

Libertà d'espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale

di Carla Maria Reale e Marta Tomasi*

Abstract: *Freedom of expression, new media, and artificial intelligence: looking for a new balance in the constitutional ecosystem* – New media and artificial intelligence are challenging the traditional legal categories applied in the field of communication and information. This paper discusses how and to what extent artificial intelligence is changing freedom of expression as we know it, based on the assessment of three guiding principles in this area (truth and transparency, pluralism, and non-discrimination). The authors combine a more theoretical analysis with the discussion of some crucial examples and case-studies from a comparative law perspective. Finally, some general coordinates of the main opportunities and challenges involved in this field will be presented.

325

Keywords: New media; Artificial intelligence; Media pluralism; Transparency; Hate speech.

1. Il perimetro dei nuovi ruoli “tecnologici” nel mondo della comunicazione e dell'informazione

Pur con significative divergenze nella sua conformazione, la libertà di espressione costituisce uno dei punti nodali delle democrazie contemporanee. Essa rappresenta, secondo la lettura fornita da diverse giurisprudenze, la matrice, la condizione necessaria per l'esercizio di ogni altra libertà fondamentale¹, per il progresso delle società e lo sviluppo di ogni essere umano², il “più alto, forse,” dei “diritti primari e fondamentali”³, la

* Il seguente articolo è frutto di riflessioni condivise. Tuttavia, i paragrafi 1 e 2 sono attribuibili a Marta Tomasi, mentre il 3 e 4 a Carla Maria Reale. Le conclusioni sono state scritte congiuntamente dalle Autrici.

¹ Nel caso *Palko v. State of Connecticut*, 302 U.S. 319 (1937), Justice Cardozo ha definito la libertà di espressione come “the matrix, the indispensable condition, of nearly every other form of freedom” (par. 11).

² La Corte Europea dei Diritti Umani, nel caso *Handyside v. the United Kingdom*, deciso il 7 dicembre 1976, si riferisce al diritto in discussione come “one of the essential foundations of such a society, one of the basic conditions for its progress and for the development of every man” (par. 49).

³ Corte cost., sentenza n. 168 del 1971.

“pietra angolare dell’ordine democratico⁴”.

Trattandosi di un diritto che, nei decenni, ha dato prova di un forte dinamismo, adeguandosi al mutare del mondo della comunicazione e dell’informazione⁵, è imprescindibile, verificare se e come le tecnologie dell’Intelligenza Artificiale (di seguito IA) – che di recente hanno “catturato” ampia parte del dibattito giuridico – interagiscano o interferiscano con questo “cardine di democrazia nell’ordinamento generale”⁶.

Si tratta di un obiettivo ambizioso, condizionato dalla multiformità di entrambi i poli che costituiscono oggetto dell’analisi. Da un lato, infatti, è difficile poter ricondurre a unità le varie forme di IA che possono essere coinvolte nei processi comunicativi e di informazione⁷. Dall’altro, il diritto alla libertà di espressione si colloca all’intersezione di una pluralità di posizioni elementari più specifiche, quali la libertà di opinione, la libertà di comunicare e di esprimere il proprio pensiero, la libertà di informare e di essere informati. Ciascuna di queste, come si è accennato, è stata oggetto, nel corso dei decenni, di significative trasformazioni, dettate principalmente dal modificarsi e dall’evolversi dei mezzi di comunicazione che hanno fatto sì che le estrinsecazioni concrete di un diritto, si è detto, profondamente radicato nella storia costituzionale di molti ordinamenti, godano in realtà di confini tutt’altro che stabili e definiti.

Queste multiformità rendono quindi quantomeno arduo, nelle poche pagine di cui questo contributo si compone, riassumere tutte le possibili intersezioni tra AI e libertà di espressione. Una semplificazione che può aiutare, ai fini di una analisi sintetica, muove dalla distinzione fra due possibili funzioni o ruoli svolti dalle tecnologie dell’IA nel mondo della comunicazione e dell’informazione. Esse, infatti, possono alternativamente agire come soggetti che comunicano – quindi che *producono* discorsi o informazioni – o come soggetti che *selezionano* e *ordinano* tali contenuti, determinandone la circolazione e la diffusione.

Quanto al primo ruolo, che è ancora a un livello di sviluppo embrionale, si pensi all’editoriale pubblicato sul Guardian nel settembre del 2020, interamente scritto da un generatore di linguaggio basato sull’IA (*robot journalism*)⁸. Nell’articolo, il sistema rassicura gli umani rispetto al fatto che robot e IA vengano in pace e non abbiano alcuna intenzione di distruggere l’umanità⁹. Nonostante le inedite capacità del sistema impiegato, che dal punto di vista della coerenza degli argomenti presentati risulta molto avanzato rispetto ai suoi predecessori, gli studiosi di scienze cognitive

⁴ Corte cost., sentenza n. 84 del 1969. Già dal 1965, peraltro, era stata indicata come condizione “del modo di essere e dello sviluppo della vita del Paese in ogni suo aspetto culturale, politico, sociale” (sentenza n. 9 del 1965).

⁵ M. Orfino, *La libertà di espressione tra Costituzione e Carte europee dei diritti. Il dinamismo dei diritti in una società in continua trasformazione*, Torino, 2014.

⁶ Corte cost., sentenza n. 126 del 1985.

⁷ Sul punto si veda M. U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, 2, 2016, 359 e, in questo fascicolo, P. Traverso, *Breve introduzione tecnica all’intelligenza artificiale*.

⁸ Secondo la definizione fornita dal Guardian, GPT-3, il Sistema impiegato, “is a cutting edge language model that uses machine learning to produce human like text. It takes in a prompt, and attempts to complete it.”

⁹ L’editoriale è accessibile al link: www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

concordano con i computer scientists nel mostrarsi ancora profondamente scettici rispetto alle sue concrete capacità di comprensione dei messaggi veicolati¹⁰.

Pur trattandosi di esperimenti preliminari, lontani ancora da una sovrapponibilità piena con il linguaggio umano, si deve nondimeno rilevare il sempre crescente numero di algoritmi che lavorano per produrre output configurabili come discorsi (si pensi, per esempio, ai chatbot, come Alexa di Amazon, Siri di Apple, l'Assistente di Google e Cortana di Microsoft). E in effetti, alcune questioni relative alla possibilità di qualificare queste esternazioni come vere e proprie forme di espressione si sono già poste. In Francia, per esempio, alcuni giudici si sono trovati a dover ragionare circa la possibilità di far rientrare la funzione di completamento automatico nell'alveo di applicazione della libertà di espressione. La risposta negativa è derivata, in un caso, da una interpretazione restrittiva, che consentirebbe di riconoscere la libertà di espressione solo in capo alle persone¹¹ e, nell'altro, dall'idea che il completamento automatico rappresenti solo un'associazione "tecnica" di parole, non assimilabile a una manifestazione di pensiero¹². Diversamente, però, il Bundesgerichtshof, nel 2013, aveva ritenuto che le associazioni di parole, come quelle risultanti da un suggerimento di completamento automatico, implicassero un significato¹³.

Oltre questi episodi isolati, le questioni che interessano i *discorsi* generati algoritmicamente andranno verosimilmente ad aumentare e a complicarsi, via via che le forme espressive prodotte da sistemi autonomi si faranno più "intelligenti" e potenzialmente foriere di implicazioni inattese.

Il secondo ruolo cruciale svolto dalle tecnologie di AI nel mondo della comunicazione e dell'informazione è quello di "selettore e ordinatore" delle espressioni, opinioni e informazioni destinate alla diffusione sul web. In questo caso l'IA non genera i contenuti, ma li filtra, stabilendo se, come, quando e dove essi possano circolare nel mondo della comunicazione e dell'informazione digitale. Si pensi alla funzione degli algoritmi nella selezione, sui social media, dei contenuti considerati rilevanti, popolari, o offensivi, oppure all'intervento di tecnologie di IA sulla selezione e prioritizzazione delle notizie, attività che sempre più spesso influenzano le decisioni editoriali¹⁴.

Nei paragrafi che seguono si cercheranno di descrivere brevemente le possibili modalità di azione dell'IA nell'ambito della comunicazione e dell'informazione – con riferimento a entrambi i ruoli descritti –, focalizzando

¹⁰ G. Marcus, E. Davis, GPT-3, *Bloviator: OpenAI's language generator has no idea what it's talking about*, in *MIT Technology Review*, 22 agosto 2020, in www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.

¹¹ Tribunal de Grande Instance de Paris, 17ème chambre, 8 settembre 2010, M. X... /Google Inc., Eric S. et Google France.

¹² Court de Cassation, 1e civ, 19 giugno 2013, Google v. Lyonnaise de Garantie. Per un approfondimento sul tema: S. Karapapa, M. Borghi, *Search engine liability for autocomplete suggestions: personality, privacy and the power of the algorithm*, in *International Journal of Law and Information Technology*, 3, 2015, 261-289.

¹³ Bundesgerichtshof, VI ZR 269/12, 14 maggio 2013.

¹⁴ Si veda il documento: Council of Europe, *Artificial Intelligence – Intelligent Politics Challenges and opportunities for media and democracy*, febbraio 2020, 9. Disponibile al seguente link: rm.coe.int/cyprus-2020-ai-and-freedom-of-expression/168097fa82.

l'attenzione su quelli che sono tradizionalmente intesi, nelle riflessioni costituzionalistiche, come limiti all'esercizio della libertà di espressione.

2. Tecnologie dell'IA, veridicità e trasparenza di espressioni e informazioni

Le nuove tecnologie hanno inciso in maniera profonda sul modo di fare comunicazione e informazione e sulla veridicità delle stesse. Viviamo nell'era della post-verità, in cui le regole che presidiano i mezzi informativi tradizionali (i limiti al diritto di cronaca, fra i quali *in primis* quello della veridicità del fatto riportato¹⁵) difficilmente trovano applicazione. Nel mondo digitale, i problemi delle informazioni false, che costituiscono una questione vecchia “quasi quanto la storia dell'essere umano”, “sono diventati onnipresenti. Ne siamo sommersi”¹⁶.

In riferimento a questo aspetto, entrambi i descritti ruoli dell'IA risultano rilevanti. Da un lato, infatti, i sistemi di intelligenza artificiale sono parsi in grado – in alcune occasioni – di intervenire nella comunicazione sociale, orientando ora il dibattito politico, ora le scelte commerciali degli utenti. I casi più noti sono quelli che riguardano alcuni delicati eventi politici (come la Brexit o le elezioni negli Stati Uniti del 2016), a fronte dei quali si è diffusa la convinzione che il dibattito fosse stato intorbidito dall'intervento di sistemi automatici che avrebbero mimato il comportamento di esseri umani (c.d. *political bots*). Al di là del concreto impatto dei sistemi intelligenti nei singoli episodi, ciò che conta è che sta prendendo piede una tendenza alla “computational propaganda”, si starebbero cioè diffondendo comportamenti che – attraverso gli strumenti digitali – mirano a modulare l'opinione pubblica, spesso distorcendo la verità. Questi comportamenti, la cui attuazione un tempo era condizionata alla disponibilità di strumenti costosi e sofisticati, sono soggetti a sempre più evidenti processi di “democratizzazione”, determinati dal crescente grado di disponibilità ed economicità di bot o programmi¹⁷. Il riferito “effetto distorsivo”, peraltro, pare potersi rivolgere non solo al pubblico generale, ma interessare in futuro anche le comunità di esperti¹⁸.

D'altro canto – e quasi ironicamente – l'AI, nella sua funzione “selettiva”, rappresenta uno dei più promettenti argini proprio rispetto alla diffusione di informazioni distorte o di vere e proprie *fake news*¹⁹. Oggi, infatti,

¹⁵ A. Pace, *Art. 21*, in A. Pace, M. Manetti (a cura di), *Rapporti civili. La libertà di manifestazione del proprio pensiero*, in *Commentario della Costituzione*, Bologna-Roma, 2006, 334.

¹⁶ Come riconosce C.R. Sunstein, *On rumors: How Falsehoods Spread, Why We Believe Them, and What Can Be Done*, trad. it. *Voci, gossip e false dicerie. Come si diffondono, perché ci crediamo, come possiamo difenderci*, Milano, 2010, 11.

¹⁷ H. Murphy, *The new AI tools spreading fake news in politics and business*, in *Financial Times*, 10 maggio 2020, in www.ft.com/content/55a39e92-8357-11ea-b872-8db45d5f6714.

¹⁸ P. Ranade et al, *Study shows AI-generated fake reports fool experts*, in *The conversation*, 7 giugno 2021, in theconversation.com/study-shows-ai-generated-fake-reports-fool-experts-160909.

¹⁹ G. Marchetti, *Le fake news e il ruolo degli algoritmi*, in *MediaLaws – Rivista di diritto dei media*, 1, 2020 29-35.

gran parte della tecnologia utilizzata per identificare e gestire la disinformazione è alimentata dall'IA che consente (a giornalisti e *fact-checkers*, ma non solo) di verificare rapidamente grandi quantità di (dis)informazioni, che non sarebbero gestibili senza l'ausilio della tecnologia.

Sul punto, si può osservare che problemi sociali estremamente umani come la polarizzazione politica, la mancanza di pensiero critico, la relazione tra falsità e verità, sono questioni che difficilmente rispondono alla mera logica dei dati e delle correlazioni, e che – di conseguenza – non sembrano poter essere interamente creati, né interamente risolti attraverso l'uso della sola IA.

Oltre a quello della veridicità dei contenuti (oggetto di espressione o di informazione), un altro valore rilevante è quello della trasparenza. Esso, in estrema sintesi, rileva: a) in una dimensione “esterna”, imponendo che gli utenti delle tecnologie digitali siano informati della presenza e dell'operatività di tecnologie di intelligenza artificiale, e b) in una dimensione “interna”, più problematica, chiedendo che le decisioni assunte sulla base delle indicazioni fornite dai sistemi di intelligenza artificiale nel mondo della comunicazione e dell'informazione siano “spiegabili” nei loro processi. Ad oggi, non sono stati sviluppati strumenti utili a esaminare e rendere trasparenti i fondamenti tecnici delle decisioni automatizzate nella sfera online²⁰. Ciò significa che i diritti di espressione o di informazione delle persone (che dipendono, ad esempio, dalla moderazione dei contenuti e dai connessi meccanismi di reclamo, dal funzionamento dei motori di ricerca, dalla prioritizzazione delle notizie o dalla personalizzazione dei contenuti) potranno occasionalmente essere intaccati, senza che vi sia possibilità di indagare o comprendere a pieno su quali basi ciò possa essere avvenuto (il riferimento è al noto problema della *black-box*)²¹. Ad oggi, tuttavia, la proposta di Regolamento europeo nell'ambito dell'intelligenza artificiale sancisce un certo dovere di trasparenza²² per quanto concerne la progettazione ed il funzionamento dei sistemi di intelligenza artificiale (Art. 13), che certamente potrà trovare applicazione anche nell'ambito dell'informazione.

3. Tecnologie dell'IA e pluralismo dell'informazione

Allo stesso tempo lo sviluppo delle tecnologie di IA va di pari passo con il tramonto progressivo del ruolo centrale dell'editoria nell'informazione, con sostanziali ripercussioni nel modo in cui l'informazione viene prodotta, distribuita e utilizzata. Questo, oggi, ha svuotato di significato il legame - un tempo sinallagmatico - fra pluralismo dell'informazione e apertura e concorrenza del mercato editoriale, in una trama che si infittisce di nodi e

²⁰ M. Ananny, K. Crawford, *Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability*, in *New Media and Society*, 3, 13 dicembre 2016. Disponibile al seguente link: journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa.

²¹ F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge (MA)-Londra, 2015, 3 e ss.

²² Si veda in questo fascicolo C. Casonato, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, e M. Fasan, *I principi costituzionali nella disciplina dell'intelligenza artificiale. Nuove prospettive interpretative*.

ramificazioni, difficili da districare. Tale intreccio, pertinente soprattutto alla funzione selettiva/organizzativa dell'intelligenza artificiale, ha inciso in maniera parimenti significativa sia sui profili attivi, sia su quelli passivi del diritto all'informazione, toccando aspetti cruciali delle democrazie contemporanee.

Per quanto concerne il profilo relativo al diritto di cronaca, in quella che è stata oggi definitiva come *network information economy*, assistiamo ad una produzione radicalmente decentrata dell'informazione, cui ogni utente è chiamata/o a contribuire; una distribuzione ampia e capillare dei mezzi materiali di informazione e di produzione culturale grazie al basso prezzo dei dispositivi elettronici e all'esistenza di flussi di comunicazione con carattere aperto e globale²³. Positivamente (dal punto di vista del pluralismo), si parla di ricchezza delle reti, che può potenzialmente ampliare la sfera di libertà e la partecipazione alla vita democratica.

Nella sentenza *Sergejs Buivids*, la Corte di Giustizia dell'Unione Europea²⁴, con la valorizzazione della finalità della divulgazione al pubblico dell'informazione, ha assimilato, ai fini dell'applicazione della normativa alla protezione dei dati, la posizione del comune cittadino che diffonde informazione a quella di un/una giornalista. I fatti riguardavano un cittadino lettone che veniva fermato dalla polizia e condotto in caserma, ed in seguito al rilascio pubblicava sulla piattaforma Youtube alcuni video degli scambi con gli agenti. In questa sentenza la Corte ha statuito che, l'applicazione della deroga per finalità giornalistica del trattamento dei dati previste nella direttiva 95/46/CE (ora Regolamento UE 2016/679, General Data Protection Regulation), non potesse essere esclusa per il mero fatto che Buivids non fosse un giornalista professionista. Questo implica sicuramente la necessità di portare la riflessione ad alcuni profili di tutela che gli ordinamenti predispongono nei confronti delle/dei giornaliste/i, fra i quali, per esempio, quello dell'organizzazione come professione protetta, del segreto sulle fonti, della clausola di coscienza ecc. come anche quelle previsioni deontologiche e non solo, che disciplinano l'attività giornalistica.

Il c.d. giornalismo diffuso²⁵ presenta intrinseci profili di problematicità, che tuttavia non ne intaccano in toto il potenziale espansivo rispetto alla vivacità del dibattito democratico.

Questa propensione all'apertura può invece venire meno, in concreto, nel momento in cui passiamo dal profilo attivo a quello passivo del diritto all'informazione e interroghiamo la posizione dell'utente. Infatti, sebbene la produzione di informazione sia, ad oggi, in mano a molteplici soggetti e diffusamente in capo ai singoli, questa mole di informazioni viene comunque ordinata e resa accessibile da un ristretto gruppo di soggetti che possiedono la capacità di creare ponti e collegamenti tramite piattaforme, social network, motori di ricerca. Se la rete è aperta e ha moltiplicato le voci nello spazio pubblico, i paletti di questo spazio, il volume e la risonanza che le diverse

²³ Y. Benkler, *The Wealth of Networks. How Social Production Transforms Markets and Freedom*, Yale, 2006.

²⁴ Corte di Giustizia dell'Unione Europea, C-345/17, 14 febbraio 2019. Per un commento si veda F. Danini, *La nuova frontiera dell'informazione: giornalismo online e protezione dei dati personali*, in *DPCE online*, 2, 2019, 1749-1752.

²⁵ A. Lauro, *Siamo tutti giornalisti? Appunti sulla libertà di informazione nell'era social*, in *Media Laws – Rivista di diritto dei media*, 2, 2021, 1-24.

parole hanno, sono in mano ad una ridotta cerchia di soggetti con elevato potere economico (es: Google, Facebook, ecc.). Questo processo è stato descritto come un vero e proprio potenziale *gatekeeping*. In tale panorama sono due gli attori principali coinvolti nell'operazione di organizzazione e proposta di informazioni all'utenza: da una parte i motori di ricerca e dall'altra i social network.

Per quanto concerne i motori di ricerca questi fungono da tramite fra i bisogni informativi dell'utente e il materiale disponibile online, filtrando i contenuti e determinando l'ordine di presentazione degli stessi sulla base di ciò che è ritenuto maggiormente utile e confacente all'utenza. I social network, quali invece Facebook, ma anche Instagram (come luogo preferenziale del fenomeno dell'attivismo digitale), si fanno portatori di un consumo preterintenzionale²⁶ di informazioni, quello dell'utente che si imbatte "casualmente" in un contenuto informativo. In entrambi i casi, alla base della proposta e dell'organizzazione delle informazioni vi sono algoritmi elaborati da sistemi di Machine Learning che "personalizzano" l'informazione sulla base delle impronte che le singole persone lasciano sul web (*big data*). A tal fine, per esempio, Google utilizza RankBrain capace (ipoteticamente) di interpretare il senso della richiesta fatta dall'utente sulla base della conoscenza che ha della/o stessa/o, cioè sui suoi interessi e sulle sue preferenze; un meccanismo analogo è quello che utilizza Facebook per creare le "bacheche" (Newsfeed) di ciascun/a utente.

Questi funzionamenti sono alla base dei fenomeni conosciuti come "filter bubble"²⁷ o anche "echo chambers", che richiamano l'idea di un pluralismo ovattato, filtrato attraverso i gusti e le preferenze espresse da chi fruisce delle informazioni, che non viene mai esposto a sollecitazioni che fuoriescono dalla propria opinione precedentemente costituitasi, da stimoli innovativi.

Al di là della riflessione su come questi algoritmi possano di fatto incidere sull'esito di alcuni processi, già accennata nel precedente paragrafo²⁸, con a mente il valore del pluralismo, dobbiamo osservare come ci si trovi dinanzi un'utenza che si vede restituita dalle nuove fonti di informazione un'immagine fissa, statica, somigliante al proprio riflesso. Un meccanismo ormai consolidato al punto che in dottrina si è avanzata l'idea di poter elaborare, a fronte di ciò, un diritto del singolo alla discontinuità e all'incoerenza²⁹. Ma ancora, una persona che si fa individuo avulso rispetto al pubblico confronto e al *Government by discussion*³⁰, che dovrebbe

²⁶ G. Pitruzzella, *La libertà di informazione nell'era di Internet*, in *Media Laws – Rivista di diritto dei media*, 1, 2018, 25

²⁷ Il concetto è stato elaborato per la prima volta da Elis Pariser. Si veda: E. Pariser, *Filter Bubble How in the New Personalized Web Is Changing What We Read and how We Think*, New York, 2011.

²⁸ Per approfondire si veda M. Fasan, *Intelligenza Artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico*, in A. D'Aloia (a cura di), *Intelligenza artificiale e diritto*, Milano, 2021, in particolare 306-307.

²⁹ Così C. Casonato, relazione *The Rise of New (and old) Rights in the Age of AI*, presso l'incontro inaugurale del corso *Constitutional Law of Technologies*, Firenze, 6 ottobre 2021 ed in questo fascicolo C. Casonato, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*.

³⁰ Il termine è inizialmente coniato da Mill, ma il concetto viene ripreso da molti pensatori e pensatrici nella nostra epoca. Si pensi all'elaborazione di A. Sen, *Collective*

caratterizzare le moderne democrazie, e diviene invece incline alla polarizzazione e alla radicalizzazione³¹.

I processi democratici virtuosi muovono all'interno di una prospettiva di empowerment collettivo, coltivando la capacità dei singoli e delle loro forme di organizzazione di incidere sui processi decisionali tramite il dibattito pubblico. Questo presuppone tuttavia un ruolo cruciale dell'educazione e dell'informazione, strumenti di crescita e formazione di coscienza critica individuale e collettiva che garantiscono che ciascuna/o possa contribuire sostanzialmente al dibattito pubblico. Tale ruolo è fortemente minacciato oggi, a causa di un pluralismo informativo opaco, che paventa la via verso quella che è stata definita *bubble democracy*³².

4. Tecnologie dell'IA e principio di eguaglianza e non discriminazione

Molti degli aspetti che ruotano attorno al rapporto fra diritti fondamentali e nuovi media nelle democrazie contemporanee si intrecciano e convergono su quello che costituisce, come vedremo, un vero e proprio nodo da districare: il rispetto del principio di eguaglianza e non discriminazione, principalmente in merito al ruolo dell'AI nella selezione e moderazione dei contenuti online.

Una considerazione preliminare necessaria parte dall'assunto che l'intelligenza artificiale distingue. Lo fa perché programmata per raccogliere dati e operare scelte e distinzioni sulla base delle informazioni selezionate. Nel distinguere è accertato, ormai, che essa sia capace di discriminare e che lo faccia a partire dai bias e dagli stereotipi di chi la programma, ma seguendo poi processi difficilmente decifrabili. Questo accade ai danni di soggetti e gruppi di soggetti che già si collocano in assi d'oppressione all'interno delle nostre società, ma anche a scapito di "nuove minoranze" e sulla base di nuovi fattori di discriminazione, che indubbiamente necessitano di essere guardati tramite una lente intersezionale³³. Le discriminazioni sottese all'AI sono reputate insidiose perché "difficili da prevenire; difficili da rilevare; difficili da sanare, annidate come sono nelle pieghe degli algoritmi difficilmente ricostruibili dai loro stessi artefici³⁴". Lungi dal promuovere decisioni più giuste e imparziali, molti casi di discriminazione conclamata generati da tecnologie di *deep learning* e *machine learning* sono stati largamente documentati e discussi, fra cui per esempio quello riguardante il sistema di reclutamento di Amazon che penalizzava le donne³⁵, o ancora il sistema

Choice and Social Welfare, Londra, 1970.

³¹ Parlamento Europeo, *Polarisation and the use of technology in political campaigns and communication*, marzo 2019, in [www.europarl.europa.eu/RegData/etudes/STUD/2019/634414/EPRS_STU\(2019\)634414_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634414/EPRS_STU(2019)634414_EN.pdf).

³² D. Palano, *Bubble democracy: la fine del pubblico e la nuova polarizzazione*, Brescia, 2020.

³³ C. Nardocci, *Intelligenza artificiale e discriminazioni*, in *La rivista "Gruppo di Pisa"*, 5 ottobre 2021, 9.

³⁴ M. Luciani, *Relazione Conclusiva*, in *La rivista "Gruppo di Pisa"*, 1 ottobre 2021, 6.

³⁵ La notizia fu in primo luogo diffusa da J. Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, in *Reuters.com*, 11 ottobre 2018. L'articolo è disponibile al seguente link www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

COMPAS di predizione della recidiva utilizzato negli Stati Uniti che si è rivelato essere razzista nei confronti delle persone Afroamericane³⁶. Per questo motivo, la contingenza di riaffermare il principio di non discriminazione nell'ambito delle nuove tecnologie e dell'intelligenza artificiale è stata più volte e in diverse sedi affermata, si veda per esempio l'operato della European Commission for the Efficiency of Justice (CEPEJ) che ha sottolineato come l'emersione e la cristallizzazione delle discriminazioni presenti nella società tramite l'uso di AI debba essere contrastata. Inoltre, la già citata Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale, nei suoi considerando fa più volte riferimento all'attuazione del principio di non discriminazione in questo settore (cfr. 1.2; 3.5) anche se, solo parzialmente, affronta il tema con riguardo ai dati all'articolo 10 della proposta.

Anche all'interno dei motori di ricerca e dei social media, dove l'IA esercita il proprio ruolo di selezione dei contenuti espressivi e informativi, sono stati mostrati analoghi processi discriminatori, per esempio nell'ambito pubblicitario. L'autorità olandese per la protezione dei dati ha, per esempio, accertato una violazione delle leggi nazionali sulla protezione dei dati da parte di Facebook, documentando, fra le altre cose, come l'azienda consentisse la targetizzazione della pubblicità sulla base di dati sensibili, fra i quali le preferenze sessuali³⁷, allo stesso modo in cui ProPublica ha mostrato che in questi processi di *advertising* taluni ambiti escludevano gruppi razzializzati, quali persone ispaniche e nere³⁸. Altro aspetto riguarda invece il meccanismo di selezione della visibilità dei contenuti e dell'oscuramento degli stessi all'interno delle piattaforme tramite, per esempio, il meccanismo dello Shadow Ban, che pur non cancellando determinati contenuti ne limita la visibilità, sulla base di criteri non esplicitati, impedendo che gli stessi raggiungano una ampia platea. Il think-tank Australian Strategic Policy Institute (ASPI) ha mostrato come molti hashtag connessi alla comunità e alle questioni LGBT+ siano stati oscurati da TikTok in Bosnia, Giordania e Russia³⁹, in un peculiare intreccio con le questioni geopolitiche sottese ai diritti umani. Oppure, si pensi al movimento

³⁶ Per una spiegazione dei fatti e dell'indagine, a partire dal lavoro di ProPublica condotta al riguardo si veda: W. Dietrich, C. Mendoza, T. Brennan, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, 8 luglio 2016, in go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf; Per un commento della dottrina italiana: G. Contissa, G. Lasagni, G. Sartor, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Diritto di internet*, 4, 2019, 621.

³⁷ Autoriteit Persoonsgegevens, 23 febbraio 2017. Una sintesi delle conclusioni del rapporto è disponibile in inglese a questo link: autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/conclusions_facebook_february_23_2017.pdf.

³⁸ Si veda J. Angwin, T. Parris, *Facebook Lets Advertisers Exclude Users by Race*, in *ProPublica.org*, 28 ottobre 2016. Disponibile al seguente link: www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

³⁹ F. Ryan, A. Fritz, D. Impiombato, *TikTok and WeChat. Curating and controlling global information flows. Report No. 37/2020*, settembre 2020. Il rapporto è disponibile al seguente link: s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-09/TikTok%20and%20WeChat.pdf?VersionId=7BNJWaoHImPVE.6KKcBP1JRD5fRnAVTZ.

online *#Freethenipple*, che lamentava i diversi standard di censura fra capezzoli maschili e capezzoli femminili, in seguito al quale Facebook ed Instagram hanno cessato di rimuovere foto di donne che allattano al seno o di mastectomie⁴⁰.

Un'altra questione particolarmente dibattuta nell'ambito dell'informazione riguarda i discorsi d'odio online definibili come "l'istigazione, la promozione o l'incitamento alla denigrazione, all'odio o alla diffamazione nei confronti di una persona o di un gruppo di persone, o il fatto di sottoporre a soprusi, molestie, insulti, stereotipi negativi, stigmatizzazione o minacce tale persona o gruppo"⁴¹ sulla base di fattori quale sesso, genere ed identità di genere, orientamento sessuale, razza/etnia, religione, disabilità, età. Particolarmente controverso è il ruolo che le piattaforme quali Twitter, Instagram, Facebook ecc. hanno e dovrebbero avere nelle rispettive comunità digitali nei confronti di simili condotte. Parte di questo dibattito, che in questa sede non può essere approfondito, riguarda la natura di queste piattaforme, che pur avendo un carattere eminentemente privato si trovano a svolgere funzioni di rilevanza pubblica. Si pensi alla qualificazione di Twitter come *public forum operata* da Justice Thomas nella *concurring opinion* alla sentenza della Corte Suprema degli Stati Uniti sul caso riguardante a la moderazione dei commenti nell'account dell'allora presidente Donald Trump⁴². A prescindere da ciò, si discute molto circa le possibilità di azione e i doveri degli Internet Service Provider di fronte alla diffusione di discorsi d'odio. All'interno dell'Unione Europea la disciplina di riferimento è quella contenuta nella direttiva 2000/31/CE che pur in assenza un obbligo specifico di sorveglianza in capo ai providers, limato tuttavia dalla giurisprudenza della CGEU, stabilisce l'obbligo di rimozione di contenuti in presenza di un ordine dell'autorità pubblica. Nel 2016, inoltre, la Commissione UE ed i principali internet service providers hanno sottoscritto un Codice di condotta per il contrasto all'hate speech che sottolinea la necessità che siano le piattaforme stesse ad autoregolarsi e sorvegliare le proprie community, adottando delle linee-guida indirizzate all'utenza. Non solo queste devono adottare procedure chiare ed efficaci per esaminare le segnalazioni provenienti da utenti, ma si impegnano ad analizzare le stesse entro 24 ore ed eliminare, ove necessario, i contenuti nel medesimo lasso di tempo. In questo campo è ancora fortemente discusso in qual misura e come il diritto debba operare, si pensi per esempio alla recente legge francese sui discorsi d'odio che imponeva l'obbligo di rimozione dei contenuti di hate speech e discriminatori entro 24 ore da una qualsiasi segnalazione prevedendo una sanzione a presidio di tale obbligo, dichiarata in larga parte incostituzionale dal *Conseil Constitutionnel* perché reputata non

⁴⁰ T. Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, New Haven, 2018.

⁴¹ Raccomandazione di politica generale n.15 della Commissione europea contro il razzismo e l'intolleranza del Consiglio d'Europa (ECRI) del 21 marzo 2016.

⁴² M. Monti, *La Corte Suprema statunitense e il potere delle piattaforme digitali: considerazioni sulla privatizzazione della censura a partire da una concurring opinion*, in *DPCE online*, 2, 2021, 2781-2791; L. Rinaldi, *Le piattaforme tra diritto pubblico e diritto privato. Libertà di espressione, discorso politico e social network in alcuni casi recenti tra Italia e Stati Uniti*, in *La Rivista "Gruppo di Pisa"*, Quaderno monografico 3, 2021, 211 ss.

proporzionata in relazione alla compressione della libertà di espressione⁴³. Il rischio, chiaramente è che gli strumenti algoritmici per contrastare il dilagare di forme di odio online, in assenza di un *online due process*⁴⁴, vengano utilizzati per silenziare forme di espressione che rientrano invece nell'ombrello dei diritti nell'ambito dell'informazione, o che gli stessi meccanismi di moderazione della comunità online producano e riproducano diseguaglianze e discriminazioni⁴⁵.

5. Verso un'espressione di "forma" tecnologica e di "senso" umano

Le tecnologie di IA oggi, nella loro duplice funzione da un lato di produzione e dall'altro di selezione, moderazione e ordine dei contenuti, difficilmente possono essere considerate come meri strumenti attraverso i quali si attua la libertà di espressione e informazione; essi, piuttosto, si qualificano quali vettori di potenziali cambiamenti strutturali dei paradigmi del mondo della comunicazione e dell'informazione, essendo portatrici di mutamenti di carattere non solo quantitativo, ma anche e soprattutto qualitativo, essendo capaci di incidere attivamente su contenuti, significati, connessioni e ordini di priorità.

Ne consegue un dibattito vivo, del quale si sono tracciate solo alcune direttrici di fondo, che si snoda all'intersezione fra quesiti di natura giuridica (relativi per esempio alla possibilità di individuare nuovi titolari diritti o di rideterminare riparti di competenze e responsabilità) e tecnica (di comprensione di meccanismi di funzionamento, selezione e ordinamento di contenuti e informazioni).

La tecnologia in questione, tuttavia, non è intrinsecamente caratterizzata da valori assoluti e non si presta a essere schiacciata in giudizi dicotomici (quali "bene/male", "positivo/negativo"), abbisognando prima di tutto di un'analisi che voglia cogliere la complessità del fenomeno. Come è emerso, l'intelligenza artificiale nel mondo della comunicazione e dell'informazione può agire in modo ambivalente: essa, infatti, ha il potenziale per essere una chiave di promozione per il diritto alla comunicazione, ampliando le arene e la loro risonanza e democratizzando i processi alla base di creazione dei contenuti, e al diritto all'informazione, aumentandone le fonti, selezionando contenuti rilevanti e dando preferenza a informazioni veritiere. Al contempo tuttavia, i medesimi meccanismi, possono produrre un eccesso di informazione (frammentata o contraddittoria) o un'informazione univoca, incentivare la diffusione di contenuti discriminatori ed ostili, determinare censure ingiustificate o erronee, produrre manipolazioni o distorsioni del dibattito pubblico e persino di quello scientifico.

È bene peraltro tenere a mente che la maggior parte dei timori legati alle turbative che l'IA potrebbe produrre nella comunicazione e nell'informazione riguardano oggi il perseguimento di logiche e fini che di

⁴³ Décision n° 2020-801 DC du 18 juin 2020.

⁴⁴ M. Husovec, *Why There is No Due Process Online?*, in *Balkanization*, 7 giugno 2019.

⁴⁵ Si veda per esempio il caso studio relativo all'algoritmo di Instagram in prospettiva di genere riportato nel seguente articolo: C. Are, *How Instagram's algorithm is censoring women and vulnerable users but helping online abusers*, in *Feminist Media Studies*, 20, 2020, 741-744.

“artificiale” hanno ben poco: essi sono radicati in questioni profondamente umane, quali quelle della propaganda politica, il fine commerciale e gli scopi di profitto, la perpetrazione di pregiudizi e stereotipi. L’intelligenza artificiale, insomma, in quanto strumento creato dall’essere umano, non è una tecnologia neutra né neutrale, ma è piena dei significati che come singoli e come collettività vogliamo conferirle.

Il problema più delicato, oggi, non sembra tanto quello di una macchina che tenti, nelle sue espressioni, di emulare l’essere umano, quanto quello di un essere umano che si arrenda alle logiche della tecnologia e della macchina nell’esercizio di un diritto che da sempre sta al cuore della democrazia. Questo rischio è reso evidente, per esempio, dall’emersione di poteri “tecnologici” che fuoriescono dagli assi su cui il costituzionalismo moderno si è incentrato (esecutivo, legislativo, giudiziario). Sono poteri, la cui natura giuridica è strettamente privata, che intrecciano la capacità di immagazzinare, gestire ed utilizzare una mole significativa di dati con ingenti disponibilità di denaro e con l’obiettivo di generare profitto. Questi soggetti sono in grado di condizionare profondamente l’esercizio di diritti di comunicazione e informazione, i quali invece sono stati costruiti tradizionalmente come libertà negative e nell’ottica di una tutela dall’ingerenza del potere pubblico. Sarà dunque necessario ragionare, soprattutto, circa la necessità di risignificare la dicotomia poteri pubblici/poteri privati, che nella realtà odierna presenta innumerevoli sfumature e non si presta più ad un ragionamento binario⁴⁶.

336

Più in generale, il rischio è che il mondo della comunicazione e dell’informazione si automatizzi, alternativamente si standardizzi o si frammenti in una serie di bolle non interoperabili, privandosi di quel dinamismo e di quella forza trasformativa necessari a che la libertà di espressione e informazione continui a rappresentare un motore di crescita individuale e collettiva.

Il modo in cui utilizziamo le tecnologie dell’AI, l’impronta e l’indirizzo che poniamo alla base delle loro capacità di auto-apprendimento ed elaborazione sono dunque cruciali per la costruzione di un nuovo ecosistema all’interno del quale la libertà di espressione e informazione possa mantenere la propria funzione di veicolo per lo sviluppo di orizzonti di significato che sono, ancora, profondamente umani.

Carla Maria Reale
Facoltà di Giurisprudenza
Università degli Studi di Trento
Carlamaria.reale@unitn.it

Marta Tomasi
Facoltà di Giurisprudenza
Università degli Studi di Trento
Marta.tomasi@unitn.it

⁴⁶ Cfr. C. O’Neil, *Armi di distruzione matematica*, Firenze-Milano, 2017 ed anche al concetto di capitalismo della sorveglianza elaborato da S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell’umanità nell’era dei nuovi poteri*, Roma, 2019.