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

1. La manifestazione del pensiero al centro della rivoluzione digitale. – 2. L'esperienza privata di moderazione del dibattito interno ai social network. – 2.1 Il gruppo Meta e il sistema misto di controllo. – 2.2 Il "*visibility filtering*" predisposto da X. – 2.3 L'approccio autogestionale: Reddit, 4chan e Truth. – 3. Le soluzioni pubbliche al problema. – 3.1 Gli interventi degli Stati sul tema: un passaggio sul Network Enforcement Act tedesco. – 3.2 La prospettiva europea per una costruzione di un ecosistema digitale a misura dell'Essere umano: DSA e AI Act - 3.2.1 Digital Services Act. – 3.2.2 AI Act – 3.2.3 Riscontri positivi e negativi di DSA e AI Act. – 4. Cosa resta dell'Uomo? L'educazione ai diritti fondamentali come soluzione tecnica e sociale.

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.