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.