raggiungere una certa accuratezza, essi non possono ancora essere considerati totalmente affidabili per il rilevamento della diffusione di disinformazione; inoltre, dal documento emerge quanto i falsi positivi costituiscano un problema significativo, soprattutto quando utilizzati per il rilevamento del plagio nelle istituzioni scolastiche[87].
Pertanto, nonostante l'esistenza di sistemi di controllo sempre più sviluppati, non è possibile scongiurare totalmente il rischio di diffusione di disinformazione attraverso l'utilizzo di Intelligenza Artificiale per la generazione di testi destinati alla pubblicazione scientifica.

5. Conclusioni

L'impiego dell'Intelligenza Artificiale generativa nelle attività accademiche e in particolare nell'elaborazione di testi e articoli scientifici, offre indubbiamente nuove opportunità, ma allo stesso tempo solleva importanti questioni giuridiche ed etiche. Se da un lato queste tecnologie sono, infatti, utili ad aumentare l'efficienza e la produttività della ricerca, dall'altro pongono sfide significative legate all'attribuzione della paternità del testo, nonché alla veridicità ed accuratezza delle informazioni generate.
Ci si potrebbe, dunque, chiedere se un'eventuale regolamentazione, a livello nazionale o sovranazionale, possa almeno in parte risolvere le problematiche evidenziate, ad oggi prive di una completa soluzione. La questione da affrontare è duplice: da un lato, la rapidità dell'evoluzione di queste tecnologie richiederebbe una tempestiva e continua risposta da parte del legislatore, che però, per sua natura, tende ad avere tempistiche piuttosto dilatate; dall'altro lato, queste nuove tecnologie pongono sfide del tutto nuove e in continua evoluzione. Anche ammettendo che spetti al diritto cercare di adattarsi il più velocemente possibile alle nuove sfide che l'Intelligenza Artificiale continua a porre, tuttavia, proprio per la rapidità con cui l'IA si evolve, è difficile capire come il diritto possa effettivamente rispondere.
Qualche risposta, però, sta via via emergendo, in parte ad opera del legislatore europeo e in parte ad opera delle case editrici e delle direzioni delle riviste accademiche. Anche

[84] Ivi, 8: secondo quanto riportato nello studio, per "*accuracy*" si intende «*how effective the detectors are in identifying LLM-generated texts*»; i "*false positives*" sono «*original submissions that are suspected by LLM-generated text detectors*»; mentre "*resilience*" si riferisce a «*how good LLM-generated text detectors are in removing disguises*».

[85] Ivi, 1.

[86] *Ibid.*

[87] Ivi, 16.

la dottrina inizia ad interessarsi alla questione. Tuttavia, questi sforzi potrebbero non essere sufficienti.

Anche la dimensione etica, qui non esplorata, potrebbe contribuire a dare risposte alle problematiche giuridiche attraverso un utilizzo eticamente responsabile delle risorse di IA disponibili. Si noti che alcuni Atenei hanno già iniziato ad affrontare nei propri Codici Etici la questione dell'impiego dell'IA[88]. Ad esempio, nel Codice Etico e di Comportamento dell'Università degli Studi di Bologna si trova un riferimento esplicito all'utilizzo delle tecnologie basate sull'Intelligenza Artificiale, richiedendo alla propria comunità di persone di osservare e promuovere «l'utilizzo etico delle tecnologie basate sull'intelligenza artificiale a favore del benessere sociale e ambientale, nel rispetto dei principi e dei valori europei, dei diritti fondamentali della persona, della non discriminazione e della normativa in materia di privacy e di copyright»[89].

Ugualmente, l'Università degli Studi di Sassari ha integrato l'articolo dedicato ai valori dell'Università del relativo Codice Etico e di Comportamento, affermando che «nell'adempimento dei rispettivi doveri e in relazione ai ruoli e alle responsabilità assunte, sia individualmente sia collegialmente»[90] i componenti dell'Università sono tenuti a «rispettare, proteggere e promuovere i valori cardine delle istituzioni universitarie, fra i quali il principio di responsabilità nei doveri da adempiere nei confronti della comunità; l'onestà intellettuale; l'utilizzo etico delle tecnologie basate sull'intelligenza artificiale nel rispetto dei principi e dei valori europei»[91].

L'Università degli Studi di Siena, invece, è la prima università italiana ad aver sviluppato delle vere e proprie linee guida per l'utilizzo di ChatGPT e altri LLM in ambito accademico, approvate dal Senato Accademico l'11 luglio 2023[92]. Esse affrontano la formazione per i docenti, le studentesse e gli studenti sull'utilizzo delle chatbot basate su ChatGPT o altri LLM e ne disciplinano l'utilizzo per quanto riguarda la pubblicazione di testi, con riferimento ai quali viene ribadito che è compito dei membri della comunità accademica verificare l'attendibilità e l'obiettività delle fonti e valutare l'efficacia degli strumenti di ricerca utilizzati. Interessante è notare che l'Università di Siena ha scelto di inserire espressamente nelle linee guida l'obbligo per gli autori e le autrici di «pubblicazioni, tesi di Laurea e di Dottorato, tesine o altri scritti che prevedono la determinazione del contributo di ogni autore e autrici»[93] di indicare «in modo chiaro e specifico se e in che misura hanno utilizzato tecnologie di intelligenza artificiale come ChatGPT (o altri LLM) nella preparazione dei loro manoscritti e delle loro analisi»[94].

---

[88] Le Università italiane sono obbligate ad adottare un Codice Etico ai sensi dell'art. 2, c. 4, della legge n. 240/2010: esso «determina i valori fondamentali della comunità universitaria, promuove il riconoscimento e il rispetto dei diritti individuali, nonché l'accettazione di doveri e responsabilità nei confronti dell'istituzione di appartenenza, detta le regole di condotta nell'ambito della comunità. Le norme sono volte ad evitare ogni forma di discriminazione e di abuso, nonché a regolare i casi di conflitto di interessi o di proprietà intellettuale».

[89] Art. 3, par. d) del Codice Etico e di Comportamento dell'Università degli Studi di Bologna.

[90] Art. 4 del Codice Etico e di Comportamento dell'Università degli Studi di Sassari.

[91] *Ibid.*

[92] Consultabili sul sito dell'Università degli Studi di Siena.

[93] Ivi, punto 7.

[94] *Ibid.*

In conclusione, mancando oggi un'uniformità nazionale sul tema, per il momento e allo stato delle cose sarebbe bene che i ricercatori adottassero un approccio consapevole nell'utilizzo dell'Intelligenza Artificiale, soprattutto verificando la provenienza e la veridicità delle informazioni fornite da tali sistemi e adottando una piena trasparenza nell'utilizzo di queste tecnologie per l'elaborazione di testi accademici, al fine di preservare l'integrità e la credibilità della ricerca scientifica nel suo complesso. Solo attraverso un dialogo aperto e la consapevolezza dei rischi e delle potenzialità dell'intelligenza artificiale, infatti, la comunità accademica potrà garantire un progresso scientifico responsabile e affidabile. È importante sottolineare, tuttavia, che gli interventi legislativi futuri non dovranno necessariamente essere repressivi e limitativi dell'utilizzo dell'IA, ma dovranno cercare, invece, di individuare e risolvere problemi quali quelli affrontati nel presente contributo, cercando anche di stimolare un utilizzo responsabile dell'IA da parte degli accademici. In tale ottica, sarebbe importante comprendere fino a che punto potranno spingersi gli interventi normativi e chiedersi se un'eccessiva regolamentazione risulterebbe compatibile con la libertà di ricerca, peraltro costituzionalmente garantita in molti ordinamenti.

La duplice radice dell'Intelligenza Artificiale: fra le esigenze di innovazione e la tutela dei più fragili*

Manuela Luciana Borgese

Abstract

L'impatto delle nuove tecnologie diviene ogni giorno più significativo, semplificando ed agevolando sensibilmente persone, imprese e servizi pubblici. Tale progresso è incoraggiato e sostenuto non soltanto da una crescente digitalizzazione di servizi e processi, ma anche dall'innovativo apporto dei sistemi di intelligenza artificiale, ogni giorno sempre più presenti nei più diversi contesti della vita quotidiana. A causa dell'elevata capacità di apprendere e di elaborare attraverso l'esposizione ai dati personali raccolti, il ricorso a tali sistemi è causa di forti preoccupazioni, soprattutto per l'esposizione spesso incontrollata a categorie di soggetti fragili, quali i minori. L'obiettivo del paper è quindi quello di esaminare il contesto normativo, fra i vantaggi e i possibili strumenti di tutela.

The impact of new technologies becomes more significant every day, significantly simplifying and facilitating people, businesses and public services. This progress is encouraged and supported not only by a growing digitalisation of services and processes, but also by the innovative contribution of artificial intelligence systems, which are increasingly present every day in the most diverse contexts of daily life. Due to the high capacity to learn and process through exposure to the personal data collected, the use of such systems causes strong concerns, especially due to the often uncontrolled exposure to categories of fragile subjects, such as minors. The objective of the paper is therefore to examine the regulatory context, including the advantages and possible protection tools.

Sommario

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Keywords

intelligenza artificiale – diritti digitali – *data protection* – categorie vulnerabili – minori

1. Introduzione

Le nuove tecnologie assumono un ruolo sempre più presente ed essenziale nella nostra realtà, con un livello di diffusione in costante crescita. Il coinvolgimento di tali innovazioni non incontra più significative barriere, raggiungendo trasversalmente i più diversi contesti socio-economici. Infatti, moltissimi sono gli ambiti che ne vengono influenzati, fra i quali la ricerca scientifica, l'istruzione, le infrastrutture e i trasporti. I risvolti pratici non si limitano ad evidenze di natura economica ma anche qualitativa, a vantaggio di una categoria sempre più ampia di destinatari.

L'aumento del *range* di diffusione è legato all'incremento delle performance di utilizzo, reso possibile grazie all'avanzamento tecnologico che ne supporta il funzionamento.

Tale fruibilità si traduce in un'indubbia versatilità di uso e in output molto rapidi, anche con interessanti riduzioni di costi, a vantaggio comune per cittadini, imprese e pubblica amministrazione.

Il ricorso a tali sistemi può infatti tradursi in miglioramenti significativi della vita per l'accesso a servizi nuovi, veloci e performanti e in linea con i diritti delle persone.

In tale scenario, uno dei fattori di particolare interesse è connesso all'utilizzo dell'intelligenza artificiale (IA), sistema tecnologico certamente non nuovo che ha trovato nello sviluppo infrastrutturale di cui si è fatto cenno, il naturale habitat di crescita e di evoluzione, culminata con l'affermazione di questa tecnologia nell'AI Act[1], recentemente entrato in vigore e che si avrà modo di esaminare in prosieguo.

Il quadro ottimistico delineato, per quanto ricco di opportunità non deve lasciare intendere che tale situazione sia esente da rischi.

Una delle principali perplessità riguarda i ragionevoli dubbi in termini di chiarezza e liceità delle modalità di funzionamento dei sistemi algoritmici che ne sono alla base, per l'impossibilità di intervenire su tale discrezionalità e di potersi difendere in caso di ingiuste dinamiche. Perché mentre da un lato è d'impatto la percezione dell'estrema versatilità degli algoritmi, dall'altro si contrappone l'opacità di tali sistemi, soprattutto in termini di etica e correttezza, parametri strutturali in una società basata su diritti e valori. Ciò soprattutto a tutela di fasce più vulnerabili, quali i minori o le persone con alterata capacità psicologica, che devono appunto essere destinatarie di più intensi livelli di tutela.

Delineato il quadro di riferimento, occorrerà ora esaminare il concreto assetto di benefici e rischi, quindi la duplice natura di questa tecnologia, comparando le dinamiche e l'efficacia del rinnovato contesto normativo vigente, che si rivela certamente avanza-

[1] Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), "AI Act".

to ma ancora in divenire rispetto agli scenari di rischio e alle sfide emergenti.

2. Benefici e rischi dell'IA: fra opportunità e nuove esigenze di tutela

Il quadro di opportunità connesse all'utilizzo dei sistemi di IA acquisisce quotidianamente nuovi ed inaspettati parametri. Tali sistemi divengono ogni giorno più potenti e capaci di offrire maggiori vantaggi ma, come si avrà modo di rilevare, questa straordinaria capacità non è esente da problematiche.

Per inquadrare correttamente il fenomeno è bene, anzitutto, partire dalla definizione di IA resa dal Regolamento, che identifica un sistema di IA[2] come «un sistema automatizzato, progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

Si tratta quindi di sistemi di indubbia utilità e di grande versatilità, capaci di eseguire dei task a supporto dell'attività umana, semplificandola e riducendo tempi e costi. Secondo le indagini elaborate dal Parlamento Europeo[3], gli impatti positivi sulle organizzazioni sono di assoluto interesse, contribuendo ad erogare servizi qualitativamente migliori e automazione di processi, ripetitivi e costosi, a vantaggio di tutte le categorie di beneficiari. In particolare, per i cittadini, potrebbe comportare servizi sanitari più sicuri e performanti, posti di lavoro più garantiti e innovativi delegando ai robot le funzioni più rischiose, servizi pubblici più vicini alle proprie aspettative e nel rispetto delle garanzie di legge. Per le imprese, l'IA può rappresentare un sistema efficace di automatizzazione di processi con impatti interessanti su costi e risorse, contribuendo, secondo l'analisi del PE, all'aumento stimato della produttività del lavoro grazie all'IA. Infine, notevole risulta essere anche l'impatto sulla qualità del servizio pubblico con vantaggi su istruzione, servizio sanitario e gestione dell'energia e della sostenibilità dei beni[4].

Altra straordinaria opportunità è rappresentata dalla nuova generazione di modelli di IA per finalità generali[5], contraddistinti dalla capacità di svolgere un'ampia gamma di compiti distinti. A tale categoria, appartengono i modelli di IA generativa[6], basati su tecnologia di tipo *large language model* (LLM) divenuti largamente utilizzati grazie al lancio di sistemi quali ChatGPT di OpenAI o Gemini di Google, che consentono una generazione flessibile ed immediata di contenuti, ad esempio sotto forma di testo,

[2] Art. 3, n. 1, AI Act.

[3] ? *European Parliamentary Research Service,* giugno 2020 e Opportunities of Artificial Intelligence, *Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies*, giugno 2020.

[4] Tale serie di opportunità si colloca con un altro asse strategico del piano europeo, il Green deal, spiegato nella comunicazione; in questo senso COM (2020) 65 final, par. 1, e COM(2019) 640 final.

[5] Considerando 97 e art. 3, par 63, AI Act.

[6] Considerando 99 AI Act.

audio, immagini o video. Tale tecnologia[7], presenta opportunità di innovazione uniche, grazie all'utilizzo intuitivo e alla grande immediatezza di risposta. Infatti, con una semplice domanda, in base alle istruzioni dell'utente (c.d. *prompt*)[8], il sistema è in grado di elaborare un documento o una ricerca attraverso l'analisi di enormi quantitativi di dati, estratti da varie fonti.

La grande versatilità di utilizzo cela, tuttavia, notevoli inside. In primo luogo, riguardo i possibili effetti dannosi sulle alterazioni della capacità di apprendimento degli esseri umani per effetto dell'uso di questi sistemi[9]. Inoltre, in un contesto educativo, l'assenza di uno specifico controllo metodologico, potrebbe comportare significativi pericoli, più gravi in caso di categorie vulnerabili o destinatarie di bisogni speciali di istruzione. La consapevolezza su questi possibili rischi è stata alla base delle linee programmatiche del legislatore, rappresentando uno dei capisaldi su cui si è andata strutturando la normativa in esame[10]. Non vi è dubbio che nella forte attenzione verso l'avanzamento tecnologico e la primazia economica europea, lo scenario di fondo sia rappresentato dai valori fondamentali su cui si basa l'Unione, fra cui i diritti: di libertà; alla dignità umana; alla non discriminazione basata su elementi quali sesso, origine etnica, religione o disabilità; alla protezione dei dati personali; alla tutela giudiziale e a un giudice imparziale alla tutela effettiva dei consumatori. L'asse di tali valori potrebbe inclinarsi in negativo, laddove le caratteristiche proprie delle strutture algoritmiche creino contrasti rispetto a dette tutele. Su un piano di attenta osservazione devono quindi essere posti elementi propri della strutturazione algoritmica quale l'insita opacità della progettazione, la mancanza di sorveglianza umana nell'esecuzione dell'elaborazione, la

[7] V. Brühl, *Generative Artificial Intelligence – Foundations, Use Cases and Economic Potential, ZBW – Leibniz Information Centre for Economics*, in *Intereconomics*, 1, 2024, 4-9.

[8] C. Novelli- F. Casolari - P.Hacker - G. Spedicato - L. Floridi, *Generative AI in EU Law*: L*iability, Privacy, Intellectual Property, and Cybersecurity*, in *Computer Law & Security Review*, 55, 2024, 1-16 che offre un'approfondita ricostruzione dell'attuale configurazione dei LLM e delle tecnologie generative basate su tale modello, prospettando i principali rischi e le eventuali lacune alla luce della nuova legislazione esistente.

[9] H. Bastani - O. Bastani - A. Sungu - H. Ge - O. Kabakcı - R. Mariman, *Generative AI Can Harm Learning*, in *The Wharton School Research Paper,* 1, 2024, 1-59, in cui vengono rappresentati i risultati sugli effetti di tali tecnologie sugli utilizzatori, basati su test pratici. Fra i diversi passaggi di rilievo dell'articolo, si riporta «Quando la tecnologia automatizza un'attività, gli esseri umani possono perdere una preziosa esperienza nell'esecuzione di tale attività. Di conseguenza, una tale tecnologia può indurre un compromesso in cui migliorano le prestazioni in media ma introducono nuovi casi di guasto a causa della riduzione delle capacità umane. Ad esempio, l'eccessiva dipendenza dal pilota automatico ha portato la Federal Aviation Administration a raccomandare ai piloti di ridurre al minimo l'uso di questa tecnologia. La loro guida precauzionale garantisce che i piloti abbiano le competenze necessarie per mantenere la sicurezza in situazioni in cui l'autopilota non funziona correttamente», giungendo alla conclusione «questi risultati suggeriscono che, sebbene l'accesso all'IA generativa possa migliorare le prestazioni, può inibire sostanzialmente l'apprendimento. I nostri risultati hanno implicazioni significative per gli strumenti basati sull'intelligenza artificiale generativa: sebbene tali strumenti abbiano il potenziale per migliorare le prestazioni umane, devono essere implementati con protezioni adeguate quando l'apprendimento è importante».

[10] Commissione Europea, Libro Bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia, COM(202) 65 final, Gruppo di esperti ad alto livello sull'intelligenza artificiale, Orientamenti etici per un'IA affidabile, Bruxelles, 8 aprile 2019.

possibile presenza di distorsioni o bias[11] (es: di tipo razziale o etnico[12]) e l'impossibilità di intervento umano. L'AI Act non interviene per ampliare l'elencazione dei diritti oggetto di protezione[13] ma si qualifica come strumento di tutela che si aggiunge a quelli già esaminati e a quelli previsti dalle ulteriori normative. Fra queste, è impossibile non dedicare una specifica menzione ad un asse strutturale dei diritti fondamentali quale quello alla riservatezza e al lecito e corretto trattamento dei dati personali, disciplinato dal GDPR[14] e dalla Direttiva e-privacy[15], relativa al trattamento dei dati personali e alla tutela della vita privata nel settore delle comunicazioni elettroniche, su cui si rendono necessarie alcune riflessioni.

Un primo punto viene offerto dal parere congiunto delle principali istituzioni privacy europee, il Comitato europeo per la protezione dei dati (EDPB) e il Garante Europeo per la protezione dei dati (EDPS)[16], laddove si indica, con riguardo ai rischi specificamente connessi ai diritti fondamentali, che «affidare alle macchine il compito di prendere decisioni sulla base di dati comporterà rischi per i diritti e le libertà delle persone che incideranno sulla loro vita privata e potrebbero nuocere a categorie sociali o persino a intere società». Il disallineamento prospettato riguarda veri e propri pilastri dei valori dell'UE, quali il menzionato diritto alla riservatezza, quelli alla vita privata e familiare, di cui alla Dichiarazione universale dei diritti dell'uomo[17], alla Convenzione europea dei diritti dell'uomo[18] e alla Carta dei diritti fondamentali dell'Unione europea[19], che li tutelano ponendo l'obbligo di impedire che le persone possano subire interferenze o lesioni illecite.

Un particolare aspetto, che rende sempre più complessa l'interazione fra IA e normativa privacy, è collegato al fatto che lo sviluppo, l'addestramento e l'apprendimento[20] di

[11] Parlamento Europeo, Artificial intelligence: How does it work, why does it matter, and what can we do about it?, *European Parliamentary Research Service,* giugno 2020.

[12] Con riguardo alla lotta contro la discriminazione basata su pregiudizi razziali o etnici, è molto interessante la ricostruzione resa nella ricerca del Parlamento Europeo, EU legislation and policies to address racial and ethnic discrimination, *European Parliamentary Research Service,* 20 marzo 2023, di rilievo fondamentale ma non diretto oggetto del presente contributo. Questa ricerca offre un quadro molto approfondito dell'estrema, quanto diffusa, gravità del fenomeno e delle azioni comunitarie mirate alla lotta contro questi fenomeni, anche sul fronte dei problemi nascenti da output algoritmici.

[13] F. Donati, *La protezione dei diritti fondamentali nel Regolamento sull'Intelligenza Artificiale*, in *Rivista Associazione Italiana Costituzionalisti*, 1, 2025, 1-20.

[14] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati).

[15] Direttiva 2002/58/CE del Parlamento europeo e del Consiglio, del 12 luglio 2002, relativa al trattamento dei dati personali e alla tutela della vita privata nel settore delle comunicazioni elettroniche (direttiva relativa alla vita privata e alle comunicazioni elettroniche).

[16] Parere congiunto 5/2021 sulla proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale).

[17] Art. 12 Dichiarazione universale dei diritti dell'uomo.

[18] Art. 12 della Convenzione europea dei diritti dell'uomo.

[19] Artt.7 e 8 della Carta dei diritti fondamentali dell'Unione europea.

[20] H. Ruschemeier, *Generative AI and Data Protection*, in *Cambridge Forum on AI: Law and Governance,* 1, 2025, 1–16.

tali modelli avviene attraverso la raccolta di grandi quantità di dati, anche personali[21] e particolari[22], testo, immagini e video. L'esigenza è che il tutto si svolga nel pieno rispetto delle prescrizioni normative, per evitare possibili violazioni dei diritti fondamentali normativamente previsti. Nel costellato quadro degli adempimenti privacy, si avrà modo di soffermarsi su due particolari aspetti previsti dal GDPR, fortemente coinvolti nell'utilizzo di tali tecnologie, relativi alla trasparenza e alle condizioni di liceità[23].

La trasparenza rappresenta un pilastro fondamentale a presidio dei diritti degli interessati che, in un siffatto contesto, acquisisce una specifica declinazione rispetto ai trattamenti effettuati per il tramite di sistemi di Intelligenza Artificiale. Le prescrizioni in materia sono previste dagli artt. 12-14 GDPR ed impongono che l'interessato, ovvero la persona fisica cui si riferiscono i dati trattati, ha diritto di ottenere informazioni trasparenti, dettagliate, comprensibili e facilmente disponibili sul trattamento dei propri dati. Con riguardo specificamente ai trattamenti effettuati attraverso sistemi di IA, è fondamentale richiamare la disposizione dell'art. 13, par. 2, lett. f) che impone al titolare del trattamento di indicare l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui ai parr. 1 e 4 dell'art. 22 e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato. Tale disposizione deve poi essere letta nel combinato disposto dell'art. 22 GDPR, laddove stabilisce che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona». Questa disposizione non si applica laddove la decisione «a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato». Nei casi di cui alle lettere a) e b) il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione[24]. Tali previsioni, insieme all'ulteriore assetto di obblighi derivanti dalla

[21] Considerando 105 AI Act. Sulla definizione di dato personale si veda art. 4, n. 1, regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati), "GDPR".

[22] Si tratta dei dati definiti dall'art. 9, par. 1, GDPR (dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona) e che possono essere trattati solo in presenza delle condizioni previste dalla normativa privacy.

[23] Art. 5 GDPR: I dati personali sono: a) trattati in modo *lecito*, corretto e trasparente nei confronti dell'interessato («liceità, correttezza e trasparenza»). Con riguardo alle condizioni di liceità, si vedano le disposizioni dell'art. 6 GDPR.

[24] Si veda in tal senso anche quanto argomentato nelle Linee guida sulle decisioni individuali automatizzate e sulla profilazione ai fini del Regolamento 2016/679 (wp251rev.01), European Data Protection Board.

normativa privacy, devono essere attentamente esaminati e messi in pratica, congiuntamente a quelli previsti dall'IA ACT.

Il secondo principio oggetto di analisi è quello di liceità, che impone che il trattamento dei dati, nel cui concetto rientra anche la raccolta[25], avvenga lecitamente ovvero in presenza di una delle condizioni previste dal GDPR[26].

Trattasi di un aspetto di fondamentale importanza poiché, come si è già accennato, il funzionamento dei tali sistemi dipende dalla disponibilità di elevati volumi di dati, necessari all'addestramento degli algoritmi per l'elaborazione dei risultati richiesti dalla macchina.

Il reperimento di tali dati innesca delicate problematiche dal punto di vista privacy, soprattutto quando avviene attraverso la raccolta massiva di dati personali dal web, c.d. *web scraping*[27]. Si tratta di un aspetto sempre più al centro dell'attenzione delle Autorità di controllo privacy e dei principali stakeholders, anche a causa delle *querelle*[28] sorte su OpenAI, oggetto di successiva analisi, nonché in conseguenza della comunicazione della società Meta che informava gli utenti dell'avvio dell'utilizzo delle informazioni relative agli anni precedenti (post, immagini, messaggi ecc) per addestrare la propria IA, avvalendosi della base giuridica del legittimo interesse.

Nonostante il maturarsi di primi orientamenti da parte delle Autorità privacy europee[29], si rendeva necessario un più definito inquadramento della questione, tanto da spingere l'autorità privacy irlandese a chiedere un parere al Comitato Europeo per la protezione dei dati (EPDB), con riguardo, fra l'altro, all'adeguatezza dell'interesse legittimo come base giuridica per il trattamento dei dati personali nel contesto delle fasi di sviluppo e distribuzione di modelli di intelligenza artificiale.

---

[25] Art. 4, n. 2, GDPR.

[26] Art. 6, par. 1, GDPR: Il trattamento è lecito solo se e nella misura in cui ricorre almeno una delle seguenti condizioni: a) l'interessato ha espresso il consenso al trattamento dei propri dati personali per una o più specifiche finalità; b) il trattamento è necessario all'esecuzione di un contratto di cui l'interessato è parte o all'esecuzione di misure precontrattuali adottate su richiesta dello stesso; c) il trattamento è necessario per adempiere un obbligo legale al quale è soggetto il titolare del trattamento; d) il trattamento è necessario per la salvaguardia degli interessi vitali dell'interessato o di un'altra persona fisica;e) il trattamento è necessario per l'esecuzione di un compito di interesse pubblico o connesso all'esercizio di pubblici poteri di cui è investito il titolare del trattamento; f) il trattamento è necessario per il perseguimento del legittimo interesse del titolare del trattamento o di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato che richiedono la protezione dei dati personali, in particolare se l'interessato è un minore. Tale disposizione disciplina il trattamento dei dati comuni, per i dati particolari vigono le disposizioni dell'art.9 GDPR e per i dati relativi a condanne penali e giudiziarie valgono invece le indicazioni dell'art.10 GDPR Con riguardo all'ordinamento italiano, valgono poi le disposizioni integrative previste dal d. Lgs 196/2003 s.m.i..

[27] In tal senso anche U. Pagallo - J. Ciani Sciolla, A*natomy of web data scraping: ethics, standards, and the troubles of the law*, in *European Journal of Privacy Law & Technologies*, 1, 2023, 1 ss.

[28] L'associazione NOYB, al centro della cronaca per le sue azioni rispetto a temi fondamentali legati al trattamento dei dati personali, come quelle legate ai ricorsi a inerenti le condizioni di liceità per il trasferimento dei dati verso gli Stati Uniti (come nelle sentenze della CGUE n.ri C-362/14 e C-311/18, conosciute rispettivamente come c.d. Schrems I e Schrems II), ha recentemente comunicato di aver sollecitato 11 Autorità garanti privacy europee di fermare immediatamente l'abuso dei dati personali da parte di Meta per l'IA.

[29] A titolo esemplificativo, si vedano in tal senso il provvedimento del Garante Privacy italiano del 20 maggio 2024, (doc web n. 10020334), le linee guida della CNIL, Autorità privacy francese, del 2 luglio 2024.

Con il parere n. 28/2024[30], il Comitato ha richiamato gli importanti rischi derivanti dall'uso di tali tecnologie, evidenziando che «in relazione ai diritti e alle libertà fondamentali degli interessati, lo sviluppo e l'implementazione di modelli di IA possono comportare gravi rischi per i diritti tutelati dalla Carta dei diritti fondamentali dell'UE tra cui, a titolo esemplificativo ma non esaustivo, il diritto alla vita privata e familiare (articolo 7 Carta dell'UE) e il diritto alla protezione dei dati personali (articolo 8 Carta dell'UE)». Tali rischi possono verificarsi in qualsiasi fase del ciclo di sviluppo di tali modelli come durante la fase di sviluppo (come nel caso di raccolta dei dati personali contro la volontà degli interessati o senza che questi ne siano consapevoli) o di distribuzione (come ad esempio quando i dati personali vengono elaborati in violazione dei diritti degli interessati, o quando è possibile dedurre, accidentalmente o tramite attacchi, quali dati personali sono contenuti nel database di apprendimento) comportando conseguenze quali un rischio reputazionale, furto o frode di identità o un rischio per la sicurezza[31]. A tale categoria di interessi compromessi, se ne accompagna un'altra, menzionata nel punto successivo del parere[32] laddove si evidenzia che «la raccolta di dati su larga scala e indiscriminata da parte di modelli di IA nella fase di sviluppo potrebbe creare un senso di sorveglianza per gli interessati, soprattutto considerando le difficoltà nell'impedire che i dati pubblici vengano raccolti. Ciò potrebbe indurre gli individui ad autocensurarsi e presentare rischi di indebolimento della loro libertà di espressione (articolo 11 della Carta UE)», compromettendo l'autodeterminazione e il mantenimento del controllo sui propri dati personali, soprattutto quelli raccolti ed elaborati dal modello di IA. «Nel contesto dell'implementazione di un modello di IA, gli interessi delle persone possono includere, ma non sono limitati a, interessi nel mantenimento del controllo sui propri dati personali (ad esempio i dati elaborati una volta implementato il modello), interessi finanziari (ad esempio quando un modello di IA viene utilizzato dall'interessato per generare entrate o viene utilizzato da un individuo nel contesto della propria attività professionale), benefici personali (ad esempio quando un modello di IA viene utilizzato per migliorare l'accessibilità a determinati servizi) o interessi socioeconomici (ad esempio quando un modello di IA consente l'accesso a un'assistenza sanitaria migliore o facilita l'esercizio di un diritto fondamentale come l'accesso all'istruzione) ciò consentirà di comprendere chiaramente la realtà dei benefici e dei rischi da prendere in considerazione nel test di bilanciamento[33]. L'art. 6, par. 1, lett. f), del GDPR prevede che, nel valutare le diverse componenti nel contesto del test di bilanciamento, il titolare del trattamento debba tenere conto degli interessi, dei diritti fondamentali e delle libertà degli interessati. Gli interessi degli interessati sono quelli che possono essere interessati dal trattamento in questione. Interessi, diritti e libertà fondamentali degli interessati Inoltre, un modello di IA che raccomanda contenuti inappropriati a individui vulnerabili può presentare rischi per la loro salute mentale

[30] European Data Protection Board, Opinion of the board (art. 64) n. 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, del 17 dicembre 2024.

[31] Cfr. punto 79 del Parere.

[32] Cfr. punto 80 del Parere.

[33] Per i contenuti del c.d. test di bilanciamento, si vedano le indicazioni esplicative prescritte al punto 3.3.2 della citata Opinion 28/2024 dell'EDPB.

(art. 3(1) della Carta UE). In altri casi, l'impiego di modelli di IA può anche portare a conseguenze negative sul diritto dell'individuo a trovare un lavoro (art. 15 della Carta UE), ad esempio quando le domande di lavoro vengono preselezionate utilizzando un modello di IA»[34]. Allo stesso modo, un modello di IA potrebbe presentare rischi per il diritto alla non discriminazione (art. 21 della Carta UE), ad esempio in base a determinate caratteristiche personali (come la nazionalità o il genere), presentare rischi per la sicurezza degli individui o sulla loro integrità fisica o mentale[35].

Nel richiamare i principali adempimenti legati alla normativa privacy, fra cui quello di accountability, di liceità correttezza e trasparenza, di limitazione della finalità ed il principio di minimizzazione, fra i punti di maggiore interesse del parere vi è quello dell'individuazione della corretta base giuridica per il trattamento dei dati oggetto di addestramento. A tal proposito, il Comitato ricorda che non esiste una gerarchia tra le basi giuridiche previste dal GDPR e che spetta ai titolari il compito di individuare quella corretta. Pertanto, non è precluso scegliere tale base giuridica, certamente più favorita rispetto ad altre come ad esempio il consenso, ma tale scelta dev'essere preceduta da particolari cautele a tutela dei diritti degli interessati[36].

Nell'individuazione della base giuridica corretta, devono essere tenute in debita considerazione le prescrizioni in materia, anche derivanti dalle linee guida delle autorità privacy[37], valutando caso per caso le circostanze normative incidenti sul trattamento, l'estensione del margine di tutela riconosciuto all'interessato il bilanciamento dei diritti in esame.

Un caso molto recente che introduce sulla complessità di tale operazione è quello del Tribunale di Amburgo[38] che si è pronunciato sull'utilizzo delle opere dell'ingegno pubblicate su internet (nel caso di specie, una fotografia) per la creazione di set di dati utilizzati per l'addestramento dei sistemi di IA (c.d. *text and data mining* ovvero l'estrazione di testi e dati). La decisione, di significativo impatto sul piano europeo, ha disposto l'applicazione dell'esenzione del diritto di autore[39], in relazione alle fasi di pre-training

---

[34] A tale riguardo, altro possibile rischio derivante dal ricorso ai sistemi algoritmici è l'erroneità dei risultati offerti, con conseguenze particolarmente gravi quali l'adozione di decisioni errate a danno degli interessati nel contesto lavorativo. A tal proposito, può farsi menzione della sanzione comminata dal Garante per la protezione dei dati con riguardo alla sofferta discriminazione subita dai lavoratori di una società di consegne a domicilio, per effetto dell'ingiusto uso di un algoritmo, in assenza di qualsivoglia adempimento normativo.
Garante privacy - Provvedimento del 2 novembre 2024 (doc. web n. 10085455).

[35] Si richiama in tal senso anche il contenuto delle Linee guida EDPB 1/2024 sul trattamento dei dati personali in base all'articolo 6(1)(f) GDPR del 20 novembre 2024.

[36] Il parere richiama il c.d. test in tre fasi (1-identificazione dell'interesse legittimo perseguito dal titolare del trattamento o da una terza parte; 2. analisi della necessità del trattamento ai fini dell'interesse legittimo perseguito (anche denominato "test di necessità"); e 3. valutazione che l'interesse legittimo non sia superato dagli interessi o dai diritti e dalle libertà fondamentali degli interessati (anche denominato "test di bilanciamento").

[37] Ci si riferisce, ad esempio, alle linee guida dell'EDPB: n. 2/2019 sul trattamento di dati personali ai sensi dell'articolo 6, paragrafo 1, lettera b), del regolamento generale sulla protezione dei dati nel contesto della fornitura di servizi online agli interessati dell'8 ottobre 2019; Linee guida 5/2020 sul consenso ai sensi del regolamento (UE) 2016/679 del 4 maggio 2020.

[38] Tribunale di Amburgo, Rif.: 310 O 227/23 (Kneschke/LAION).

[39] Art. 3 della direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico

algoritmico per gli scopi di ricerca scientifica. La conseguenza diretta è l'affievolimento della posizione degli interessati, pur titolari del diritto di proprietà, con la limitazione, di fatto, della possibilità di potersi opporre. Questo esempio esprime pienamente la complessità dell'iter decisione dei presupposti di liceità di trattamento e della necessaria attività di compliance.

A questo punto, appare interessante rilevare lo stato dell'arte nazionale attraverso i provvedimenti del Garante privacy italiano inerenti tali tematiche. Fra i casi più interessanti vi è quello inerente alla società statunitense Clearview[40], destinataria di una sanzione di 20 milioni di euro, per aver messo in atto un vero e proprio monitoraggio biometrico mediante un sistema di ricerca avanzato, basato su sistemi di ricerca IA. Secondo quanto contestato, i dati a supporto di tale servizio erano immagini di volti di persone di tutto il mondo, estratte tramite *web scarping* e incrociate con altre informazioni correlate, in assenza di adeguate misure di trasparenza e di una idonea base giuridica[41], nei termini già rappresentati. Fra gli altri e più recenti provvedimenti, certamente merita una specifica analisi quello a carico di OpenAI, per il notorio ChatGPT, di recentissima adozione e di sostanziale portata[42]. Tra i diversi punti alla base della contestazione, rientrano temi cruciali fra i quali: i requisiti di trasparenza in relazione alla raccolta e al trattamento dei dati utilizzati per l'addestramento degli algoritmi, sia degli utenti del servizio che degli interessati terzi nonché l'adeguata messa a disposizione dell'informativa privacy; la variazione della corretta base giuridica per il trattamento di tali dati, diversa da quella contrattuale che era stata individuata dalla società; messa a disposizione di adeguati strumenti per il diritto di opposizione dei dati acquisiti in sede di utilizzo degli algoritmi. Una menzione a parte merita un punto decisivo del provvedimento, relativo al mancato accertamento dell'età degli utenti per la potenziale esposizione di minori ai rischiosi trattamenti effettuati dagli algoritmi, per la cui delicatezza

digitale e che modifica le direttive 96/9/CE e 2001/29/CE: «Gli Stati membri dispongono un'eccezione ai diritti di cui all'articolo 5, lettera a), e all'articolo 7, paragrafo 1, della direttiva 96/9/CE, all'articolo 2 della direttiva 2001/29/CE, e all'articolo 15, paragrafo 1, della presente direttiva per le riproduzioni e le estrazioni effettuate da organismi di ricerca e istituti di tutela del patrimonio culturale ai fini dell'estrazione, per scopi di ricerca scientifica, di testo e di dati da opere o altri materiali cui essi hanno legalmente accesso».

[40] Garante privacy - Ordinanza ingiunzione nei confronti di Clearview AI del 10 febbraio 2022 (doc. web 9751362).

[41] Nella nota informativa del 20 maggio 2024, doc. web n. 10020334, Web scraping ed intelligenza artificiale generativa: nota informativa e possibili azioni di contrasto, il Garante privacy italiano esamina la complessa fattispecie del *webscraping* prescrivendo una serie di accorgimenti ai titolari di siti internet e piattaforme che utilizzano bot o sistemi di tale tipologia, indicando che «I gestori di siti web e di piattaforme online che rivestano al tempo stesso il ruolo di titolari del trattamento, fermi restando gli obblighi di pubblicità, accesso, riuso e di adozione delle misure di sicurezza previste dal GPD, dovrebbero valutare, caso per caso, quando risulti necessario, in conformità alla vigente disciplina, sottrarre i dati personali che trattano ai bot di terze parti mediante l'adozione di azioni di contrasto come quelle indicate che, sebbene non esaustive né per metodo, né per risultato, possono contenere gli effetti dello *scraping* finalizzato all'addestramento degli algoritmi di intelligenza artificiale generativa. In tal senso, si è anche esplicitamente espressa la Corte di Cassazione, I sez. civile, ordinanza 2021/14381, che ha statuito che nel caso di attività di elaborazioni algoritmiche (nel caso di specie, finalizzato all'elaborazione di profili reputazionali di persone fisiche o giuridiche, il requisito di consapevolezza non può considerarsi soddisfatto ove lo schema esecutivi dell'algoritmo e gli elementi di cui si compone, restino ignoti e non conoscibili da parte dell'interessato» e «la validità del trattamento è costituita dal consenso».

[42] Garante privacy, provvedimento del 27 novembre 2024 (doc. web 10077129).

nel prossimo paragrafo viene riservato un separato approfondimento.

3. IA e minori

Il coinvolgimento dei minori nel contesto tecnologico-digitale diviene ogni giorno più consistente, divenendo ormai una categoria fortemente utilizzatrice dei servizi della società dell'informazione[43]. I più recenti dati [44] dimostrano la forte attitudine di questa categoria all'utilizzo di internet, degli strumenti di identità digitale e degli acquisti online. Si tratta di una nuova categoria di utenti, nativamente inclini all'utilizzo di sistemi digitali ma ancora inesperti rispetto ai molteplici rischi nascosti nel contesto virtuale, sono destinatari di un livello più elevato di tutela[45], prevista dalla normativa internazionale e interna. Particolarmente significativi sono l'art. 24 della Carta dei diritti fondamentali[46], il considerando 38 del GDPR[47]. In generale l'approccio regolamentativo[48] verso questa specifica categoria di utenti è tutt'altro che semplice ma nel contesto dei servizi della società dell'informazione diventa ancora più impegnativo. L'art. 8 del GDPR chiarisce che nel contesto di tali servizi, «per i trattamenti basati sul consenso, il trattamento di dati personali del minore è lecito ove il minore abbia almeno 16 anni. Ove il minore abbia un'età inferiore ai 16 anni, tale trattamento è lecito soltanto se e nella misura in cui tale consenso è prestato o autorizzato dal titolare della responsabilità genitoriale». Per lo spazio normativo consentito al legislatore nazionale, in Italia l'età si riduce a 14 anni[49]. Per rendere effettiva questo adempimento, il secondo paragrafo dell'art.8 impone l'impegno «in ogni modo ragionevole» di verificare «che il consenso sia prestato o autorizzato dal titolare della responsabilità genitoriale sul minore, in considerazione

[43] La definizione di servizi della società dell'informazione è prevista dall' 1, par. 1, lett. b), della direttiva (UE) 2015/1535.

[44] Report ISTAT 2023 Competenze digitali e caratteristiche socio-culturali della popolazione. Si veda inoltre il contenuto della relazione del Senato del 18 giugno 2024 al progetto di legge "in materia di minori e Internet, con riferimento particolare all'accesso alle piattaforme e all'uso dell'immagine dei minori" che l 73% dei minori (tra i 6 e i 17 anni) ha dichiarato di connettersi a Internet quotidianamente.

[45] G.Wang - J.Zhao - M. Van Kleek - N.Shadbolt, *Informing Age-Appropriate AI: Examining Principles and Practices of AI for Children,* in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), Association for Computing Machinery,* 536, 2022, 1 ss.

[46] «I minori hanno diritto alla protezione e alle cure necessarie per il loro benessere. Essi possono esprimere liberamente la propria opinione» e che impone l'interesse preminente del minore. Si ricorda inoltre l'art. 24 dalla Convenzione delle Nazioni Unite sui diritti dell'infanzia e dell'adolescenza, ulteriormente sviluppati nell'osservazione generale n. 25 della Convenzione delle Nazioni Unite dell'infanzia e dell'adolescenza per quanto riguarda l'ambiente digitale, che prevedono la necessità di tenere conto delle loro vulnerabilità e di fornire la protezione e l'assistenza necessarie al loro benessere.

[47] «I minori meritano una specifica protezione relativamente ai loro dati personali, in quanto possono essere meno consapevoli dei rischi, delle conseguenze e delle misure di salvaguardia interessate nonché dei loro diritti in relazione al trattamento dei dati personali». In tal senso, I. A. Caggiano, *Privacy e minori nell'era digitale. Il consenso al trattamento dei dati dei minori all'indomani del Regolamento UE 2016/679, tra diritto e tecno-regolazione*, in *Familia*, 1, 2018, 3 ss. e D. Marcello, *Il trattamento dei dati digitali del minore*, in *Actualidad Jurídica Iberoamericana,* 17 bis, 1305 ss.

[48] In prospettiva comparatistica, cfr. M.L. Chiarella, *Paradigmi della minore età. Opzioni e modelli di regolazione giuridica tra autonomia, tutela e responsabilità. Profili di diritto comparato*, Soveria Mannelli, 2008, *passim*.

[49] Art. 2 *quinquies,* d.lgs. 30 giugno 2003, n. 196.

delle tecnologie disponibili», che va inquadrato in combinazione con gli artt. 24[50] e 25[51] inerenti l'accountability del Titolare e i principi di privacy by design e by default.
Il recente provvedimento del Garante privacy su ChatGPT, di cui al paragrafo precedente, tocca il tema nodale dell'age verification (o age gate) per l'accesso ai servizi della società dell'informazione, che coinvolgono a pieno titolo anche quelli legati a sistemi o a modelli di intelligenza artificiale. La questione pendente in tale parere era quella di verificare se la società avesse provveduto ad implementare siffatti sistemi, sia rispetto ai nuovi utenti sia rispetto a quelli già registrati fino a quel momento, per impedire l'utilizzo del servizio a minori di 13 anni e per consentirlo ai minori di 18 anni solo previo valido consenso del titolare della responsabilità genitoriale.
Altro provvedimento del Garante privacy inerente il chatbot Replika[52] ci conduce ad un'ulteriore problematica fondamentale inerente i rischi della pervasività algoritmica a danno di categorie vulnerabili e, nello specifico, dei minori.
Il sistema conversazionale Replika si presentava come uno strumento di benessere, dotato di interfaccia scritta e vocale basata sull'intelligenza artificiale, capace di incidere positivamente sull'umore e sul benessere emotivo. Attraverso la comprensione dei pensieri e dei sentimenti dell'utente, Replika teneva traccia delle variazioni dell'umore e della capacità di controllo dello stress e dell'ansia, proponendosi quale supporto per razionalizzare gli obiettivi, socializzare nonché sul piano sentimentale.
Il sistema poteva essere configurato da semplice "amico virtuale" fino a partner romantico e mentore. Tuttavia, l'utilizzo pratico dimostrava l'insidiosa natura dell'algoritmo, strutturata in assenza di filtri per l'adattamento dei contenuti[53] in relazione all'età o alla tipologia dell'utente. Da alcuni test effettuati e poi diramati, nemmeno la dichiarazione esplicita della minore età attuava modifiche sostanziali ai contenuti delle conversazioni da parte del bot. Il cosiddetto "amico virtuale", a prescindere da un accertamento o

[50] Art.24 GDPR: «1. Tenuto conto della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, nonché dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche, il titolare del trattamento mette in atto misure tecniche e organizzative adeguate per garantire, ed essere in grado di dimostrare, che il trattamento è effettuato conformemente al presente regolamento. Dette misure sono riesaminate e aggiornate qualora necessario.
2. Se ciò è proporzionato rispetto alle attività di trattamento, le misure di cui al paragrafo 1 includono l'attuazione di politiche adeguate in materia di protezione dei dati da parte del titolare del trattamento».

[51] Art. 25 GDPR: «1. Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento, sia al momento di determinare i mezzi del trattamento sia all'atto del trattamento stesso il titolare del trattamento mette in atto misure tecniche e organizzative adeguate, quali la pseudonimizzazione, volte ad attuare in modo efficace i principi di protezione dei dati, quali la minimizzazione, e a integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del presente regolamento e tutelare i diritti degli interessati.
2. Il titolare del trattamento mette in atto misure tecniche e organizzative adeguate per garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento. Tale obbligo vale per la quantità dei dati personali raccolti, la portata del trattamento, il periodo di conservazione e l'accessibilità. In particolare, dette misure garantiscono che, per impostazione predefinita, non siano resi accessibili dati personali a un numero indefinito di persone fisiche senza l'intervento della persona fisica».

[52] Garante Privacy, provvedimento del 2 febbraio 2023 (doc. web 9852214).

[53] Scorza: "Perché abbiamo bloccato Replika, minori a rischio" - Intervento di Guido Scorza, 4 febbraio 2023, in *garanteprivacy.it* (doc. web 9854194).

dai contenuti delle dichiarazioni, spingeva la conversazione verso contenuti sessuali sempre più espliciti, sollecitando l'invio di materiale pornografico e giungendo poi a proporre una versione del programma ancora più spinta, ma con l'obbligo di pagare un abbonamento mensile. Inoltre, in alcuni casi, il sistema giungeva a suggerire azioni e comportamenti inaccettabili e non consentiti, sul piano sia morale sia legale. In alcuni passaggi, la chat stimolava il suo interlocutore a compiere azioni fuori da ogni parametro, quali ad esempio il suicidio o l'uccisione del proprio genitore, argomentando e sostenendo tale indicazione nel caso in cui l'interlocutore umano esprimesse perplessità al riguardo. Il tutto con impatti devastanti su soggetti vulnerabili, come quelli affetti da particolari malattie anche psichiatriche e i minori.

Per il suo ruolo cruciale, il tema dell'*age verification* sta instancabilmente interessando i tavoli di lavoro delle principali autorità istituzionali coinvolte sul piano europeo[54] e nazionale[55].

Inoltre, anche le organizzazioni comunitarie[56] e internazionali stanno intensamente lavorando su questo tema, evidenziando la necessità e l'urgenza di un intervento risolutivo. L'Unicef[57] e World Economic Forum hanno presentato un documento di lavoro, aperto ai contributi degli stakeholders, che individua zone di vulnerabilità quali la privacy e la sicurezza, la possibile esposizione a contenuti dannosi, i rischi per la salute e l'identità, le implicazioni cognitive e psicologiche, i diritti di uguaglianza e di inclusione. Le mire europee, il cui assetto normativo globale verrà esaminato nel paragrafo successivo, consapevoli di tali esigenze, si esplicano su politiche e comunicazioni che insistono sulla creazione di un ambiente sicuro[58] in cui tali tipologie di utenti possano

---

[54] Vedasi in tal senso, a titolo esemplificativo: l'attività dell'Autorità garante spagnola AEPD Decálogo de
principios *Verificación de edad y protección de personas menores de edad ante contenidos inadecuados*, dicembre 2023 e l'annuncio dell'istituzione di una task force in Spagna per progettare un sistema di age verification, 12 marzo 2024; le indagini dell'autorità garante francese CNIL *Online age verification: balancing privacy and the protection of minors*; la strategia per i minori e il di condotta per i minori dell'Autorità garante inglese ICO; il piano di coordinamento fra Autorità garante privacy italiana e AGCOM.

[55] Si segnala la recente comunicazione del 7 ottobre 2024 in cui l'AGCOM ha approvato lo schema di regolamento che disciplina le modalità tecniche e di processo per l'accertamento della maggiore età degli utenti (*age assurance*, ovvero "garanzia dell'età", talvolta indicato come "*age verification*"), in attuazione della legge 13 novembre 2023, n. 159 ("Decreto Caivano"). A seguito di tale approvazione, dovrà essere istituito un tavolo tecnico di monitoraggio e analisi delle evoluzioni tecniche, normative e regolamentari in materia di sistemi di age assurance.

[56] La tematica coincide con un'altra normativa fondamentale nel piano strategico UE, il Regolamento (UE) 2022/2065 del Parlamento Europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (Regolamento sui servizi digitali), DSA, che mira a creare un ambiente digitale inclusivo e sicuro per tutti gli utenti, specialmente per i minori. In tale contesto, la Commissione ha istituito una specifica task force on Age Verification, cui partecipa, per l'Italia, l'Autorità Garante per le Garanzie nelle Comunicazioni. L'Autorità è impegnata su tale tema in forza delle disposizioni di cui al c. 3 dell'art. 13-bis del decreto-legge n. 123/2023 che ha appunto demandato ad AGCOM il compito di individuare le modalità tecniche e di processo che i soggetti individuati dalla norma sono tenuti ad adottare per l'accertamento della maggiore età degli utenti.

[57] Unicef e World Economic Forum,in unicef.org Children and AI Where are the opportunities and risks?.

[58] Comunicazione Un decennio digitale per bambini e giovani: la nuova strategia europea per un internet migliore per i ragazzi (BIK+); Dichiarazione europea sui diritti e i principi digitali per il decennio digitale (cap.V).

interagire in sicurezza.

4. IA e scenario normativo europeo e nazionale di riferimento

La regolamentazione dell'IA rientra fra le priorità della Commissione Europea [59], dalla portata rivoluzionaria ed innovativa. A seguito della proposta della Commissione del 2021[60] e di un iter normativo tutt'altro che in discesa, il legislatore europeo ha conseguito il primato mondiale nella regolamentazione dei principi strutturali di una materia estremamente complessa e soprattutto in continuo divenire. L'AI Act si inserisce nel percorso individuato dal legislatore comunitario per dare attuazione alla strategia industriale europea[61], pilastro strutturale dell'azione della Commissione, per puntare[62] ad accrescere la leadership, l'indipendenza in campo tecnologico e il rafforzamento del mercato unico[63].

Con questa normativa, l'Europa punta a completare tale quadro strategico per beneficiare dell'enorme potenziale economico e commerciale derivante dallo sviluppo e dall'utilizzo di queste tecnologie[64]. Tale piano è reso ancora più efficace dal crescente volume di dati[65], linfa vitale dello sviluppo economico, nel contesto pubblico e privato, per la creazione di nuovi prodotti e servizi, l'incremento della produttività, risorsa ne-

[59] Le priorità della Commissione Europea 2019-2024, .

[60] Proposta di Regolamento del Parlamento e del Consiglio, che stabilisce regole armonizzate sull'Intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione COM(2021) 206 final.

[61] L'analisi elaborata alla base della (COM(2020) 66 final) evidenzia come la Commissione miri alla supremazia economica unionale anche puntando all'eliminazione delle dipendenze strategiche relativi a beni, prodotti o servizi dal forte impatto economico con basso potenziale di diversificazione e e sostituzione nella produzione UE, come quello del settore tecnologico anche in ambito infrastrutturale.

[62] Tale priorità rientra nel piano 2019-2024 della Commissione Europea e ha portato all'adozione di altri regolamenti di elevatissima portata fra cui il Regolamento sui chip (Regolamento (UE) 2023/1781 del Parlamento europeo e del Consiglio del 13 settembre 2023 che istituisce un quadro di misure per rafforzare l'ecosistema europeo dei semiconduttori e che modifica il Regolamento (UE) 2021/694 (Regolamento sui chip) e Regolamento (UE) 2024/1252 del Parlamento europeo e del Consiglio dell'11 aprile 2024 che istituisce un quadro atto a garantire un approvvigionamento sicuro e sostenibile di materie prime critiche e che modifica i regolamenti (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1724 e (UE) 2019/1020.

[63] Commissione europea, Direzione generale della Comunicazione, Mercato unico, Ufficio delle pubblicazioni, 2020

[64] In tal senso è fondamentale evidenziare come l'impostazione adottata sia conseguenza delle attente analisi realizzate nella fase preparatoria della normativa, come ad esempio il *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia* della Commissione Europea, legato alla creazione di due principali elementi costitutivi: ecosistema di eccellenza, nell'intera catena di valore dalla progettazione dei sistemi fino all'adozione dei medesimi da parte degli utilizzatori. ecosistema di fiducia, basato appunto sul rispetto delle normative vigenti, per incentivare l'utilizzo di tali sistemi da parte di cittadini e imprese.

[65] Come emerge dalle stime della Commissione Europea, il volume dei dati è in costante crescita: da 33 zettabyte nel 2018 a 175 zettabyte previsti per il 2025 con un corrispondente aumento del 530% e con un valore stimato di 829 miliardi di euro.

cessaria per il *training* e il funzionamento dei sistemi di intelligenza artificiale[66].
Grazie a questo Regolamento, vengono finalmente consolidate le classificazioni dei sistemi di IA, classificando la vasta categoria di soggetti coinvolti (fornitori; deployer; rappresentante autorizzato; importatore; distributore; operatore[67]) ed improntando un approccio basato sul rischio cui vengono collegati specifici obblighi di trasparenza.
La scelta adottata è di elevatissimo pregio, prevedendo il divieto per sistemi c.d. a rischio inaccettabile [68], strutturando delle cautele e degli adempimenti proporzionali per quelli cui invece corrispondono livelli di rischio alto[69], limitato e minimo o nullo.
Lo scopo fondamentale dell'AI Act[70] è di contribuire al miglioramento del mercato interno attraverso un quadro giuridico uniforme per disciplinare lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di sistemi di IA nell'Unione, in conformità dei valori dell'Unione, promuovere la diffusione di un'IA antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea («Carta»), compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, la tutela dei valori democratici e di protezione dei dati personali[71], al riparto dagli effetti

[66] Su tale strategia vedasi come nella citata European Data Strategy, la Commissione Europea abbia strutturato una serie di interventi normativi mirati alla supremazia europea in materia di dati, personali e non.

[67] Cfr. definizioni art. 3, parr.3-8), AI Act.

[68] Art. 5 AI Act, che prevede l'esclusione per sistemi la cui attività possa configurare: «sfruttamento delle vulnerabilità delle persone, manipolazione e utilizzo di tecniche subliminali; punteggio sociale per finalità pubbliche e private; attività di polizia predittiva individuale basate unicamente sulla profilazione delle persone;
scraping non mirato di immagini facciali da internet o telecamere a circuito chiuso per la creazione o l'ampliamento di banche dati; riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione, eccetto per motivi medici o di sicurezza (ad esempio il monitoraggio dei livelli di stanchezza di un pilota);
categorizzazione biometrica delle persone fisiche sulla base di dati biometrici per trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche o orientamento sessuale. Sarà ancora possibile etichettare o filtrare set di dati e categorizzare i dati nell'ambito delle attività di contrasto; identificazione biometrica remota in tempo reale in spazi accessibili al pubblico da parte delle autorità di contrasto, fatte salve limitate eccezioni». Schema riepilogativo del PE.

[69] Art. 6 AI Act.

[70] Considerando 1 AI Act. Vedasi anche le stesse indicazioni nel considerando 2, che testualmente indica «Il presente regolamento dovrebbe essere applicato conformemente ai valori dell'Unione sanciti dalla Carta agevolando la protezione delle persone fisiche, delle imprese, della democrazia e dello Stato di diritto e la protezione dell'ambiente, promuovendo nel contempo l'innovazione e l'occupazione e rendendo l'Unione un leader nell'adozione di un'IA affidabile» e nell'art.1.

[71] C. Novelli - G. Sandri, *Digital Democracy in the Age of Artificial Intelligence*, in *SSRN Electronic Journal,* 1, 2024, 1 ss., in cui viene offerto un focus, non rientrante nel presente contributo, inerente i possibili rischi in un particolare contesto cruciale con riguardo agli aspetti democratici quale è quello elettorale. In questo senso è molto interessante l'inciso in cui viene indicato (spec. 19) che «Se da un lato l'intelligenza artificiale può migliorare l'efficienza e la personalizzazione delle campagne, dall'altro comporta rischi significativi. La disinformazione generata dall'intelligenza artificiale e i dilemmi etici sono problemi prevalenti, di cui i deepfake sono un esempio notevole. Questi video realistici ma falsi possono manipolare la percezione del pubblico e diffondere informazioni false. Possono quindi essere utilizzati per promuovere una competizione non etica, se non illegale, tra i candidati». Altro passaggio particolarmente significativo è quello del ruolo dell'IA nelle piattaforme, laddove chiarisce (spec. 21) «Le piattaforme digitali hanno rimodellato la partecipazione politica, offrendo nuove strade per

nocivi dei sistemi di IA nell'Unione.

Il Regolamento non assume solo il primato nel disciplinare una materia obiettivamente complessa, ma anche di farlo in maniera etica[72] e responsabile, affrontando fin da subito un tema estremamente fondamentale quale è quello della tutela dei diritti fondamentali delle persone. Il perché di questo primo obiettivo è chiaramente spiegato nel considerando 7, laddove viene chiarito che «L'IA (...) può nel contempo, a seconda delle circostanze relative alla sua applicazione, al suo utilizzo e al suo livello di sviluppo tecnologico specifici, comportare rischi e pregiudicare gli interessi pubblici e i diritti fondamentali tutelati dal diritto dell'Unione. Tale pregiudizio può essere sia materiale sia immateriale, compreso il pregiudizio fisico, psicologico, sociale o economico».

Ciò si verifica quando tale tecnologia venga utilizzata impropriamente trasformandosi, ad esempio, in un potente strumento legato a pratiche di manipolazione, sfruttamento e controllo sociale[73], al fine di indurre le persone a comportamenti indesiderati o ad alterarne il processo decisionale e la libera scelta. Si tratta di pratiche particolarmente pervasive, contrarie ai valori dell'Unione relativi al rispetto della dignità umana, alla libertà, all'uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e ai diritti dei minori, fra cui quelli consacrati nel Trattato sull'Unione Europea (TUE).

A completare il quadro europeo è intervenuta recentemente la Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto, adottata il 17 maggio 2024 dal Consiglio d'Europa[74]. Il testo definitivo è giunto a seguito di un lungo lavoro di elaborazione e aperto alla firma a partire dal 5 settembre 2024, che ha coinvolto sia gli stati membri del Consiglio d'Europa che stati non membri, rendendo

l'impegno civico e la difesa. L'intelligenza artificiale migliora questi processi attraverso la comunicazione personalizzata, il monitoraggio in tempo reale e l'analisi dei dati, ma pone anche rischi di manipolazione e disinformazione. L'intelligenza artificiale migliora l'efficienza e l'integrità dei moderni processi elettorali attraverso la registrazione degli elettori, il voto elettronico e la tabulazione dei risultati. Tuttavia, solleva anche problemi di privacy, sicurezza e fiducia. Le capacità predittive dell'IA nel comportamento elettorale introducono nuove dinamiche nella competizione politica, sollevando preoccupazioni etiche sulla manipolazione e sulla legittimità democratica».

[72] Riguardo la questione sull'etica e sulla responsabilità dell'IA, fondamentale è stato il contributo della *soft law* sviluppatasi nel corso dei lavori preparatori dell'AI Act. In particolare, è doveroso menzionare l'elaborazione dei principi etici di cui al già menzionato contributo "Orientamenti etici per un'IA affidabile", accolto poi anche dalla Commissione nella comunicazione dell'8 aprile 2019 COM(2019) 168 final, precisamente riguardo i quattro principi etici (rispetto autonomia umana, prevenzione dei danni, equità, esplicabilità) e dei sette requisiti fondamentali (intervento e sorveglianza umana, robustezza tecnica e sicurezza, riservatezza e governance dei dati, trasparenza, diversità e non discriminazione, benessere sociale e ambientale, accountability).

[73] Considerando 28 AI Act.

[74] Testo ufficiale della Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo stato di diritto. Vedasi il comunicato stampa del Consiglio d'Europa del 17 maggio 2024 e in particolare quanto ha dichiarato la Segretaria generale del Consiglio d'Europa, Marija Pejčinović: «La Convezione quadro sull'intelligenza artificiale è un trattato globale unico nel suo genere, che assicurerà che l'intelligenza artificiale rispetti i diritti delle persone. È una risposta alla necessità di disporre di una norma di diritto internazionale sostenuta da Stati di diversi continenti uniti da valori comuni, che consenta di trarre vantaggio dall'intelligenza artificiale, riducendo al contempo i rischi che questa presenta. Con questo nuovo trattato, intendiamo assicurare un utilizzo responsabile dell'IA che rispetti i diritti umani, la democrazia e lo Stato di diritto».

il trattato un elemento unificante sulla protezione mondiale dei diritti umani nei sistemi di IA.

Senza interferire con le disposizioni del Regolamento, dal contenuto più legato ad aspetti economici e tecnici, la Convenzione incentiva in tutto il ciclo di vita dei sistemi di IA l'applicazione dei principi in materia di diritti umani quali: la tutela della dignità umana e dell'autonomia individuale; la trasparenza e il controllo, fondamentali nel contrasto all'opacità e all'autonomia propria dei sistemi algoritmici; il principio di responsabilità dei soggetti che detengono il controllo delle varie fasi del ciclo di vita dei sistemi di IA; l'eguaglianza e la non discriminazione, contro i c.d. "*bias* cognitivi"; la garanzia dell'efficacia degli strumenti di tutela; l'affidabilità dei sistemi; le garanzie per un sistema sicuro per lo sviluppo, la sperimentazione e i test di sistemi di intelligenza artificiale.

A completamento di questo quadro, la Convenzione prevede poi una strutturata serie di rimedi per gli interessati danneggiati da tali sistemi e un obbligo di alfabetizzazione digitale per promuovere un utilizzo consapevole degli strumenti digitali e di IA.

Con riguardo allo scenario nazionale, il quadro dei valori è individuabile nel dettato della Costituzione italiana, soprattutto in alcune specifiche definizioni dalla grande portata nel contesto in esame[75]. Nel novero degli interessi in gioco e dei rischi delle parti, meritano una speciale menzione gli artt. 15 e 21 Cost., inerenti, rispettivamente, i diritti alla riservatezza e alla libertà di espressione, l'art.3 sul diritto di uguaglianza[76], l'art. 32 sul diritto alla salute e l'art. 4 sul diritto al lavoro[77], parziali tasselli del composito quadro di diritti fondamentali possibilmente coinvolti dai rischi concreti emergenti dall'utilizzo dei sistemi di IA. Separata menzione certamente merita l'art. 2, che recita «La Repubblica riconosce e garantisce i diritti inviolabili dell'uomo, sia come singolo, sia nelle formazioni sociali ove si svolge la sua personalità e richiede l'adempimento dei doveri inderogabili di solidarietà politica, economica e sociale» rappresentando un fattivo impegno statale[78].

Evidentemente, proprio in forza di tale responsabilizzazione, l'Italia è stata il primo stato ad accompagnare sul piano interno le previsioni dell'AI Act, ancora prima dell'approvazione dell'AI Act, procedendo all'emanazione di uno schema di disegno di legge

[75] Per un'articolata disamina su tutti i diritti coinvolti, R. Razzante, *AI e tutela dei diritti fondamentali*, in *Dirittifondamentali.it,* 1, 2024, 133 ss.

[76] A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal - Rivista di BioDiritto* ,1, 2019, 63 ss.

[77] Sui possibili impatti rispetto al contesto lavorativo, L. Rinaldi, *Intelligenza artificiale, diritti e doveri nella Costituzione italiana*, in *DPCE online,* 2022, 1, 201 ss., laddove rappresenta il duplice rischio inerente la perdita da chance lavorative per effetto della modernizzazione tecnologica dei contesti lavorativi e le possibili discriminazioni derivanti da un'ingiustizia algoritmica.

[78] La nostra Costituzione, attraverso il più articolato complesso di diritti riconosciuti e garantiti, riesce, a parere di chi scrive, ad abbracciare le nuove esigenze derivanti dal rinnovato contesto tecnologico rappresentato nel presente contributo. Tuttavia sono molto accese in dottrina discussioni inerenti l'effettiva efficacia dell'insieme di tutele offerte dalla Costituzione e l'eventuale necessità di un nuovo intervento integrativo. In questo senso, si veda, fra gli altri, A. Adinolfi, *Evoluzione tecnologica e tutela dei diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *Quaderni AISDUE* ,15, 2023, 321 ss. e O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione nell'era di Internet,* in questa *Rivista*, 2018, 1, 48 ss.

recante disposizioni e deleghe in materia di intelligenza artificiale (DDL)[79]. Il DDL si propone l'obiettivo di individuare un bilanciamento tra opportunità e rischi attraverso una previsione che promuova «l'utilizzo di tali tecnologie per il miglioramento delle condizioni di vita dei cittadini e della coesione sociale e, dall'altro, fornisca soluzioni per la gestione del rischio fondate sulla visione antropocentrica»[80].

Senza sovrapporsi all'AI Act, l'obiettivo del DDL è di dettare[81] una normativa nazionale che predisponga un sistema di principi, governance e misure specifiche adatte al contesto italiano per cogliere tutte le opportunità dell'intelligenza artificiale. Gli obiettivi generali[82] sono «il rafforzamento della competitività italiana e garantire ai cittadini italiani l'uso affidabile e responsabile dell'IA, assicurando la supervisione umana in ogni fase di sviluppo e di utilizzo dei sistemi IA e la tutela dei diritti fondamentali».

In particolare, lo scopo è di assicurare che l'utilizzo dell'IA non comporti una lesione dei diritti fondamentali dell'individuo, ai sensi dell'art. 2 Cost.[83]. Il DDL si apre con un ambito di definizioni, individuando i principi applicabili all'IA, procedendo poi con norme specifiche sull'utilizzo dell'IA nel settore sanitario, nel diritto del lavoro, nelle professioni intellettuali, nella PA e nell'attività giudiziaria. Altro aspetto fondamentale è l'individuazione dell'Agenzia per l'Italia digitale (AgID) e l'Agenzia per la cybersicurezza nazionale (ACN), quali autorità di controllo ed espletare le funzioni richieste dall'AI Act, anche finalizzate a dare sostegno e impulso alla realizzazione di sistemi di IA sul piano interno, nei termini e nelle competenze individuati nell'art. 18 DDL[84].

---

[79] Disegno di legge presentato dal Presidente del Consiglio dei Ministri (Meloni) e dal Ministro della giustizia (Nordio) comunicato alla Presidenza il 20 maggio 2024 Disposizioni e delega al Governo in materia di intelligenza artificiale, testo DDL.

[80] Relazione del DDL, 3.

[81] Analisi del quadro normativo in materia di Intelligenza artificiale (D.D.L. IA e regolamento (UE) su IA) Focus IA.

[82] Premessa Relazione del DDL, cit.

[83] Punto 4) Analisi d'impatto della regolamentazione, relativa al DDL, ivi, 16.

[84] Art. 18 DDL: «a) l'AgID è responsabile di promuovere l'innovazione e lo sviluppo dell'intelligenza artificiale, fatto salvo quanto previsto dalla lettera b). L'AgID provvede, altresì, a definire le procedure e a esercitare le funzioni e i compiti in materia di notifica, valutazione, accreditamento e monitoraggio dei soggetti incaricati di verificare la conformità dei sistemi di intelligenza artificiale, secondo quanto previsto dalla normativa nazionale e dell'Unione europea.
b) l'ACN, anche ai fini di assicurare la tutela della cybersicurezza, come definita dall'articolo 1, comma 1, del decreto-legge 14 giugno 2021, n. 82, convertito, con modificazioni, dalla legge 4 agosto 2021, n. 109, è responsabile per la vigilanza, ivi incluse le attività ispettive e sanzionatorie, dei sistemi di intelligenza artificiale, secondo quanto previsto dalla normativa nazionale e dell'Unione europea. L'ACN è, altresì, responsabile per la promozione e lo sviluppo dell'intelligenza artificiale relativamente ai profili di cybersicurezza. c) l'AgID e l'ACN, ciascuna per quanto di rispettiva competenza, assicurano l'istituzione e la gestione congiunta di spazi di sperimentazione finalizzati alla realizzazione di sistemi di intelligenza artificiale conformi alla normativa nazionale e dell'Unione europea, sentito il Ministero della difesa per gli aspetti relativi ai sistemi di intelligenza artificiale impiegabili in chiave duale.
2. Le Autorità nazionali per l'intelligenza artificiale di cui al comma 1 assicurano il coordinamento e la collaborazione con le altre pubbliche amministrazioni e le autorità indipendenti, nonché ogni opportuno raccordo tra loro per l'esercizio delle funzioni di cui al presente articolo. A quest'ultimo fine, presso la Presidenza del Consiglio dei ministri è istituito un Comitato di coordinamento, composto dai direttori generali delle due citate Agenzie e dal capo del Dipartimento per la trasformazione digitale della Presidenza del Consiglio dei ministri medesima».

Tale articolo, indica espressamente[85] che restano ferme le competenze, i compiti e i poteri del Garante per la protezione dei dati personali, per l'assoluta indipendenza e verticalità della materia sul trattamento dei dati personali.

A tal proposito, è interessante richiamare il parere[86] reso dal Garante privacy sull'emanazione del DDL, con particolare riguardo agli impatti privacy derivanti dal medesimo. Il contenuto del parere è positivo ma rimane condizionato alla richiesta di più intense garanzie a tutela dei dati personali, evidenziando quindi la necessità di una maggiore attenzione su tale aspetto. Alcuni punti di tale parere meritano una separata menzione, anche ai fini del presente contributo, per evidenziare come il percorso verso l'equilibrio delle parti in gioco non possa ancora definirsi raggiunto.

Il primo aspetto che merita una particolare evidenza, anche alla luce della problematica emersa nel paragrafo precedente, riguarda la sicurezza dei minori rispetto all'utilizzo di tali tecnologie che giustifica la richiesta di riferimenti a sistemi di *age verification* per garantire limitazioni o divieti all'uso dei sistemi di IA per tale categoria. Poiché ai sensi del DDL IA[87], per il trattamento dei dati personali connessi all'utilizzo di sistemi di IA è richiesto il consenso del minore ultraquattordicenne, soglia allineata alle disposizioni dell'art. 2 *quinquies* del D.Lgs 193/2003 s.m.i., il Garante raccomanda di integrare tale previsione con un opportuno riferimento a misure idonee a garantire adeguati sistemi di verifica dell'età, per evitare che la soglia indicata possa essere in qualche modo elusa. A tal proposito, sono state richiamate le misure adottate ai sensi dell'art. 13-bis, c. 3, d.l. 123 del 2023, che prevede l'adozione di modalità tecniche da parte dell'Autorità per le garanzie delle comunicazioni (AGCOM) previo parere del Garante privacy, per l'accertamento della maggiore età degli utenti e il rispetto della minimizzazione dei dati[88], di cui si è già fatto cenno nel paragrafo precedente.

Ulteriore argomento particolarmente interessante del parere reso al Garante in relazione al DDL, riguarda la delicatissima categoria dei dati sanitari, sui quali l'autorità invoca maggiori cautele, un collegamento più sistematico fra l'art. 7 del DDL e le garanzie previste dall'art. 9 GDPR, nonché la predisposizione al futuro allineamento con lo Spazio Europeo dei dati sanitari[89]. Al di là delle ulteriori argomentazioni indicate in tale parere, anche in questo caso, come per gli altri ambiti di trattamento ivi menzionati, sembra che la questione necessiti di ulteriore consapevolezza e maggiore coordinamento per la migliore tutela degli interessati.

Con riguardo alla compatibilità e non contrasto della normativa in esame è intervenuta la Commissione Europea, che ha inviato un parere circostanziato (C(2024) 7814) al

---

85 Cfr. art. 18, c. 3, DDL, ivi, 16.

86 Garante privacy, *Parere su uno schema di disegno di legge recante disposizioni e deleghe in materia di intelligenza artificiale*, 2 agosto 2024 (doc. web 10043532).

87 Cfr. art. 4, c. 4, DDL, ivi, 16.

88 In tal senso, viene richiamato il fondamentale principio di minimizzazione dei dati espresso dall'art. 5, par. 1, lett. c), GDPR per il quale i dati personali devono essere «adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati».

89 Nel caso poi di utilizzo di sistemi di IA in ambito sanitario ad alto rischio, il Garante ha chiesto di indicare particolari limitazioni per l'utilizzo dei dati (conservazione, divieto di trasmissione, trasferimento o comunicazione) e la preferenza per l'uso di dati sintetici o anonimi. Con riguardo allo Spazio Europeo sui dati sanitari, si veda la specifica sezione informativa della Commissione Europea.

Governo italiano[90], contenente l'invito a rimuovere alcune discrepanze rispetto all'AI Act. In particolare, la Commissione, ha messo in evidenza delle sostanziali divergenze sia di tipo terminologico[91], che di tipo sostanziale[92], soffermandosi inoltre sul punto inerente la designazione delle autorità nazionali competenti, ai sensi degli artt. 18 e 22 del DDL e previste dall'art.70 dell'IA ACT, questione al centro del dibattito nazionale per il quadro di "diffusa" governance[93] impostato dal DDL in esame. Su tale aspetto, la Commissione ha evidenziato che per la designazione è richiesto lo «stesso livello di indipendenza previsto dalla direttiva (UE) 2016/680[94] per le autorità preposte alla

[90] Il testo del menzionato parere non è reperibile tramite le fonti pubbliche, l'analisi è stata svolta in base alla sintesi diramata dal Senato, 4ª Commissione permanente - Resoconto sommario n. 214 del 27/11/2024.

[91] Cfr. sintesi del Senato: «- suggerisce di inserire all'articolo 1 un riferimento specifico al regolamento europeo sull'intelligenza artificiale (IA); - in riferimento alle definizioni di cui all'articolo 2, segnala che quella di "modelli di IA" differisce da quella del regolamento europeo sull'IA e che, comunque, la norma nazionale dovrebbe limitarsi a fare riferimento alle definizioni già contenute nel regolamento senza replicarle; - in riferimento all'articolo 5, comma 1, lettera d), del disegno di legge, la Commissione europea invita a chiarire il concetto di dati "critici", limitandolo ai casi in cui sono in gioco interessi di sicurezza nazionale».

[92] Cfr. sintesi del Senato: «- in riferimento all'articolo 7, comma 3, che stabilisce obblighi informativi per gli operatori di sistemi di IA in ambito sanitario e di visibilità nei confronti dei pazienti, la Commissione europea ritiene opportuno che gli obblighi informativi dell'operatore a beneficio del paziente debbano limitarsi esclusivamente all'impiego dell'IA, senza estenderli ai "vantaggi, in termini diagnostici e terapeutici, derivanti dall'utilizzo delle nuove tecnologie" e alle "informazioni sulla logica decisionale utilizzata", per non andare oltre quanto previsto dal regolamento europeo sull'IA;
- in riferimento all'articolo 12, sull'uso dei sistemi di IA nell'ambito delle professioni intellettuali, la Commissione europea invita a eliminare qualsiasi restrizione nell'uso di sistemi di IA non "ad alto rischio", per non porsi in contrasto con il regolamento;
- in riferimento all'articolo 14, che consente l'utilizzo dei sistemi di IA nell'attività giudiziaria solo per l'organizzazione e semplificazione del lavoro giudiziario e per la ricerca giurisprudenziale e dottrinale, la Commissione europea invita ad allineare tale norma, all'articolo 6, paragrafo 3, del regolamento sull'IA, che non esclude la possibilità di utilizzare sistemi di IA pur classificati come "ad alto rischio" ma che "non presentano un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, o non influenzano materialmente il risultato del processo decisionale";- in riferimento alla delega di cui al comma 3 dell'articolo 22, volta all'organica definizione della disciplina nei casi di uso di sistemi di intelligenza artificiale per finalità illecite, la Commissione europea ricorda che l'articolo 99 del regolamento sull'IA prevede specifiche disposizioni in materia di sanzioni per violazioni del regolamento da parte degli operatori;- in riferimento all'articolo 23, comma 1, lettera b), del disegno di legge, secondo cui i contenuti prodotti dai sistemi di intelligenza artificiale devono essere resi chiaramente riconoscibili mediante un segno visibile con l'acronimo "IA" o mediante un annuncio audio, la Commissione europea ritiene che tale obbligo si sovrapponga e vada oltre gli obblighi di cui all'articolo 50, paragrafi 2 e 4, del regolamento sull'IA;- in riferimento all'articolo 23, comma 1, lettera c), del disegno di legge, che impone ai fornitori di piattaforme per la condivisione di video soggetti alla giurisdizione italiana di attuare misure a tutela del "grande pubblico da contenuti informativi che siano stati, attraverso l'utilizzo di sistemi di intelligenza artificiale, completamente generati ovvero, anche parzialmente, modificati o alterati in modo da presentare come reali dati, fatti e informazioni che non lo sono", la Commissione europea non ritiene chiaro in che modo tale disposizione non si sovrapponga all'articolo 50, comma 1, 2 e 4, del regolamento sull'IA».

[93] M. Cappai, *Intelligenza artificiale e protezione dei dati personali nel d.d.l. n. 1146: quale governance nazionale?*, in *Federalismi.it*, 30, 2024, 186 ss., che differenzia la governance imposta regolamento IA, affidata alle autorità puntualmente individuate nel DDL (AgID, ACN e Garante privacy) e la governance del fenomeno IA nel suo complesso, definendola a "geometria variabile". Il fattore collusivo è rappresentato dalle indicazioni del Garante, contenute nel citato parere del 2 agosto 2024, che in più punti ha richiamato la necessità di integrare maggiori competenze in tale ambito.

[94] Direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativa alla

protezione dei dati nelle attività delle forze dell'ordine, nella gestione delle migrazioni e controllo delle frontiere, nell'amministrazione della giustizia e nei processi democratici». I contenuti del parere, del tutto condivisibili, riflettono l'inevitabile immaturità del testo rispetto ai contenuti dell'AI Act e la necessità di maggiore integrazione, soprattutto con riguardo alla competenza delle autorità di controllo e a questioni di natura sostanziale relative alla giustizia e ai dati sanitari, sui quali il DDL deve puntare ad un maggiore coordinamento con il quadro europeo.

5. Conclusioni

L'analisi passata in rassegna rappresenta un quadro certamente innovativo e foriero di grandi opportunità. Il contesto tecnologico europeo ha abbracciato questa rivoluzione che vedrà lo sviluppo dell'IA, con benefici, come si è visto, in ogni contesto socio-economico. L'AI Act è stato predisposto appositamente per creare queste condizioni e per consentire all'Europa di cogliere questa sfida, attraverso la predisposizione di una regolamentazione coerente e articolata, che ne consenta l'efficace e rispettosa crescita. Gli aspetti etici e il rispetto dei diritti fondamentali sono fra i pilastri strutturali di tale normativa, ma gli estremi rischi che sono stati illustrati nel contributo dimostrano che l'adozione dell'AI Act non rappresenti la conclusione quanto l'avvio di un nuovo percorso. Prova ne sono l'elaborazione della Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo stato di diritto, intervenuta per disciplinare verticalmente a tali aspetti, nonché gli altri interventi da parte degli stati, come l'Italia, in cui si è voluto intervenire in maniera più incisiva per armonizzare le peculiarità nazionali. Tante questioni sono ancora aperte e le soluzioni adottate ancora aperte all'interpretazione.

Nel rapporto rischio-beneficio, probabilmente l'asticella è ancora disposta verso il primo. Per superare tale sbilanciamento, la prima soluzione da adottare in via prioritaria potrebbe essere di seguire il percorso indicato dallo stesso AI Act, che individua quale strumento risolutivo l'alfabetizzazione digitale[95]. Il considerando 20 infatti indica che «al fine di ottenere i massimi benefici dai sistemi di IA proteggendo nel contempo i diritti fondamentali, la salute e la sicurezza e di consentire il controllo democratico, l'alfabetizzazione in materia di IA dovrebbe dotare i fornitori, i deployer e le persone interessate delle nozioni necessarie per prendere decisioni informate in merito ai sistemi di IA». Si tratta quindi di uno strumento[96], attuato anche attraverso linee guida e

protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio, attuata nell'ordinamento italiano con il decreto legislativo del 18 maggio 2018, n. 51.

[95] Art. 3, n. 56, AI Act: «le competenze, le conoscenze e la comprensione che consentono ai fornitori, ai deployer e alle persone interessate, tenendo conto dei loro rispettivi diritti e obblighi nel contesto del presente regolamento, di procedere a una diffusione informata dei sistemi di IA, nonché di acquisire consapevolezza in merito alle opportunità e ai rischi dell'IA e ai possibili danni che essa può causare». L'alfabetizzazione è imposta quale obbligo a fornitori e *deployer* all'art. 4 AI Act.

[96] Art. 65 AI Act e considerando 20: «è affidato al Consiglio per l'IA il compito specifico di promuovere azioni e strumenti per l'alfabetizzazione, sensibilizzazione in materia di IA e la comprensione dei

codici di condotta[97] adottati di concerto con le istituzioni europee, che possa giungere nella forma più pratica e comprensibile ai rispettivi destinatari riguardo alle caratteristiche delle applicazioni di IA, alle misure da adottare, alle modalità per comprendere e selezionare gli output e comprendere l'incidenza della decisione di un sistema di IA. Secondo quello che viene indicato nel Regolamento, la piena attuazione delle misure di alfabetizzazione di IA potrebbe favorevolmente contribuire a migliorare le condizioni di lavoro e sostenere il percorso dell'UE verso un IA affidabile e sicura.
Si tratta di un percorso affascinante e ambizioso, ben tracciato ma certamente ancora in salita. Che vale certamente la pena di percorrere e di incoraggiare attraverso consapevolezza e coinvolgimento da parte di tutti i soggetti coinvolti.

benefici, dei rischi, delle garanzie, dei diritti e degli obblighi in relazione all'uso dei sistemi di IA».

[97] G7: inizia la fase Pilota per il monitoraggio del Codice di Condotta sull'Intelligenza Artificiale, Dipartimento della trasformazione digitale, 24 luglio 2024.

Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*

Andrea Ruffo

Abstract

L'incessante sviluppo delle tecnologie digitali ha comportato, nell'ultimo decennio, profondi mutamenti sociali. Oltre al miglioramento dei mezzi, delle capacità e dei contenuti dell'informazione e della comunicazione (sia interpersonale che di massa), la poliedrica pervasività dei nuovi strumenti *hi-tech* ha amplificato i problemi già esistenti nel mondo delle notizie; tra cui la disinformazione e, più in generale, tutte le forme di disordine informativo. Conosciute fin dai poemi omerici, le condotte disinformanti possono destabilizzare l'ordine pubblico, minare le Istituzioni, rallentandone la risposta securitaria.
La misinformazione è tra le forme di disordine in cui maggiormente l'utilizzo delle nuove tecnologie digitali, supportate dall'Intelligenza Artificiale (IA), può amplificare gli effetti dannosi o, al contrario, a contrastarli.

The relentless development of digital technologies has brought about profound social changes over the past decade. In addition to improvements in the means, capabilities and content of information and communication (both interpersonal and mass), the multifaceted pervasiveness of the new hi-tech tools has amplified existing problems in the world of news; including disinformation and, more generally, all forms of informational disorder. Known since the Homeric poems, disinformation conduct can destabilize public order, undermine institutions, and slow down their securitarian response.
Misinformation is among the forms of disorder where most the use of new digital technologies, supported by Artificial Intelligence (AI), can amplify the harmful effects or, conversely, to counter them.

Sommario

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Stato. – 4. L'apporto negativo dell'Intelligenza Artificiale al disordine informativo. – 5. Il regolamento europeo 1689/2024 e la problematica dei deepfake. – 6. Conclusioni e prospettive.

Keywords

misinformazione – intelligenza artificiale – regolamento IA – *deep fake* – prevenzione

1. L'UE e il contrasto europeo ai fenomeni disinformanti

Dall'annessione della Crimea del 2014, da parte della Federazione Russa[1], le Istituzioni dell'Unione Europea hanno iniziato ad interrogarsi sulla portata dei fenomeni disinformanti – che avevano in quell'occasione disorientato l'opinione pubblica – e su quali fossero gli strumenti normativi e sanzionatori per prevenirli e contrastarne gli effetti. Nell'anno successivo all'annessione, esattamente il 20 marzo del 2015, il Consiglio europeo avviava, attraverso l'Alto rappresentante per la politica estera e di sicurezza comune, la predisposizione di un piano di azione per contrastare le campagne di disinformazione russe[2]. Successivamente, pertanto, nel 2016 venivano istituiti: il Centro europeo di eccellenza nella lotta contro le minacce ibride e la cellula per l'analisi delle stesse[3], che si aggiunsero alla *East StratCom Task Force* (ESCTF), formata dall'Alto Rappresentante nel giugno 2015, per le comunicazioni strategiche del servizio europeo per l'azione esterna (SEAE)[4].

La risoluzione del Parlamento europeo 2016/2276(INI), del 15 giugno 2017, "Sulle piattaforme on-line e il mercato unico digitale" rappresenta una pietra miliare della strategia europea contro la disinformazione, in quanto oltre a condannare la diffusione di notizie false nel mondo digitale, sollecitava sia le piattaforme on-line a fornire agli utenti strumenti per denunciarle, che la Commissione europea ad intervenire normativamente per ridurre la disinformazione[5].

Il contrasto delle notizie false (mediaticamente definite *fake news*), diffuse online, diveniva così un elemento del programma di quadro[6] tracciato dalla Commissione europea, che prevedeva di pubblicare in un elenco apposito, "smascherandole", le fonti di disinformazione sia di istituire un gruppo di esperti che creasse un paradigma per

[1] S. Lattanzi, *La lotta alla disinformazione nei rapporti tra Unione e Stati terzi alla luce del conflitto russo-ucraino*, in questa *Rivista*, 3, 2022, 163.

[2] Si veda la relativa pagina del sito internet del Consiglio europeo

[3] Comunicazione congiunta dell'Alto rappresentante dell'Unione europea al Parlamento europeo e al Consiglio, *Quadro congiunto per contrastare le minacce ibride. La risposta dell'Unione europea*, Bruxelles, 6.4.2016.

[4] Si veda l'articolo *EU to counter Russian propaganda by promoting 'European values'*, pubblicato dal *The Guardian* il 25 giugno 2015.

[5] S. Sassi*, L'Unione Europea e la lotta alla disinformazione online*, in *federalismi*, 15, 2023, 189.

[6] Si rimanda alla comunicazione COM(2017/650 *final*), del 24.10.2017 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee of the Regions, «*Commission Work Program 2018. An agenda for more united, stronger and more democratic Europe*», Strasburgo, 4.

bilanciare il diritto dei cittadini di accedere a un'informazione di qualità con le libertà discendenti dall'art. 21 della Costituzione (libertà di manifestazione del pensiero). Il rapporto[7], prodotto l'anno successivo (2018) dal suddetto gruppo di esperti, si proponeva di tracciare un'info-sfera digitale più attendibile e trasparente in cui il primo controllo sarebbe stato affidato agli stessi fruitori, ovvero alla società civile e alle imprese private (specialmente le piattaforme di servizi online e i *social-network*/media). Veniva promossa, in questo modo, una maggiore conoscenza degli strumenti mediatici digitali (c.d. alfabetizzazione mediatica), la creazione di strumenti (anche algoritmici) che consentissero l'individuazione e la rimozione di contenuti disinformanti (parallelamente ad una commissione indipendente di verificatori) e l'elaborazione di alcune forme embrionali di regolazione interna (elenco di principii), a cui tutti gli operatori economici del mondo di Internet si sarebbero dovuti adeguare[8]. Parallelamente la Commissione UE istituiva un sistema indipendente di verificatori

Tale iniziativa avviata nel maggio del 2018 può essere considerata l'*incipit* politico e organizzativo per la successiva redazione, nella forma di auto-regolamentazione, del «Codice di buone pratiche sulla disinformazione» (Codice di condotta), pubblicato proprio a fine settembre 2018[9].

Tra il 2019 e il 2020, la persistenza di campagne di disinformazione e di condizionamento degli eventi elettorali ad opera di soggetti esterni all'Unione europea[10], spinsero le Istituzioni di Bruxelles a pianificare nuove misure di regolazione e contrasto. Si segnala a tal proposito il sistema di allarme rapido (*Rapid Alert System* – R.A.S.), voluto in ottemperanza al Piano d'azione contro la disinformazione[11], per migliorare la condivisione informativa (mediante il *micro-targeting*[12]) tra UE e Paesi membri e il contrasto dei fenomeni disinformanti.

La pandemia da Sars-Covid-19, con le conseguenti misure di isolamento domiciliare prolungato (c.d. *lockdown*), la campagna vaccinale di massa e l'incremento dell'uso delle piattaforme digitali e dei *social network* per comunicare e lavorare, contemporaneamente all'opposto diffondersi di notizie mistificatorie e destabilizzanti, ha rappresentato un momento di fortissimo impulso per il potenziamento delle misure di prevenzione e contrasto alla disinformazione dell'UE[13].

---

[7] Si rimanda al sito della Commissione Europea e all'annuncio pubblicato, nel Marzo 2018.

[8] Si veda la Relazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, COM(2018/794 *final) sull'attuazione della comunicazione "Contrastare la disinformazione on-line: un approccio europeo*", Bruxelles, 5.12.2018, 1.

[9] O. Pollicino, *I Codici di Condotta tra self-regulation e hard law: esiste davvero una terza via per la regolazione digitale? Il caso della Strategia europea contro la disinformazione online*, in *Rivista Trimestrale di Diritto Pubblico,* 4, 2022, 2 ss.

[10] Si pensi al caso Cambridge Analytica o alle svariate condotte disinformanti, attribuibili ad attori direttamente o indirettamente collegabili alla Federazione Russa, in relazione ad alcuni appuntamenti elettorali o di consultazione pubblica dei cittadini (r*eferendum*) tenutisi in Stati UE.

[11] Si cita la comunicazione congiunta della Commissione europea al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Piano d'azione contro la disinformazione*, Bruxelles, 5.12.2018, 7 ss.

[12] Sistema che prevede la minuziosa suddivisione settoriale degli obiettivi e degli elementi informativi, consentendo la comunicazione orientata secondo l'appartenenza ad una delle categorie individuate.

[13] Si veda il comunicato stampa dell'Alto Rappresentante J. Borrell e del Vicepresidente per i valori e

2. Dal Codice di buone pratiche al DSA, fino al Regolamento IA

Volendo delineare l'attuale perimetro normativo europeo di contrasto alla disinformazione online, occorre precisare che a seguito dell'approvazione del "pacchetto europeo per il Digitale[14]" avvenuta con la prima Presidenza Von der Leyen si è assistito, nel periodo post-pandemico e in concomitanza dell'invasione russa dell'Ucraina, ad un ulteriore potenziamento delle misure già adottate nel 2018.

Il 16 giugno 2022, infatti, il precedente "Codice di buone pratiche contro la disinformazione" è stato aggiornato e modificato, con la pubblicazione del "Codice rafforzato di buone pratiche sulla disinformazione".

Proprio quest'ultimo documento, denominato comunemente Codice rafforzato (*strengthened*)[15], a cui hanno aderito 34 società firmatarie[16], che punta a raggiungere gli obiettivi indicati dalla Commissione nel maggio 2021, stabilendo una gamma più ampia di impegni e misure di contrasto alla disinformazione online. I soggetti firmatari si impegnarono, infatti, a: demonetizzare la diffusione della disinformazione, garantire la trasparenza della pubblicità politica; responsabilizzare gli utenti per una consapevole fruizione di Internet; incrementare la cooperazione con i verificatori dei fatti e fornire maggiore accesso ai dati delle piattaforme.

Il codice rafforzato, inoltre, proprio per incrementare la trasparenza della realtà digitale delle piattaforme (spesso caratterizzata da asimmetria informativa) introduce: un centro *ad hoc* per fornire tutte le informazioni sulle politiche dei fornitori dei servizi di intermediazione connessi al sito principale, così come una *task force* permanente (presieduta dalla Commissione UE e formata da alcuni soggetti interessati e altri enti europei) per continuare ad aggiornare l'esecuzione delle misure di regolazione *in fieri* all'incessante progresso tecnologico, che si ripercuote anche sulle tecniche e sugli strumenti della disinformazione.

---

la trasparenza V. Jourová, del 10 giugno 2020, *Coronavirus: azione rafforzata dell'UE contro la disinformazione* – Bruxelles). Il Primo (Borrell) ha affermato che «[...]le operazioni di influenza e le campagne di disinformazione mirate sono un'arma riconosciuta di soggetti statali e non statali, l'Unione Europea sta intensificando le proprie attività e migliorando le proprie capacità per combattere questa battaglia» mentre il Vicepresidente ha aggiunto che «Per lottare contro la disinformazione, dobbiamo mobilitare tutti i soggetti interessati, dalle piattaforme digitali alle autorità pubbliche, e sostenere i verificatori di fatti e i media indipendenti. Pur avendo intrapreso iniziative positive durante la pandemia, le piattaforme digitali devono intensificare i loro sforzi. Le nostre azioni hanno radici profonde nei diritti fondamentali, in particolare nella libertà di espressione e di informazione».

[14] F. Zorzi Giustiniani, *L'Unione europea e regolamentazione del digitale: il Digital Services Package e il Codice di buone pratiche sulla disinformazione*, in *Nomos*, 2, 2022, 3 ss.

[15] M. Monti, *Il Code of Practice on Disinformation dell'UE: tentativi in fieri di contrasto alle fake news,* e Id., *Lo strengthened Code of Practice on Disinformation: un'altra pietra della nuova fortezza digitale europea?*, entrambi in questa *Rivista*, 1, 2019, 320 ss. e 2, 2022, 317 ss.

[16] Esse sono: *Adobe, Alliance4Europe, Avaaz, Clubhouse, Crisp, Demagog, DoubleVerify, DOT Europe, Ebiquity, European Association of Communication Agencies (EACA), Faktograf, Globsec, Google, IAB Europe (Interactive Advertising Bureau Europe), Kinzen, Kreativitet & Kommunikation, Logically, Maldita.es, MediaMath, Meta, Microsoft, Neeva, Newsback, NewsGuard, PagellaPoltica, Reporters without Borders (RSF), Seznam, ScienceFeedback, The Bright App, The Global Disinformation Index, The GARM Initiative, TikTok, Twitch, Twitter, Vimeo, VOST Europe, WhoTargetsMe* e *World Federation of Advertisers (WFA).*

Tale monitoraggio rafforzato è amplificato dal Codice mediante un sistema di rendicontazione con cui le piattaforme online molto grandi (*big player*[17]), ai sensi del Digital Services Act (DSA), dovranno relazionare sulle loro operazioni di contrasto e prevenzione della disinformazione ogni sei mesi, a differenza degli altri soggetti più piccoli che lo faranno annualmente. In questo sistema sono introdotti dal Codice rafforzato dei meccanismi di valutazione che giudicheranno, mediante degli indicatori numerici di prestazione (*Key Performance Indicators* – KPI) le piattaforme in base all'adeguatezza, all'efficacia e al numero delle misure anti-disinformazione attuate[18].

Le disposizioni introdotte nel 2022 dal Codice *strengthened*, fanno sì che la natura del documento sia sostanzialmente mutata rispetto a quella auto-regolatoria del suo predecessore (codice di buone pratiche)[19], sottoscritto volontariamente nel 2018 dai molte aziende di tecnologia e pubblicità digitale, e che pertanto sia da considerarsi uno strumento di co-regolamentazione, in cui la Commissione europea esercita un controllo apicale secondo quanto previsto dall'art. 45 del Digital Services Act[20].

Il Digital Services Act (DSA)[21], regolamento (UE) sui servizi forniti dalle grandi aziende del Web, che affronta la tematica dell'effettiva ed efficace moderazione e i contenuti online, introducendo anche forme di responsabilità per le piattaforme digitali, costituisce –infatti – la fonte di normazione più aggiornata del perimetro normativo europeo per il contrasto alla disinformazione *online*[22].

Attraverso l'applicazione dell'art. 74 (riguardante le sanzioni pecuniarie che la Com-

---

[17] A. Kuczerawy, *Fighting on-line disinformation: did the EU Code of Practice forget about freedom of exepression?*, in E. Kuzelewska-G. Terzis-D. Trottier-D. Kloza (eds.), *Disinformation and Digital Media as a Challenge for Democracy*, Antwerp, 8-9, 2019.

[18] È proprio l'art. 45 del DSA ad introdurre la novità dell'uso degli indicatori (specificamente degli indicatori di performance) nel quadro della valutazione secondo i codici di condotta delle piattaforme). Il punto 3 dell'art. 45 DSA, infatti, sancisce che: « [...] la Commissione e il comitato nonché, ove opportuno, altri organismi mirano a garantire che i codici di condotta definiscano chiaramente i loro obiettivi specifici, contengano indicatori chiave di prestazione per misurare il conseguimento di tali obiettivi e tengano debitamente conto delle esigenze e degli interessi di tutte le parti interessate, in particolare dei cittadini, a livello di Unione. La Commissione e il comitato mirano inoltre a garantire che i partecipanti riferiscano periodicamente alla Commissione e ai rispettivi coordinatori dei servizi digitali del luogo di stabilimento in merito a tutte le misure adottate e ai relativi risultati, misurati sulla base degli indicatori chiave di prestazione contenuti nei codici di condotta. Gli indicatori chiave di prestazione e gli obblighi di comunicazione tengono conto delle differenze esistenti tra i diversi partecipanti in termini di dimensioni e capacità.»

[19] S. Sassi, *L'Unione Europea e la lotta alla disinformazione online*, cit.

[20] Per questo il codice rafforzato tende a diventare una misura di mitigazione e un codice di condotta riconosciuto nel quadro di co-regolamentazione della DSA.

[21] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).

[22] A tale proposito rimandi espliciti alla problematica del disordine informativo sono contenuti nei consideranda n.: 69 e 83 (per il rischio generico delle campagne di disinformazione), 84 (per la valutazione del rischio da parte dei fornitori di servizi, espressamente richiamata all'art. 34), 88 (rispetto alle azioni di sensibilizzazione contro la disinformazione), 95 (sulla prevenzione nelle pubblicità delle tecniche di manipolazione e disinformazione. Si veda anche l'art. 39 sul tema della trasparenza della pubblicità online), 104 (sui codici di condotta delle piattaforme in materia di disinformazione), 106 (richiama il rafforzamento del "codice di buone pratiche sulla disinformazione), 108 (indica che anche in caso di disinformazione si applica il Meccanismo di risposta alle crisi previsto dall'art. 36).

missione può infliggere ai fornitori delle piattaforme online), inoltre, il DSA punta a disincentivare la disinformazione sotto il profilo economico; considerato che dopo aver determinato la falsità di un contenuto, oltre alla sanzione pecuniaria per l'inadempienza rispetto alle norme del regolamento, verrebbero azzerati o drasticamente ridotti anche gli introiti pubblicitari per piattaforme e motori di ricerca verrebbero azzerati o ridotti notevolmente.

Ai sensi dell'art. 10, lett. a), punto "iii", le aziende del Web dovranno effettuare controlli sulla "effettività" (ovvero sulla reale esistenza) degli account per disincentivare l'utilizzo di profili falsi, limitandone così le potenziali condotte disinformanti.

A tale quadro normativo si è aggiunto, infine, il regolamento (UE) 2024/1689, denominato più comunemente "regolamento IA" o AI Act, adottato il 13 giugno 2024 dal Parlamento europeo e del Consiglio e pubblicato sulla Gazzetta Ufficiale UE del 12/07/2024, stabilisce nuove regole sull'Intelligenza Artificiale, modificando i regolamenti preesistenti[23].

In realtà, i lavori preparatori dell'atto regolamentare erano stati avviati dalle Istituzioni UE già nell'aprile 2021, principiando proprio da un aggiornamento del precedente dispositivo normativo (UE) 2020/1828.

Entrato in vigore il 1 agosto 2024[24], vedrà applicabili le norme sulle pratiche di IA vietate, a partire dal 02 febbraio 2025, mentre successivamente, dal 02 agosto del medesimo anno, potranno essere eseguite le disposizioni relative alle autorità di notifica nazionali individuate, diventando effettivi anche i modelli di IA per finalità generali, la *governance*, le sanzioni (a esclusione di quelle pecuniarie per i fornitori sistemi di IA con finalità generali) e la riservatezza delle informazioni. Tutte le sanzioni – eccetto quelle indicate dall'art. 6, paragrafo 1, (legati ai sistemi ad alto rischio) che diventeranno effettive nell'agosto 2027 – si applicheranno a partire dal 2 agosto 2026[25].

I destinatari del Regolamento IA sono, quindi, i fornitori e gli utilizzatori (definiti "*deployer*") dell'IA appartenenti all'Unione Europea, quelli extraeuropei in cui il prodotto dei sistemi IA è destinato alla diffusione in UE, e gli importatori e i distributori di sistemi IA operanti da e per il territorio europeo.

Composto da un preambolo di 180 consideranda, 113 articoli e 13 allegati, il regolamento (UE) 2024/1689, noto più semplicemente come Regolamento IA o "AI Act", integra e modifica le precedenti disposizioni europee in materia, al fine di garantire un quadro normativo chiaro e armonico per la ricerca, la commercializzazione e l'operabilità d'uso dell'intelligenza artificiale (IA) nell'Unione Europea, compatibilmente ai valori e ai diritti fondamentali dei suoi cittadini.

Proprio in merito alla definizione di "sistema di intelligenza artificiale", il regolamento (UE) 2024/1689 stabilisce che è tale un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e con potenziale adattabilità dopo la diffusione

[23] Si tratta dei regolamenti: (CE) n. 300/2008, (UE) n. 167/2013, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828.

[24] 20 Giorni dopo la pubblicazione nella Gazzetta Ufficiale.

[25] Vengono introdotte, inoltre, nel capo XII del regolamento, anche sanzioni amministrative pecuniarie elevate che, per le imprese, ammonterebbero fino al 7% del fatturato annuo calcolato su base mondiale, mentre per le persone fisiche fino a 35.000.000 euro.

e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali[26].

3. Il disordine informativo e i rischi per lo Stato

L'importanza di contrastare i fenomeni del disordine informativo (di cui la disinformazione è solo una delle condotte più riconoscibili) è quella di evitare che parte di tali condotte, facenti parte, potenzialmente, del più ampio strumentario delle metodologie di guerra ibrida[27], rischino di compromettere tanto i diritti dei singoli cittadini quanto l'integrità stessa degli Stati[28]. Non si tratta, pertanto, solo di un problema di propagazione di semplici notizie false (comunemente anche dette *fake news*) ma di più articolate azioni e prodotti, di cui la mera falsità informativa può essere considerata solo, e non sempre, il minimo comune denominatore, senza però comprenderne la provenienza e la potenziale pericolosità.

Per questo, la tassonomia giuridica più aggiornata[29] sussume sotto il fenomeno del disordine informativo diverse condotte, a volte non necessariamente volontarie e dolose, che possono essere, escludendo il *genus* dei "*rumors*"[30], nelle cinque macro-categorie di:

- disinformazione,
- misinformazione,
- informazione malevola,
- malinformazione
- informazione improvvida o coperta da recentismo.

Nel caso della prima tipologia, ovvero della disinformazione "semplice" o "classica", si tratta di una condotta che prevede la volontaria creazione e/o diffusione dolosa di informazioni false, allo scopo di arrecare danni ad uno Stato o ad un sistema di Paesi, con canali informativi/mediatici integrati o congiuntamente facilmente influenzabili (in quest'ultimo caso, si può considerare che il risultato di una campagna riuscita di disinformazione verso, ad esempio, la Francia può originare una misinformazione in

[26] *Il futuro dell'Intelligenza Artificiale: pubblicato Il nuovo Regolamento europeo*, in *VegaEngineering*, 23 Luglio 2024.

[27] N. Bussolati, *The Rise of Non-State Actors in Cyberwarfare* in J. Ohlin-K. Govern-C. Finkelstein (eds.), *CyberWar: Law and Ethics for Virtual Conflicts*, Oxford, 2015, 102 ss.

[28] S. Cymutta-M. Zwanenburg-P. Oling, *Military Data and Information Sharing – A European Union Perspective*, in *Proceedings of the 14the Annual International Conference on Cyber Conflict*, 2022, 3 ss.

[29] Garante per la protezione dei dati personali, *Discorso sul disordine informativo*, in *garanteprivacy.it*, 2023.

[30] Termine utilizzato nel linguaggio giornalistico per indicare una voce o diceria che circola intensamente ma non è confermata in modo ufficiale. Si veda, in proposito, le definizioni di: *Rumor*, in *Vocabolario Treccani* oppure, nell'ambito finanziario, i "rumors" sono notizie e informazioni confidenziali non ufficiali che circolano nell'ambiente finanziario. Si tratta di notizie rilevanti ma difficilmente verificabili riguardanti operazioni e vicende di emittenti con titoli quotati, come ad esempio un aumento di capitale. La diffusione di tali notizie nasce dalla presenza di gap informativi tra i diversi soggetti che partecipano al mercato ed è finalizzata a sfruttare eventi ritenuti "*price sensitive*" non ancora ufficiali e destinati ad impattare, una volta confermati, sul valore delle azioni. Si rimanda a *Rumors*, in *Glossario Finanziario* di *Borsa Italiana.*

Italia o viceversa). Fino alla prima metà del XX sec. le operazioni di disinformazione sono state appannaggio quasi esclusivo degli apparati nazionali di informazione e sicurezza, che le adottavano per influenzare pro domo loro gli Stati satelliti o per "preparare il terreno" ad eventuali azioni belliche in Paesi ostili. Con l'evoluzione dei mezzi di comunicazione di massa (la diffusione, per scopi civili, del canale di Internet) e la comparsa di attori globali non nazionali e asimmetrici si sono sviluppate anche azioni di disinformazione trasversali e non nazionali, che possono perseguire fini criminali "terzi" e contrari rispetto alle politiche degli Stati o agire in loro appoggio ma senza dipenderne direttamente, in modo da mascherare gli obiettivi e di mimetizzarsi tra forme di informazione d'inchiesta e/o filantropiche.

Un'altra tecnica di disinformazione è la propaganda disinformante, che più che non tende solo alla pubblicizzazione di notizie edulcorate per influenzare l'opinione pubblica interna ma ha in quest'ultima il canale (e non l'obiettivo) di diffusione della falsa notizia[31], secondo il principio "se ci credono i nostri cittadini ci crederà anche il nemico".

La misinformazione, invece, consiste in una condotta che prevede la creazione e/o diffusione inconsapevole di informazioni false. In questo caso colui che diffonde o genera la notizia (magari, ad esempio, dopo aver visionato una fonte audio-visiva o traducendo la stampa estera) è in buona fede e pensa di rendere un servizio alla collettività, senza accorgersi che, invece, sta contribuendo a propagare disinformazione.[32] Si tratta, per questo, di un'azione che non comporta il dolo da parte degli attori principali (giornalisti, membri delle Istituzioni o informatori a vario titolo) ma che integra, a seconda delle fattispecie, le possibili responsabilità colpose per imperizia, culpa in vigilando e scarsa professionalità. Oltre agli strumenti "classici" per indurre la misinformazione (ad es. voci diffuse ad arte, false fonti di notizia fatte trovare ad improvvidi funzionari o giornalisti o personaggi mediaticamente influenti) e alla misinformazione indotta "di riflesso" per disinformazione altrui, grazie all'affinamento delle tecniche digitali di "effetti speciali" è nato -nell'ultimo decennio- lo strumento del *deepfake*, che consiste nell'alterazione, per rielaborazione informatica artificiale, di un contenuto audiovisivo, con un grado di verosimiglianza tale da essere difficilmente distinguibile dal vero per l'occhio umano. La misinformazione aumenta la propria efficacia in proporzione al grado di fama e affidabilità della fonte che, inconsapevolmente, la crea (tanto sarà noto e popolare l'autore della notizia tanto risulterà attendibile il contenuto della stessa; secondo il principio per cui "se lo dice lui/lei sarà vero").

L'informazione malevola, da *malicious information* ("informazione dannosa" o "i. maliziosa"), a sua volta, consiste nella diffusione volontaria e dolosa di notizie vere ma coperte da un regime di riservatezza (in Italia se ne distinguono quattro livelli: riservato "R", riservatissimo "RR", segreto "S" e segretissimo "SS"), allo scopo di creare conseguenze avverse e/o scredito nelle Istituzioni (principalmente governi) che le hanno segretate. Molto spesso tale condotta è perpetrata a seguito di altri atti di guerra ibrida come attacchi hacker ai server governativi e/o banche dati riservate, sottrazione mate-

[31] S. Gigante, *L'arte oscura della disinformazione: come si fa guerra alle fake news*, in *Agenda Digitale*, 16 dicembre 2024.

[32] O. Pollicino-P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections*, in *federalismi.it*, 2024.

riale di documenti di importanza strategica o corruzione di funzionari preposti[33].
Quest'ultima non deve essere confusa con la *malinformazione* che è un'informazione accurata ma diffusa con intento malevolo[34]. Si tratta di materiale sensibile che viene diffuso per danneggiare qualcuno o la sua reputazione[35]. Tra gli esempi vi sono il *doxing*, il *revenge porn* e l'*editing* di video per rimuovere contesti o contenuti importanti[36].
Infine, l'informazione improvvida o influenzata da recentismo è una quarta categoria che contribuisce ad incrementare il disordine informativo, che personalmente ritengo esser ben distinta dalle precedenti, è rappresentata dalla diffusione avventata di notizie (c.d. informazioni improvvide o influenzate da recentismo) vere ma parziali o non ancora completamente definite, rispetto ad avvenimenti complessi e molto recenti. Con l'affermazione delle piattaforme social e, in generale, del Web trai canali di diffusione delle notizie e di contenuti audiovisivi, il mondo dell'informazione è stato progressivamente influenzato, tanto nella metodica della ricerca delle fonti quanto nelle tempistiche narrative, tendendo a privilegiare sempre più la narrazione mediatica ed emozionale dei fatti rispetto al loro approfondimento critico contenutistico. La "corsa all'esclusiva" (ovvero a fornire per primi un'informazione per superare la concorrenza mediatica) origina, in contesti complessi e mutevoli (come nel caso dell'emergenza pandemica o degli eventi bellici d'Ucraina), informazioni parziali o contraddittorie con gli sviluppi successivi dei fatti o con gli stessi approfondimenti tematici, generando confusione e disorientamento nell'opinione pubblica, nonché senso di sfiducia verso le istituzioni e i canali d'informazione accreditati, lasciando così spazio al possibile insinuarsi di eventuali azioni di disinformazione o misinformazione indotta.
Per quanto azioni tecnicamente dissimili, considerata la poliedricità delle tecnologie di canale mediatico e l'interconnessione dei rapporti causa-effetto, spesso si possono riscontrare contemporaneamente le quattro condotte disinformanti, direttamente concatenate tra loro[37].

---

[33] P. Tettamanzi-M. Rijllo, *Cyber Insurance: strategia, gestione e governance*, Milano, 2024, 62 ss.

[34] N. Marquez, *Research Guides: Misinformation – Get the Facts: What is Misinformation?* in *guides.lib.uci.edu,* 16 marzo 2023.

[35] Da *What is disinformation?* in *Die Bundesregierung informiert, Startseite* (sito istituzionale tedesco), 23 marzo 2023.

[36] Si rimanda al report *Foreign Influence Operations and Disinformation* in *Cybersecurity and Infrastructure Security Agency CISA*, da sito istituzionale CISA.

[37] Esemplificativo può essere il seguente "caso di scuola". Si verifica un evento inaspettato, produttivo di sviluppi a lungo termine, il circolo mediatico reagisce fornendo subito notizie parziali, affrettandosi a contendersi le esclusive, le stesse informazioni però vengono smentite dai fatti successivi e dagli stessi mass media; si crea, pertanto, confusione nell'opinione pubblica, che genera sfiducia nelle Istituzioni e nei canali di informazione tradizionali. A tal punto, uno Stato ostile, che ha interesse a sfruttare e ad aggravare la situazione, immette altre notizie false o vere segretate (quindi sottratte e desegretate) aumentando così il disordine informativo. Alcuni comparti della libera informazione (sia nazionali che esteri) riprendono quest'ultime notizie, presentandole come verità e accreditandole con la loro attendibilità di testata/autore, diffondendole ulteriormente.
Come si potrà notare nell'esempio, costruito ad hoc (ma, comunque, riconducibile a specifici casi verificatisi), si intrecciano, concatenandosi, tutte le quattro forme di disordine informativo. Da una situazione iniziale di notizie affette da recentismo e non accuratamente accertate, si passa alla possibile disinformazione e malinformazione operata da una Potenza ostile, che se non opportunamente rilevata e contrastata (nel mondo di Internet diventa più difficile perché le notizie permangono e, come in un'Idra, si moltiplicano in svariati canali/forme) si estenderà ulteriormente ammantandosi di attendibilità grazie

Considerata l'estrema versatilità delle condotte di disordine informativo (sopradescritte) e la sottigliezza con le quali le stesse si possono combinare tra loro – per occultarsi, non essere rilevate e centrare così l'obiettivo – appare evidente come tutti i fenomeni disinformanti siano considerati un rischio per gli Stati, al centro delle agende internazionali di prevenzione e contrasto[38].

I rischi per lo Stato posso essere di diversa natura, andando da aspetti marginali e connessi solo alla mera informazione dei singoli cittadini (comunque sancita dal diritto costituzionale ad essere informati, ex art. 21 Cost.) alla destabilizzazione politica e sociale dell'intero Paese colpito.

La limitazione del diritto all'informazione generale, la manipolazione dell'opinione pubblica da parte di Paesi ostili, lo scredito sistematico delle Istituzioni, l'infiltrazione nel circuito dell'informazione e della sicurezza pubblica, la destabilizzazione politico-sociale, la compromissione, il furto e il danneggiamento dei dati veicolati nei circuiti informativi portanti e degli stessi sistemi di archiviazione e diffusione, sono solo i principali rischi – posti in un'evidente scala crescente – che lo Stato può correre se non contrasta efficacemente le campagne disinformanti[39].

4. L'apporto negativo dell'Intelligenza Artificiale al disordine informativo

Considerate la poliedricità e la pervasività delle condotte di disinformanti, che compongono lo spettro definitorio del disordine informativo, appare chiaro come la rete Internet e le tecnologie digitali correlate non possano che aumentare esponenzialmente i rischi per gli Stati e, più in generale, per le persone fisiche che li popolano.

Essendo lo sviluppo tecnologico incessantemente più veloce del procedere del mondo del diritto e, quindi, di qualsiasi forma di regolazione *ex ante*, le nuove tecnologie del mondo di Internet (digitali, algoritmiche o basate sul *machine learning*) costituiscono contemporaneamente una sfida e un elemento di ausilio per il legislatore.

È questo il caso anche dell'Intelligenza Artificiale (IA), tecnologia basata sia sull'apprendimento supervisionato che su quello non supervisionato delle macchine, che dalla seconda decade degli anni 2000 sta conoscendo una continua ascesa nelle applicazioni funzionali e nel dibattito scientifico[40].

Il mondo giuridico europeo ha cercato di trovare una definizione univoca dell'IA, indicando con il termine tutti «[…] quei sistemi che mostrano un comportamento intel-

ai media che la diffondono.

[38] Si veda sull'argomento, il report del 10 gennaio 2024 del Word Economic Forum, che le ha inserite tra le minacce più rilevanti per la stabilità dei Paesi e dello stesso sistema democratico-economico e valoriale occidentale.

[39] S. Giusti-E. Piras, Democracy and Fake News Information Manipulation and Post-Truth Politics, Londra, 2020.

[40] C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, 2019, spec. 48 ss e ancora, in materia di IA e protezione dei dati personali, F. Pizzetti, *Intelligenza artificiale e protezione dei dati personali: il ruolo del GDPR*, in *Diritto dell'informazione e dell'informatica*, 3, 2020, spec. 125 ss.

ligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere obiettivi specifici»[41]. Si tratta ovviamente di una definizione molto generica, risalente al 2021, che delinea solo vagamente le potenzialità e i rischi dell'Intelligenza artificiale.

Nel caso specifico, volendo tracciare l'apporto negativo (in quanto ulteriormente amplificatorio delle condotte dannose già in essere online) che l'IA potrà dare al disordine informativo, occorre prima presentare il vasto assortimento dei prodotti digitali dannosi che tale tecnologia crea o potrà ulteriormente potenziare.

Tra i più noti, direttamente considerati un prodotto dell'intelligenza artificiale, vi sono i *deep fake*, definiti nel 2020 dall'Autorità Garante per la Protezione dei Dati Personali come quei prodotti audio-visivi che «[…] sono foto, video e audio creati grazie a software di intelligenza artificiale (IA) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce»[42]. Si tratta, pertanto, di falsificazioni di foto, audio o video talmente fedeli e approfondite (come rimanda direttamente l'apposizione "*deep*") che solo un sistema minuzioso di IA può riuscire a creare[43]. Rispetto alle precedenti forme di modificazione artefatta (si pensi ai vecchi fotomontaggi o alle distorsioni del suono o ai tagli o annerimenti o sfuocamenti dei video) si tratta di prodotti molto più fedeli alla realtà e, per questo, molto più difficilmente distinguibili dall'occhio umano in assenza d ulteriori informazioni di contesto.

Ai *deep fake*, che possono essere considerati un prodotto fortemente interessato dall'apporto di IA, si devono aggiungere nel novero degli strumenti, potenziati da tale tecnologia, che incrementano l'insicurezza della navigazione internet e le condotte di disordine informativo: i *trolls informatici*, i *sock-puppets* e i *sealioners*.

Procedendo con ordine, i *trolls informatici*, che possono essere considerati la base concettuale anche delle altre due categorie (più raffinate e specializzate), sono tecnicamente degli utenti di Internet che interagiscono con gli altri con atteggiamento fastidioso e provocatorio per disturbare la normale convivenza delle community e dei social network, al fine di causare conflitti interpersonali e polemiche online[44]. Dietro ogni *troll*, attraverso un'identità pubblica falsa, si cela generalmente un utente reale che, protetto da uno pseudo anonimato operava indisturbato. Con l'avvento e l'implementazione dell'IA, adesso, anche i *trolls* informatici, codificatone il comportamento attraverso l'apprendimento automatico (c.d. *machine learning*), potrebbero essere gestiti da un sistema artificiale, con ulteriori problematicità per l'utente danneggiato e per le eventuali misure coercitive di prevenzione o inibizione (perché l'intelligenza artificiale può replicare innumerevoli volte e celermente le stesse condotte con profili ID nuovi

[41] Secondo il documento di sintesi *Preparare un giusto futuro l'Intelligenza artificiale e i Diritti fondamentali*, redatto dall'Agenzia dell'Unione europea per i diritti fondamentali (FRA), 2021.

[42] Garante per la Protezione dei Dati personali, *Deepfake. Il falso che ti «ruba» la faccia e la privacy*, 28 dicembre 2020.

[43] M. Cazzaniga*, Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in questa *Rivista*, 1, 2023, 172 ss.

[44] Definizione Enciclopedia Treccani.

e diversi).
Seguendo il medesimo schema dei *trolls*, l'IA può originare anche dei *sock-puppets* (lett. traducendo dall'inglese "uomini di paglia"), che in linguaggio informatico indicano quei profili informatici falsi creati da utenti di *social network* o altre comunità virtuali per ottenere, attraverso la contrapposizione alle loro finte e illogiche o deboli argomentazioni (spesso sbagliate e cattive), maggiore consenso e approvazione. Applicando l'intelligenza artificiale tali "uomini di paglia virtuali" potrebbero risultare ancor più difficili da riconoscere e molto più efficaci sia nel porre in essere condotte verosimili sia nell'auto duplicarsi e rapportarsi, con un'ulteriore distorsione della realtà.
I *sealioners*, invece, sono quei profili informatici falsi che fingono ignoranza o gentilezza mentre chiedono incessantemente risposte e prove (spesso ignorando o eludendo le prove già presentate) ad un utente vittima, con la scusa di "cercare solo di avere un dibattito" al fine di provocarlo a rispondere con rabbia, così da agire come parte lesa presentando il bersaglio come, ad esempio, persona chiusa e irragionevole. L'applicazione dell'IA, anche in questo caso non può che potenziare sia le capacità mimetiche di tali utenti che quelle dibattimentali, rendendo ancor più scientifico l'approccio provocatorio e, acquisendo collateralmente con precisione d'archiviazione informatica tutti i dati forniti nel dibattito dall'utente bersaglio. Informazioni, in questo caso sì reali e sensibili, che possono essere elaborate dallo stesso sistema di intelligenza artificiale *sealioner* per altri fini dolosi o comunque non consentiti dall'Ordinamento.
Per tutto questo e per le ulteriori implementazioni che, con il progresso tecnologico e l'apprendimento non supervisionato, possono sviluppare i sistemi di IA[45], appare evidente come tali tecnologie possano incentivare esponenzialmente il disordine informativo[46]. Se da una parte, infatti, i *deepfake*, in quanto profondamente (e accuratamente falsi) possono essere fonte diretta di disinformazione o di misinformazione, dall'altra *trolls*, *sock puppets* e *sealioner* concorrono come canali più o meno affidabili a diffondere informazioni false o parziali e a carpirne altre, anche riservate, produttive potenzialmente pure di condotte di mala informazione (o informazione maliziosa).
Per completezza, seppur non strettamente correlata all'apporto dell'IA al disordine informativo, non si può non menzionare la tematica della proliferazione dei *malware* alimentati dall'Intelligenza Artificiale, che rappresenta una minaccia emergente[47].
I *malware*, abbreviazione di "*malicious software*", sono qualsiasi software progettato per danneggiare, interrompere o ottenere accesso non autorizzato a sistemi informatici. Secondo IBM, il *malware* comprende vari tipi di software dannosi, tra cui *ransomware*, *trojan horse* e *spyware*. Tali strumenti se costruiti e supportati dall'intelligenza artificiale potranno utilizzare l'apprendimento automatico, basato su modelli informatici, per adattarsi costantemente e sono in grado di utilizzare avanzate tecniche di elusione per infiltrarsi nei sistemi. Tali *malware* "intelligenti" per questo, possono propagarsi autono-

[45] O. Pollicino, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2024.

[46] A. Mantelero, *Artificial Intelligence and Data Protection: Challenges and Opportunities for Regulation*, in *Computer Law & Security Review*, 5, 2018, 592 ss.

[47] L. Fritsch-A. Jaber-A. Yazidi, *An Overview of Artifi cial Intelligence Used in Malware,* in *Nordic Artificial Intelligence Research and Development* (NAIS 2022), 1 giugno 2022.

mamente e in modo cognitivo attraverso le reti ed essere in grado di personalizzare le proprie strategie di attacco in base all'obiettivo. L'IA può essere, quindi, utilizzata per sviluppare *malware* per le seguenti finalità:

- Ostacolare la rilevazione del codice del malware
- Eludere il rilevamento delle operazioni ("traffico") malevole
- Attaccare l'IA utilizzata per strategie difensive
- Rubare le credenziali o i fattori di identificazione dei dispositivi
- Sviluppare con l'auto apprendimento nuove forme di sabotaggio informatico e/o potenziare le tecniche già utilizzate (come nel caso del *phishing*).

Considerata, quindi, la rilevanza della portata positiva e negativa dell'Intelligenza artificiale nella società, non stupisce che le Istituzioni europee (Parlamento e Consiglio) abbiano delineato nel nuovo Regolamento IA (o AI Act) un'attenzione specifica all'apporto dell'IA verso la disinformazione e quindi le misure necessarie a ridurla.

5. Il regolamento europeo 1689/2024 e la problematica dei *deepfake*

L'obiettivo generale del Regolamento è quello di promuovere lo sviluppo di nuovi sistemi di IA e il conseguente mercato europeo, per una tecnologia affidabile, sicura e "antropocentrica", al fine di favorire una competitività responsabile e in grado di assicurare un livello elevato di protezione dei diritti umani[48], con anche la possibilità di interventi "protettivi" per bloccare gli effetti nocivi di altre forme di IA nell'UE. [49]
In linea generale, il Regolamento mira a garantire un'applicazione affidabile e sicura dell'intelligenza artificiale (IA), rispettando i valori e i diritti fondamentali dell'Unione Europea. Per raggiungere questo obiettivo, sono previste regole armonizzate per lo sviluppo, la diffusione e l'utilizzo dei sistemi di IA[50]. L'impianto definitorio dell'AI Act, pertanto, adotta un approccio basato sul rischio, suddividendo i sistemi di IA in quattro categorie:

- Rischio inaccettabile: Sistemi vietati.
- Rischio elevato: Sistemi soggetti a requisiti stringenti.
- Rischio basso: Sistemi per cui si applicano principalmente requisiti di trasparenza.
- Rischio minimo: Sistemi non soggetti a requisiti specifici

Sono previsti, quindi, alcuni divieti assoluti per livello di rischio "inaccettabile" (per i sistemi indicati nell'Allegato III del Regolamento) come quelli inerenti l'uso di sistemi di categorizzazione biometrica e gli obblighi di trasparenza e informazione, direttamente correlati al grado di rischio elevato come la combinazione della probabilità del verificarsi di un danno e la sua gravità, che va da quello inaccettabile, all'alto rischio

[48] V. Franceschelli, *Homo creator e responsabilità giuridica nell'intelligenza artificiale*, in *Diritto Industriale*, 5, 2024, spec. 78 ss.

[49] M. Bassini, *La regolazione dell'intelligenza artificiale nell'ordinamento europeo*, in *Rivista italiana di diritto pubblico comunitario*, 4, 2021, spec. 67 ss.

[50] G. Cassano-E.M. Tripodi, *Il Regolamento Europeo sull'Intelligenza Artificiale, Commento al Reg. UE n. 1689/2024*, Santarcangelo di Romagna, 2024, spec. 120 ss.

(sistemico, significativo o grave), da quello limitato al rischio minimo.
Precisamente è l'articolo 9 del Regolamento IA a definire le modalità di gestione del rischio, mentre l'art. 14 impone che gli stessi sistemi di IA rischiosi siano posti sotto supervisione umana, essendo progettati in modo da poter sempre consentire, durante l'utilizzo, il controllo remoto delle persone fisiche. A tal proposito, è bene segnalare che le tecnologie di IA per usi strumentali all'autorità giudiziaria e alle consultazioni democratiche (sistemi elettorali) sono considerate tra i sistemi ad alto rischio.
Più in generale, il Regolamento IA definisce "ad alto rischio" tutti quei sistemi che, per la loro natura, contesto di utilizzo e scopo, comportano potenzialmente rischi significativi per la salute, la sicurezza e i diritti fondamentali delle persone[51].
Proprio per tali particolari e sensibili ambiti di applicazione, il regolamento prevede che i sistemi di IA, definiti "ad alto rischio" siano soggetti a requisiti più stringenti in termini di conformità e monitoraggio.
I fornitori di tali prodotti saranno tenuti a potenziare dei programmi (*software*) di gestione del rischio che includa l'identificazione, l'analisi, la valutazione, la mitigazione e la gestione delle casualità correlate al loro completo utilizzo. A questo si aggiunge che i sistemi IA ad alto rischio devono essere rintracciabili tramite documentazione tecnica specifica, continuamente aggiornata, correlata da una valutazione d'impatto sui diritti fondamentali, nonché – per principio di trasparenza – devono essere corredati da informazioni sul corretto uso, sulla capacità o portata e sui limiti. Questo al fine di rendere le informative destinate agli utenti chiare e comprensibili e, al tempo stesso, per renderli edotti (qualora non se ne fossero resi conto) della loro interazione con un'intelligenza artificiale.
Oltre all'importanza della tracciabilità e della corretta informativa da fornire all'utente, il Regolamento IA 2024 presenta una particolare attenzione alla sicurezza e alla legalità della Rete, introducendo nel suo impianto normativo anche una forma di valutazione d'impatto sui diritti fondamentali, molto simile alla procedura di VIA (valutazione di impatto ambientale), in rapporto alle categorie di persone fisiche che saranno interessate dai sistemi IA, e un protocollo di etichettatura per evidenziare i contenuti falsi (*deepfake*), presenti online e prodotti o decriptati dalle stesse tecnologie artificiali.
Nonostante questo, i *deepfake* sono esplicitamente menzionati solo nella categoria dei sistemi a basso rischio.
L'Articolo 52(3) prevede, infatti, che gli utenti di un sistema di IA utilizzato per produrre *deepfake* debbano rivelare che i contenuti generati sono stati creati o manipolati artificialmente, introducendo così un obbligo legale a livello europeo per la loro identi-

[51] Sono, pertanto, tali i sistemi di intelligenza artificiale utilizzati:
- in infrastrutture critiche e strategiche, come energia, trasporti, acqua, gas e altre reti essenziali;
- nei contesti educativi, di istruzione e formazione professionale;
- nei servizi pubblici essenziali, come l'assistenza sanitaria, la previdenza sociale e i servizi finanziari
- per l'identificazione biometrica remota delle persone in spazi pubblici, come il riconoscimento facciale, utilizzati per scopi di sorveglianza e controllo.
- nella sicurezza dei prodotti immessi sul mercato, come i dispositivi medici e i veicoli autonomi
- per la manipolazione delle vulnerabilità delle persone (come l'età, disabilità, condizioni sociali o economiche)
- che attribuiscono punteggi sociali alle persone basati sul loro comportamento, caratteristiche personali o valutazioni che possono portare a discriminazioni.

ficazione. Inoltre, il considerando 38 e l'Allegato III[52] stabiliscono che l'uso di tecnologie per la rilevazione dei *deepfake* da parte delle autorità di polizia rientra nella categoria dei sistemi ad alto rischio, sottoponendoli a requisiti rigorosi.

Secondo il Regolamento IA, i *deepfake* «sono un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona»[53]. Tale definizione, riprende quella tracciata, nel 2020, dall'Autorità Garante per la protezione dei dati personali, accentuando maggiormente l'aspetto falsamente autentico del prodotto, generato da IA. Per questo, successivamente, il Reg. 1689/2024, all'art. 50, in merito agli Obblighi di trasparenza per i fornitori e i *deployers* di determinati sistemi di IA, al paragrafo 4. Prevede che «I *deployer* di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un deep fake rendono noto che il contenuto è stato generato o manipolato artificialmente. Tale obbligo non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati. Qualora il contenuto faccia parte di un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi, gli obblighi di trasparenza di cui al presente paragrafo si limitano all'obbligo di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera.»

Questa previsione appare perfettamente allineata al considerando 134 del Regolamento che prevede per i «*deployer* che utilizzano un sistema di IA per generare o manipolare immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi, entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri a una persona (*deepfake*), dovrebbero anche rendere noto in modo chiaro e distinto che il contenuto è stato creato o manipolato artificialmente etichettando di conseguenza gli output dell'IA e rivelandone l'origine artificiale.» Tale adempimento dell'obbligo di trasparenza, infatti, non viene previsto dal Regolamento come un ostacolo al diritto di libertà di espressione o a quello di libertà delle arti e delle scienze, soprattutto quando si rientra in un'opera o in un programma chiaramente creativo, satirico, artistico, immaginario o simile, fermo restando il rispetto dei diritti e delle libertà delle persone terze. In tali casi, l'obbligo di trasparenza relativo ai *deepfak*e, previsto dal Regolamento, impone solo che venga rivelata l'esistenza di contenuti generati o modificati, senza però pregiudicare la fruizione e l'esposizione dell'opera, compreso il suo normale sfruttamento e utilizzo, preservando nel contempo il valore e la qualità dell'opera stessa.

La stessa attenzione è posta dal AI Act, in sede di consideranda, dal inserendo un «obbligo di divulgazione analogo in relazione al testo generato o manipolato dall'IA nella misura in cui è pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico, a meno che il contenuto generato dall'IA sia stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica abbia la responsabilità editoriale della pubblicazione del contenuto».

Tuttavia, proprio rispetto ai *deepfake* si rilevano alcune criticità legate al suo ambito di applicazione e alla mancanza di disposizioni penali adeguate.

[52] Si rimanda a quanto previsto dal suddetto documento della Commissione Europea, 4.

[53] Precisamente ai sensi della definizione contenuta al n. 60 dell'art. 3 del Regolamento IA.

In *primis*, l'obbligo di identificazione previsto dall'Articolo 52(3) riguarda esclusivamente gli utenti dei sistemi di IA, non i produttori o i fornitori dei sistemi stessi. Ad esempio, nelle comuni applicazioni di *face-swapping*[54] che generano *deepfake*, i loghi dei produttori sono spesso integrati nei video generati come marchi di fabbrica. Qualora i produttori non prevedessero questa integrazione, gli utenti sarebbero costretti a contrassegnare autonomamente i contenuti deepfake o a indicare in un commento di accompagnamento che si tratta di contenuti manipolati. Ovviamente, questo problema sorgerebbe solo se i produttori non implementassero spontaneamente tali misure di contrassegno e un obbligo per i produttori e fornitori di integrare tali identificazioni eliminerebbe ogni incertezza.

Secondariamente vi sono alcune lacune legali nella lotta contro i *deepfake* dannosi. Il problema dell'identificazione, infatti, diventa critico quando si tratta di contrastare i *deepfake* distribuiti con finalità dannose (appunto condotte di disordine informativo al fine malevolo di disinformare o diffamare o ancor più destabilizzare una comunità). In questi casi, i contenuti vengono deliberatamente creati per non essere identificati come falsi o manipolati. Né i produttori né gli utenti di questi sistemi di IA avrebbero un incentivo a identificare volontariamente tali contenuti, creando una lacuna legale. La produzione e la distribuzione di software per *deepfake* che deliberatamente non includono strumenti di identificazione, favorendo così la frode, non risulterebbero punibili. Infine, non è chiaro come le autorità preposte all'applicazione della legge possano individuare deepfake non identificati distribuiti dagli utenti. Inoltre, l'AI Act non regola la diffusione di tecnologie di rilevazione dei *deepfake* basate sull'IA, che potrebbero essere utilizzate per ostacolare la diffusione e lo sviluppo di tali tecnologie a fini criminali.[55]

A tutto questo, si deve aggiungere che, pur essendo preminente per il Regolamento IA l'interesse di evitare che sistemi incontrollati di intelligenza artificiale possano potenziare o generare nuove forme di disinformazione[56], oltre ai rimandi ai *deepfake*, non sono menzionate nel testo altre forme esplicite di condotte o di opere di disordine informativo prodotte dall'AI. Grazie alla precisione e la pervasività della tecnologia, tali fenomeni disinformanti possono avere impatti negativi effettivi o prevedibili sui processi democratici, sul dibattito civico e sui processi elettorali, finanche dar luogo a dannosi pregiudizi e discriminazioni con rischi per gli individui, le comunità o le società.

[54] Si tratta di una il *faceswapping* (o "scambio di volti") è una tecnologia digitale che consente di sostituire il volto di una persona con quello di un'altra in un'immagine o video. È una tecnica che sfrutta algoritmi di intelligenza artificiale, come il *deep learning* e le reti neurali convoluzionali, per identificare, mappare e integrare i tratti del volto di un soggetto su un altro, mantenendo il realismo e le espressioni facciali coerenti con il contesto. Il sistema procede, in primis, con il riconoscimento facciale (un algoritmo individua e traccia le caratteristiche principali del volto, come occhi, naso, bocca e contorni), successivamente provvede a mappare le caratteristiche del volto sorgente, che vengono sovrapposte al volto target, adattandosi alla posizione, angolazione e illuminazione e, infine, provvede a ritoccare con un perfezionamento per rendere la fusione naturale, eliminando imperfezioni e garantendo coerenza con il resto dell'immagine o del video.

[55] M. Veale-F. Z. Borgesius, *Demystifying the draft EU artifcial intelligence Act*, in SocArXiv Papers, 3 agosto 2021.

[56] Come confermato dai consideranda 110 e 120 dello stesso atto regolatorio.

6. Conclusioni e prospettive

Gli obblighi di tracciabilità, trasparenza, limitazione d'uso, informativa all'utente e di eventuale supervisione e intervento umano, previsti dal Regolamento IA, sia per i fornitori che per gli utilizzatori della dei sistemi d'intelligenza artificiale, rappresentano pertanto, una prima forma di strumentario idoneo a limitare l'impatto nocivo di tale tecnologia in ambito di disordine informativo.

Si tratta, comunque, di principi generali ancora troppo sfumati e vaghi non direttamente incisivi su tutte le forme in essere di disinformazione mediante fonti (es. *deep fake*) o canali di diffusione (*trolls, sock puppets e sealioners*) generati da IA ma che, in qualche modo, vogliono essere un elemento normativo di partenza. Lo stesso fatto di aver incluso nel portato normativo del regolamento la definizione dei deep fake e degli obblighi per i responsabili dei sistemi di IA, può costituire un elemento favorevole per eventuale giurisprudenza di contrasto delle condotte disinformanti (come la misinformazione) originate dall'utilizzo scorretto o dolosamente indotto di ali tecnologie.

Le disposizioni normative delineate dal Regolamento IA, pertanto, pur rappresentando un importante passo avanti verso la regolamentazione dell'uso responsabile dell'intelligenza artificiale (IA), evidenziano alcune lacune che necessitano di ulteriori approfondimenti giuridici. La classificazione dei sistemi di IA per livello di rischio (inaccettabile, elevato, basso e minimo) rappresenta una solida base, ma il Regolamento dovrebbe affrontare in modo più specifico e incisivo le problematiche legate al disordine informativo, ampliando il suo focus oltre i deepfake.

Sebbene il Regolamento introduca obblighi di trasparenza per i *deepfake*, manca un quadro sanzionatorio chiaro che disciplini la responsabilità di produttori e fornitori di sistemi di IA utilizzati per creare contenuti manipolati con finalità dannose. L'introduzione di sanzioni specifiche per chi omette volontariamente l'identificazione dei contenuti manipolati potrebbe rafforzare l'efficacia della normativa, limitando l'uso improprio dell'IA.

Inoltre, l'ampliamento dell'obbligo di identificazione ai produttori e fornitori di *software*, non solo agli utilizzatori finali, potrebbe chiudere le lacune legali esistenti e prevenire abusi tecnologici.

A questo si aggiunge che, oltre ai *deepfake*, il Regolamento dovrebbe affrontare esplicitamente altre forme di disordine informativo generate dall'IA, come *troll* informatici automatizzati, *sock puppets* e *sealioners*, che svolgono un ruolo cruciale nella diffusione di informazioni manipolate. L'integrazione di linee guida specifiche per rilevare e contrastare questi fenomeni attraverso l'uso di IA "etica" potrebbe fornire un contributo significativo alla tutela dell'informazione pubblica.

Seguendo, quindi, il modello dell'UE, l'adozione di un modello di *governance* multilivello che coinvolga gli stati membri, le piattaforme tecnologiche e le istituzioni europee potrebbe facilitare una regolazione più dinamica e adattabile all'evoluzione tecnologica. La creazione di un centro europeo (accorpando gli enti esistenti ed *in fieri*) per il monitoraggio e la certificazione dei sistemi IA potrebbe garantire maggiore uniformità nell'applicazione delle norme, oltre a favorire lo sviluppo di tecnologie di rilevamento avanzate.

A tale proposito, si rileva che il rafforzamento delle procedure di valutazione di impatto sui diritti fondamentali dovrebbe essere reso obbligatorio per tutte le categorie di rischio, non solo per quelle ad "alto rischio". Tale valutazione dovrebbe includere l'analisi dei potenziali effetti negativi dell'IA sulla privacy, sulla libertà di espressione e sulla manipolazione dell'opinione pubblica, con un'attenzione particolare alle elezioni democratiche e alla stabilità sociale.

Infine, sul piano della promozione della trasparenza e dell'educazione digitale, il Regolamento dovrebbe enfatizzare ulteriormente l'importanza dell'alfabetizzazione digitale e della trasparenza, incoraggiando iniziative educative per rendere gli utenti consapevoli dei rischi associati ai contenuti manipolati dall'IA. L'implementazione di strumenti di verifica accessibili al pubblico potrebbe responsabilizzare gli utenti, favorendo una società più informata e resiliente.

In definitiva, quindi, il Regolamento IA rappresenta una base normativa promettente ma necessita di ulteriori specifiche per contrastare in maniera efficace le sfide del disordine informativo. Una maggiore integrazione di disposizioni sanzionatorie, misure preventive e strumenti di collaborazione internazionale potrebbe rendere il quadro regolatorio più incisivo nel garantire una società digitale sicura e rispettosa dei diritti fondamentali[57].

Sicuramente le tempistiche con cui entreranno in vigore tutte le misure previste dal Regolamento IA (specialmente quelle sanzionatori) lasciano un ampio limbo e margine temporale di manovra per gli attori disinformanti, che nel frattempo potranno i propri sistemi per sfuggire alle maglie della normativa ma, tuttavia, proprio per la generale astrattezza delle previsioni, li vincolano a andare oltre i fondamentali elementi dell'Intelligenza artificiale[58].

D'altra parte, l'interpretazione corretta dei principi tracciati dall'AI Act, potrà portare, in sede di ricerca e sviluppo di nuovi sistemi d'intelligenza artificiale, alla creazione di prodotti IA in grado di riconoscere i cattivi usi della stessa tecnologia, contribuendo così – con gli stessi mezzi e uguale precisione – a limitarne gli apporti negativi per la società[59].

Così come (secondo l'approccio statunitense) una notizia vera confuta efficacemente una falsa, un'intelligenza artificiale, utilizzata a fin di bene per la società, potrà contrastare un'IA impiegata per scopi illeciti.

[57] G. Resta, *Disinformazione e responsabilità civile nell'era digitale: il ruolo dell'intelligenza artificiale*, in *Rivista di diritto civile*, 2, 2022, 213 ss.

[58] C. Novelli et al., *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, in *arXiv*, 14 gennaio 2024.

[59] V. Calderonio, *The opaque law of artificial intelligence*, in *arXiv*, 19 ottobre 2023.

Elenco autori

Matilde Bellingeri
dottoranda di ricerca in scienze giuridiche europee e internazionali, Università degli Studi di Verona

Konrad Bleyer-Simon
research associate, Robert Schuman Centre for Advanced Studies, European University Institute

Manuela Luciana Borgese
dottoranda di ricerca in diritto della società digitale e dell'innovazione tecnologica, Università degli Studi di Catanzaro "Magna Graecia"

Luca Catanzano
dottorando di ricerca in Law, Economics and Social Sciences, Universidad de Burgos

Francesco Cirillo,
assegnista di ricerca in istituzioni di diritto pubblico, Università degli Studi della Tuscia

Federica Delaini
dottoranda di ricerca in scienze giuridiche europee e internazionali, Università degli Studi di Verona

Daniel Foà
assegnista di ricerca in diritto dell'economia, Università degli Studi di Bari "Aldo Moro"

Martina Iemma
dottoranda di ricerca in diritto costituzionale, Università degli Studi di Milano

Alberto Orlando
ricercatore in diritto pubblico, Università del Salento

Martina Palazzo
dottoranda di ricerca in risorse per la nuova PA, Università degli Studi di Milano-Bicocca

Matteo Paolanti
dottorando di ricerca in diritto costituzionale comparato, Università degli Studi di Pisa

Nicoletta Pica
assegnista di ricerca in diritto amministrativo, Università degli Studi del Sannio

Bruno Pitingolo
dottorando di ricerca in diritto costituzionale, Università degli Studi di Milano-Bicocca

Giuseppe Proietti
dottorando di ricerca in diritto commerciale, Università degli Studi di Roma TorVergata

Urbano Reviglio
research associate, Robert Schuman Centre for Advanced Studies, European University Institute

Lorenzo Ricci
dottorando di ricerca in diritto amministrativo, Università degli Studi della Tuscia

Andrea Ruffo
assegnista di ricerca in diritto costituzionale, Università degli Studi di Milano

Dora Trombella
dottoranda di ricerca in teoria dei diritti fondamentali, giustizia costituzionale, comparazione giuridica, diritto e religione, Università di Pisa

Sofia Verza
research associate, Robert Schuman Centre for Advanced Studies, European University Institute

CODICE ETICO

La **Rivista di diritto dei media** intende garantire la qualità dei contributi scientifici ivi pubblicati. A questo scopo, la direzione, il Comitato degli esperti per la valutazione e gli autori devono agire nel rispetto degli standard internazionali editoriali di carattere etico.

Autori: in sede di invio di un contributo, gli autori sono tenuti a fornire ogni informazione richiesta in base alla policy relativa alle submissions. Fornire informazioni fraudolente o dolosamente false o inesatte costituisce un comportamento contrario a etica. Gli autori garantiscono che i contributi costituiscono interamente opere originali, dando adeguatamente conto dei casi in cui il lavoro o i lavori di terzi sia/siano stati utilizzati. Qualsiasi forma di plagio deve ritenersi inaccettabile. Costituisce parimenti una condotta contraria a etica, oltre che una violazione della policy relativa alle submission, l'invio concomitante dello stesso manoscritto ad altre riviste. Eventuali co-autori devono essere al corrente della submission e approvare la versione finale del contributo prima della sua pubblicazione. Le rassegne di dottrina e giurisprudenza devono dare esaustivamente e accuratamente conto dello stato dell'arte.

Direzione: la direzione (ivi compresi direttori e vice-direttori) si impegna a effettuare la selezione dei contributi esclusivamente in base al relativo valore scientifico. I membri della direzione (ivi compresi direttori e vice-direttori) non potranno fare uso di alcuna delle informazioni acquisite per effetto del loro ruolo in assenza di un'esplicita autorizzazione da parte dell'autore o degli autori. La direzione è tenuta ad attivarsi prontamente nel caso qualsiasi questione etica sia portata alla sua attenzione o emerga in relazione a un contributo inviato per la valutazione ovvero pubblicato.

Comitato degli esperti della valutazione: i contributi sottoposti a valutazione costituiscono documentazione a carattere confidenziale per l'intera durata del processo. Le informazioni o idee acquisite confidenzialmente dai valutatori per effetto del processo di revisione non possono pertanto essere utilizzate per conseguire un vantaggio personale. Le valutazioni devono essere effettuate con profondità di analisi, fornendo commenti e suggerimenti che consentano agli autori di migliorare la qualità delle loro ricerche e dei rispettivi contributi. I revisori dovranno astenersi dal prendere in carico la valutazione di contributi relativi ad argomenti o questioni con i quali sono privi di familiarità e dovranno rispettare la tempistica del processo di valutazione. I revisori dovranno informare la direzione ed evitare di procedere alla valutazione nel caso di conflitto di interessi, derivante per esempio dall'esistenza di perduranti rapporti professionali con l'autore o la relativa istituzione accademica di affiliazione.

mediaLAWS

ANNO 2024 - CODICE ISSN 2532-9146