ISSN 1826-3534

federalismi.it
RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO

EDITORIALE - 29 MAGGIO 2024

Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)

di Oreste Pollicino
Professore ordinario di Diritto costituzionale
Università Commerciale "L. Bocconi"

e Pietro Dunn
Research Fellow
Università Commerciale "L. Bocconi"

Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)

di Oreste Pollicino
Professore ordinario di Diritto costituzionale
Università Commerciale "L. Bocconi"

e Pietro Dunn
Research Fellow
Università Commerciale "L. Bocconi"

1. Introduzione

L'anno in corso sarà quello "più elettorale" di sempre: oltre 50 elezioni nel mondo, alle urne 76 Paesi, chiamate al voto 2 miliardi di persone (solo in Europa, tra qualche giorno, in 400 milioni). Non ci si può permettere di sottovalutare i rischi che possono emergere dall'amplificazione delle tecniche di disinformazione attraverso l'impiego di quell'ecosistema – non ci si stancherà mai di ribadire che non è soltanto una tecnologia – costituito dall'intelligenza artificiale.

Per iniziare, si può provare a trarre qualche insegnamento da quanto già accaduto.

La rappresentazione di internet quale «*new marketplace of ideas*», elaborata dalla Corte Suprema americana in una sentenza del 1997[1], la quale a sua volta adatta al mondo dei bit la leggendaria espressione alla base della *dissenting opinion* di Holmes nel 1919[2], è stata, per lungo tempo, la metafora preferita da parte di chi ha ritenuto che il fenomeno della disinformazione online potesse essere risolto grazie alle capacità autocorrettive del mercato delle idee. Purtroppo, quando nel 2018 la Commissione, nella sua prima strategia contro la disinformazione, ha di fatto importato tale idea applicandola al contesto valoriale europeo – assai differente rispetto a quello americano – la conseguenza è stata l'esito quasi fallimentare del primo codice di condotta contro la disinformazione, esclusivamente fondato su una logica di autoregolazione.

Si tornerà sul punto: intanto può già anticiparsi che se l'innesto all'interno dell'*humus* valoriale europeo della metafora del *free marketplace of ideas*, propria del costituzionalismo americano (e che peraltro

[1] Corte Suprema degli Stati Uniti, *Reno c. ACLU*, 521 US 844 (1997), p. 885.
[2] Corte Suprema degli Stati Uniti, *Abrams c. Stati Uniti*, 250 US 616 (1919), pp. 624-631.

all'interno dello stesso dibattito d'oltreoceano sulla lotta contro la disinformazione sta entrando in crisi), ha provocato una una crisi di rigetto nel contesto della tecnologia ormai in qualche modo datata di internet, si avrebbe un risultato ancor più deludente ed un rigetto ancor più evidente (e questo è stato ben compreso da parte delle istituzioni europee) se la stessa operazione di "importazione" fosse compiuta in riferimento al cocktail esplosivo disinformazione/intelligenza artificiale (IA), soprattutto durante la stagione elettorale che stiamo vivendo[3].

Cerchiamo di guardare più da vicino agli ingredienti (del cocktail): inizieremo, in particolare, dalla disinformazione.

2. Disinformazione, misinformazione, *fake news*

Al fine di comprendere l'effettiva portata dell'impatto dell'IA(anche) sul fenomeno della disinformazione, un aspetto preliminare di particolare rilevanza appare essere quello relativo alla stessa definizione del fenomeno. Invero, nonostante sia invalsa sempre più nel linguaggio comune la locuzione inglese facente riferimento alle cosiddette "*fake news*", l'utilizzo di tale espressione è in anni recenti andato sempre più incontro a critiche non solo all'interno del contesto accademico ma, soprattutto, in ambito istituzionale.

In tal senso si è espresso, in particolare, l'*High Level Group on fake news and online disinformation* (HLEG) – cui uno dei due autori di questo contributo ha partecipato – istituito dalla Commissione europea, il quale, nel suo rapporto finale[4] pubblicato nel 2018, ha suggerito l'utilizzo dell'espressione disinformazione in luogo della locuzione "*fake news*" essenzialmente per due motivi. In primo luogo, si è rilevato come il concetto di "notizie false" rischi di essere riduttivo in quanto esclude tutta una serie di manifestazioni del fenomeno[5]; in secondo luogo, l'utilizzo mediatico e popolare della stessa espressione "*fake news*" ha nel tempo condotto a una connotazione fuorviante del fenomeno, soprattutto alla luce dell'uso strumentale che ne è stato fatto da parte di alcune personalità politiche e dei loro sostenitori al fine di screditare e respingere notizie, informazioni e opinioni non volute

A fronte di ciò, appare preferibile, come si accennava, il ricorso al concetto di "disinformazione", che, secondo l'HLEG, ricomprende tutte quelle forme di manifestazione della libertà che si sostanziano di

[3] Per un'analisi più approfondita si rimanda a O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Bocconi University Press, Milano, di prossima pubblicazione (2024), con prefazione di L. VIOLANTE.

[4] M. DE COCK BUNING e altri, *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*, Ufficio delle pubblicazioni dell'Unione europea, Lussemburgo, 2018.

[5] *Ibid.*, p. 10. Per esempio, la disinformazione consta non solo di notizie completamente false, ma anche di forme di mescolanza tra elementi di informazione più o meno rispondenti al vero; inoltre, determinate fattispecie di disinformazione possono assumere forme diverse da quelle tradizionalmente associate al concetto di "notizia", comprendendo infatti anche casi quali la creazione di *account* o di reti di follower falsi, nonché la fabbricazione o manipolazione di video, il ricorso a pubblicità targettizzate, attività di *trolling* (soprattutto se organizzate e coordinate) e la diffusione di *meme*..

quelle informazioni false, inaccurate o fuorvianti, artificiosamente create, le quali siano presentate e diffuse con lo scopo precipuo di trarre un profitto di carattere economico, politico o ideologico e/o di provocare un danno a livello pubblico, ivi inclusa l'ingerenza nei processi elettorali e democratici. Occorre peraltro precisare che la nozione di disinformazione, così intesa, include non tanto quelle condotte che l'ordinamento riconosce come inerentemente illecite (si pensi, per esempio, al caso della diffamazione o a quello della calunnia), ma, piuttosto, tutti quei casi ove la produzione e diffusione di contenuto falso non integra, di per sé, una fattispecie illecita ma che produce, comunque, un potenziale danno ai principi e valori democratici. Sono peraltro proprio questi ultimi casi a rappresentare la maggiore sfida per il legislatore nazionale ed europeo, in quanto impongono la necessità di contemperare gli interessi pubblici al corretto funzionamento dello stato democratico con la necessità di non arrecare uno sproporzionato nocumento alla libertà individuale dei consociati a porre in essere condotte che l'ordinamento ritiene, di per sé, lecite.

L'elemento della volontarietà della creazione e diffusione di un contenuto falso, così come la volontà di causare un danno o di ottenere un profitto, rappresentano peraltro l'elemento di discrimine tra il fenomeno della disinformazione e quello, connesso ma ben diverso, della "misinformazione". Quest'ultimo "disordine informazionale" (*information disorder*) si caratterizza, infatti, per un elemento di carattere soggettivo, in quanto implica precisamente l'assenza della volontà di diffondere informazioni false e, pertanto, si sostanzia nella diffusione di materiale considerato genuino. È importante notare come in internet, nella maggioranza dei casi, fattispecie di disinformazione tendano a essere successivamente ricondivise e diffuse da soggetti differenti da quelli che le hanno originate: lo stesso contenuto, pertanto, rappresenta un caso di disinformazione con riferimento ai soggetti che lo hanno originato o disseminato nella consapevolezza della sua falsità e con la finalità di causare un danno o trarne beneficio, mentre costituisce un esempio di misinformazione con riferimento a quanti, appresa la notizia, abbiano contribuito a diffonderla nell'erronea convinzione della sua corrispondenza alla verità. Si tratta, come detto, di un aspetto tipico dell'ecosistema internet, ove i contenuti disinformativi hanno la capacità di propagarsi e assumere facilmente connotati virali.

Fenomeno ancora differente, ma di minor rilievo ai fini della presente trattazione, è quello della malinformazione, che consta della diffusione di informazioni rispondenti al vero (o comunque basate su elementi fattuali reali) le quali vengono tuttavia comunicate e diffuse al preciso fine di provocare un danno.

Sebbene in misura diversa, i tre menzionati fenomeni – disinformazione, misinformazione e malinformazione – pongono potenzialmente importanti pericoli per il buon funzionamento della società democratica, in quanto impattanti sui processi decisionali della stessa popolazione. In particolare, la

disinformazione mira direttamente a determinare una manipolazione o cattiva rappresentazione della realtà, al fine di alterare l'opinione pubblica ovvero di rafforzare determinate posizioni favorevoli all'agente "disinformatore": proprio per questo motivo, essa, intesa come sintesi tra l'elemento oggettivo della falsità e l'elemento soggettivo del "dolo" riferibile alla consapevolezza della falsità e volontà della diffusione di tale falsità per finalità di ordine egoistico, ha raccolto su di sé l'attenzione del legislatore euro-unitario e di quello nazionale.

Peraltro, giova sottolineare come lo stesso concetto di disinformazione possa includere al suo interno una pluralità di fattispecie differenti, più o meno gravi, le quali impongono chiaramente la necessità di adottare strategie e soluzioni differenziate e calibrate, soprattutto alla luce della necessità di garantire nel contempo la piena tutela della libertà di espressione e di informazione. Inoltre, come segnalato dallo stesso HLEG, la disinformazione, costituendo essa stessa un fenomeno multiforme, richiede di essere affrontata attraverso il ricorso a un approccio "multi-dimensionale", capace di includere, oltre a misure di carattere giuridico-sanzionatorio, anche e soprattutto interventi volti alla promozione della trasparenza, allo sviluppo dell'alfabetizzazione mediatica e informazionale dei cittadini, all'emancipazione e potenziamento di utenti e giornalisti, nonché alla promozione di un sistema mediatico pluralistico, connotato da diversità e sostenibilità.

3. Disinformazione e ingerenze di paesi terzi

È importante porre in evidenza come il tema del rapporto tra disinformazione e processi democratici non abbia soltanto una dimensione "interna", ma, al contrario, come essa abbia sempre di più una significativa importanza sul piano geopolitico e internazionale. In anni recenti, tale rilevanza sul piano internazionale e, in particolare, il ricorso all'utilizzo dello strumento della disinformazione quale mezzo di supporto di strategie di offesa, o comunque di politiche in ambito estero, hanno assunto un rilievo particolarmente significativo soprattutto alla luce dell'emergere delle tensioni, da un lato, tra Russia e Ucraina e, dall'altro lato, in Medio Oriente.

In particolare, il ricorso da parte della Russia allo strumento della disinformazione nel contesto della sua aggressione all'integrità del territorio ucraino, simboleggiato tra l'altro da episodi quale la diffusione di *deepfake* ritraenti Volodymyr Zelensky intento a chiamare i suoi concittadini a deporre le armi[6], ha posto al centro dell'attenzione internazionale il tema dell'uso offensivo della disinformazione, tanto che vi è chi ha persino definito il conflitto come la possibile "prima guerra dei *social media*" (*First Social Media War*)[7]. Come sottolineato in un rapporto del 2023 a opera dell'ONG Freedom House, lo stato e il settore privato

[6] Si veda, tra gli altri, M. HOLROYD, Deepfake Zelenskyy surrender video is the 'first intentionally used' *in Ukraine war*, in *Euronews*, 16 marzo 2022.
[7] P. SUCIU, *Is Russia's Invasion Of Ukraine The Firs Social Media War?*, in *Forbes*, 1 marzo 2022.

russi hanno collaborato strettamente al fine di diffondere a livello internazionale la propaganda del Cremlino relativa all'invasione dell'Ucraina: in un'operazione denominata "*Doppelgänger*", per esempio, si è provveduto all'imitazione di testate mediatiche statunitensi, italiane, britanniche e francesi al fine precipuo di diffondere informazioni false e narrative cospirazionistiche con riferimento alle sanzioni europee nei confronti della Russia ovvero con riferimento ai rifugiati ucraini; ancora, un *network* russo, Cyber Front Z, avrebbe utilizzato Telegram al fine di delegare a una quantità di commentatori il compito di diffondere centinaia di post al giorno sui vari *social media* a sostegno delle azioni di Vladimir Putin[8].

Non stupisce dunque che, in tal senso, la strategia dell'Unione di contrasto alla disinformazione si sia focalizzata anche (e soprattutto) sul tentativo di arginare l'impatto della disinformazione proveniente dall'estero all'interno del dibattito democratico europeo. Anzi, è precisamente in questo settore che si sono avute le prime iniziative dell'Unione in materia di contrasto alle "*fake news*". Infatti, la strategia dell'Unione europea di contrasto alla disinformazione affonda le proprie radici già nel marzo 2015 quando, a seguito dell'occupazione della penisola di Crimea da parte della Federazione Russa, il Consiglio europeo, sottolineando «l'esigenza di contrastare le campagne di disinformazione in corso da parte della Russia», invitò l'Alto rappresentante dell'Unione per gli affari esteri e la politica di sicurezza a elaborare un piano d'azione in materia di comunicazione strategica[9]. A tal fine fu costituita, in seno al Servizio europeo per l'azione esterna (SEAE) la *task force* "East StratCom", avente tra l'altro il fine di rafforzare il contesto mediatico generale nel vicinato orientale e negli Stati Membri e di migliorare le strategie euro-unitarie di previsione, contrasto e risposta alle attività di disinformazione estere.

A seguito dell'attacco russo del febbraio 2022, l'azione dell'Unione si è intensificata in modo significativo, allo scopo precipuo di ridurre quanto più possibile l'impatto della propaganda filo-putiniana nel contesto europeo. La strategia si è in questo senso mossa in molteplici direzioni, ivi inclusa, per esempio, l'adozione di sanzioni a carico di attori privati e media russi. In *RT France*, il Tribunale dell'Unione europea riconosceva la legittimità delle sanzioni poste a carico di un canale televisivo francese dedito alla promozione della propaganda russa, offrendo una sostanziale copertura "costituzionale", alla luce dei principi espressi nella Carta dei Diritti Fondamentali dell'Unione, ed in particolare quelle delle giustificate, perché proporzionali, restrizioni alla libertà di espressione, delle strategie di contrasto alla disinformazione di origine straniere adottate dalle istituzioni europee[10].

[8] A. FUNK, A. SHAHBAZ, K. VESTEINSSON, *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence*, 2023, disponibile sul sito di Freedom House, p. 9.

[9] Consiglio europeo, Conclusioni adottate nella riunione del 19 e 20 marzo 2015, EUCO 11/15, p. 5.

[10] Tribunale UE, causa T-125/22, *RT France c. Consiglio dell'Unione europea*, 27 luglio 2022. Si veda, per un commento, P. DUNN, *Il contrasto europeo alla disinformazione nel contesto della guerra in Ucraina: riflessioni a margine del caso* RT France, in *Rivista di diritto dei media*, n. 1, 2023, pp. 291-301.

Allo stesso tempo, il nuovo pacchetto sulla difesa della democrazia europea presentato a dicembre 2023[11] include una proposta di direttiva che stabilisca requisiti armonizzati nel mercato interno sulla trasparenza della rappresentanza degli interessi esercitata per conto di paesi terzi[12]. Nella relazione di accompagnamento, la Commissione giustifica come la necessità di adottare misure armonizzate finalizzate a una maggiore trasparenza rispetto all'ingresso nel dibattito pubblico di narrative provenienti da paesi terzi sottolineando come «i governi dei paesi terzi fanno un uso sempre maggiore delle attività di rappresentanza d'interessi, insieme ai canali e ai processi diplomatici formali, per promuovere i propri obiettivi strategici» e come, d'altro canto, «se svolta in maniera occulta, l'attività di rappresentanza d'interessi per conto di paesi terzi rischia di essere utilizzata come canale di ingerenza nelle democrazie dell'Unione… con ripercussioni negative sulla vita politica degli Stati membri e dell'Unione nel suo complesso»[13].

Come appare evidente dalla titolazione, la proposta di direttiva mira a garantire una maggiore trasparenza in tutti i casi in cui venga posta in essere una "attività di rappresentanza d'interessi" a favore di paesi terzi, ovverosia ogniqualvolta venga posta in essere una «attività svolta allo scopo di influenzare, nell'Unione, lo sviluppo, la formulazione o l'attuazione di politiche o normative o processi decisionali pubblici»[14]. A tal fine si prevede, tra l'altro, la predisposizione di registri nazionali concernenti quei soggetti che svolgano tali attività di rappresentanza d'interessi[15], oltre che forme di cooperazione tra le autorità di controllo nazionali[16].

La direttiva proposta, invero, non si occupa specificatamente di innovare il quadro legislativo con riferimento all'ecosistema informazionale di internet – né focalizza la propria attenzione sull'impatto dell'IA su tale ecosistema e sui processi democratici. Emerge tuttavia dalla stessa relazione di accompagnamento la complementarità del proposto testo legislativo, avente un ampio respiro nel contesto della trasparenza riguardo le ingerenze di stati terzi, con la tutela della democrazia europea e con il contrasto al fenomeno della disinformazione[17].

Lo stretto legame tra disinformazione, propaganda estera e influenza dei processi elettorali, d'altra parte, rappresenta una problematica di particolare rilevanza, la quale è stata di recente sollevata anche innanzi la Corte europea dei diritti dell'uomo. Nel 2022, infatti, la Corte di Strasburgo è stata adita da alcuni

[11] Vedi Comunicazione COM/2023/630 della Commissione del 12 dicembre 2023 al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni sulla difesa della democrazia.
[12] Proposta di direttiva del Parlamento europeo e del Consiglio del 12 dicembre 2023 che stabilisce requisiti armonizzati nel mercato interno sulla trasparenza della rappresentanza d'interessi esercitata per conto di paesi terzi e che modifica la direttiva (UE) 2019/1937. COM/2023/637.
[13] *Ibid.*, pp. 1-2.
[14] *Ibid.*, art. 2(1).
[15] *Ibid.*, artt. 3, 9-11.
[16] *Ibid.*, artt. 17-19.
[17] *Ibid.*, pp. 7-8.

membri del Parlamento britannico a seguito della pubblicazione nel 2019, da parte della Commissione per il digitale, la cultura, i media e lo sport della Camera dei comuni, di un rapporto su "disinformazione e '*fake news*"", oltre che di un rapporto sulla Russia pubblicato dalla Commissione parlamentare per l'*intelligence* e la sicurezza. Tali rapporti riferivano credibili elementi di prova a sostegno dell'ipotesi di un'interferenza russa nel contesto del referendum del 2014 sull'indipendenza scozzese, del referendum sulla Brexit del 2016 e delle elezioni politiche del 2019. Sulla base di tali elementi, i ricorrenti sostengono che il Regno Unito avrebbe fallito i propri obblighi positivi di porre in essere adeguati sistemi di investigazione e di contrasto alle ingerenze della Russia nel contesto di tali votazioni, così violando l'articolo 3 del Primo protocollo addizionale alla Convenzione: ovverosia, la norma che tutela il diritto a libere elezioni[18].

Nel momento in cui si scrive, il caso[19] non è ancora stato deciso dalla Corte: esso offre, tuttavia, un'interessante prospettiva sulla questione in quanto, in particolare, apre la strada al riconoscimento di un diritto individuale della popolazione stessa a non essere soggetta a indebite forme di influenza esterna. In altre parole, la manipolazione dei processi decisionali degli elettori non costituirebbe una violazione soltanto a livello, per così dire, "aggregato" dei principi fondanti lo stato democratico: essa, invece, si caratterizzerebbe altresì per una dimensione strettamente personale, ricollegandosi direttamente al diritto all'autodeterminazione individuale e al diritto a un'informazione corretta e pluralistica. La decisione dei giudici di Strasburgo potrebbe, in tal senso, contribuire quindi a una significativa e ulteriore evoluzione del dibattito in oggetto, riconoscendo non solo una facoltà ma, potenzialmente, anche un dovere dello stato di fare fronte ai problemi connessi al fenomeno della disinformazione.

4. L'amplificazione della disinformazione dovuta all'utilizzo di modelli di intelligenza artificiale. Rischi ma anche opportunità

In anni recenti, il fenomeno della disinformazione si è sviluppato in modo significativo non solo a causa dell'espansione di internet e delle strategie di comunicazione digitale ma anche grazie allo sviluppo e alla diffusione dei sistemi di IA. Tali sistemi, tra l'altro, ricoprono una pluralità di ruoli differenti nel ciclo di vita della disinformazione, investendone sia la produzione e creazione, sia la disseminazione.

Sotto il primo profilo, quello relativo alla produzione della disinformazione si deve fare riferimento all'emersione di innumerevoli sistemi, legati in particolare alla cosiddetta "IA generativa", ai modelli fondativi e ai *large language models* (LLM), capaci di creare immagini, video e testi sintetici altamente

[18] «Le Alte Parti Contraenti si impegnano ad organizzare, ad intervalli ragionevoli, libere elezioni a scrutinio segreto, in condizioni tali da assicurare la libera espressione dell'opinione del popolo sulla scelta del corpo legislativo».

[19] C. edu, *Bradshaw e altri c. Regno Unito*, ric. 15653/22, comunicato il 20 dicembre 2022. Si veda in tal senso J. GRIERSON, *MPs take Russian election interference case to human rights court*, in *The Guardian*, 29 marzo 2022.

realistici. In tal senso, il numero di esempi concernenti l'utilizzo di sistemi di IA per la produzione di contenuti falsi (e in particolare *deepfake*) è particolarmente allarmante nel contesto odierno, soprattutto alla luce dello scoppio di conflitti – quali quello russo-ucraino e quello israeliano-palestinese – il cui impatto sullo scenario geopolitico mondiale appare essere determinante.

A novembre 2023, negli Stati Uniti, un video veniva fatto circolare in internet raffigurante la rappresentante democratica Alexandra Ocasio-Cortez intenta a discutere in modo (a dir poco) sconclusionato il tema della richiesta di un cessate-il-fuoco a Gaza: «*Cease-fire means that somebody sees a fire. It could be any kind of fire. It could be a big fire or a small fire, a bonfire or even a candle flame. It just matters that somebody sees a fire – that's why we call it a cease-fire*». Il video, riferibile, attraverso un *watermark* visibile sullo sfondo, a uno specifico account di X (C3PMeme), era chiaramente un *deepfake* ottenuto attraverso l'alterazione di un estratto di una *live session* su Instagram, ove Ocasio-Cortez aveva precedentemente espresso le sue posizioni favorevoli al cessate-il-fuoco[20]. Peraltro, nonostante la chiara falsità del prodotto audiovisivo, esso riceveva una significativa eco non solo su X, ma anche su una pluralità di *social network* differenti e raggiungendo, pertanto, un'ampia fetta della popolazione. Se, da un lato, la chiara insensatezza del discorso può apparire sotto certi profili ai limiti del comico, dall'altro lato è chiaro come la circolazione di un simile falso abbia il potenziale di minare in modo significativo l'autorevolezza di un personaggio politico come Ocasio-Cortez e pertanto, in modo più o meno diretto, di alterare la percezione politica della popolazione stessa.

Nello stesso periodo, un video contenente un messaggio audio veniva diffuso in cui, apparentemente, il sindaco di Londra Sadiq Khan avrebbe dichiarato la volontà di posticipare l'*Armistice Day*, giornata dedicata nel Regno Unito al ricordo dei tragici eventi della prima guerra mondiale, a favore di una marcia in favore della Palestina. L'audio, originariamente diffusosi su TikTok, si era successivamente espanso anche su altre piattaforme, venendo amplificato soprattutto da account per lo più vicini a idee di destra e indignati dalle supposte dichiarazioni di Khan[21].

L'IA si presta, in sostanza, a una pluralità di possibili utilizzi finalizzati alla produzione di materiali e contenuti aventi natura inerentemente disinformativa.

Come posto in luce da un recente studio pubblicato dallo European Digital Media Observatory (EDMO)[22], le tecniche più diffuse allo stato dell'arte, con riferimento alla produzione di immagini e video *deepfake* includono, tra le altre, le seguenti: manipolazione di attributi facciali (*face attribute manipulation*), attraverso cui vengono modificate certe caratteristiche riferibili al volto della persona ritratta (per

[20] REUTERS FACT CHECK, *Fact Check: Video of Ocasio-Cortez explaining ceasefires is digitally altered*, in *Reuters*, 24 novembre 2023.
[21] M. SPRING, *Sadiq Khan says fake AI audio of him nearly led to serious disorder*, in *BBC*, 13 febbraio 2024.
[22] K. BONTCHEVA (a cura di), *Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities*, 13 febbraio 2024, disponibile sul sito dello European Digital Media Observatory, pp. 4-5.

esempio, invecchiamento o ringiovanimento del viso); tecniche di scambio volti (*face swap*), attraverso le quali al volto presente nell'immagine o video originale se ne sostituisce un altro (tecnica, peraltro, frequentemente utilizzata per la realizzazione di *deepfake* a sfondo sessuale); tecniche di *face reenactment* e *lip synching* attraverso le quali un video viene manipolato in modo tale modificare i movimenti e le espressioni del volto della persona ritratta in modo tale da far sembrare che esse abbiano reso determinate dichiarazioni (è questo, per esempio, il caso sopra menzionato relativo al video di Ocasio-Cortez).

Al tempo stesso, come evidenziato anche dall'esempio concernente il messaggio audio falsamente attribuito a Sadiq Khan, l'industria dell'intelligenza artificiale è andata evolvendosi vertiginosamente anche nel campo della sintesi vocale. In tale contesto, in particolare, sono andati diffondendosi sia, da un lato, sistemi capaci di sintetizzare contenuti audio a partire da *input* testuali (*text-to-speech*) sia, dall'altro lato, sistemi capaci di modificare tali contenuti audio in modo tale da trasferire su di essi le particolarità vocali, incluso il timbro, di una determinata persona (*voice conversion*)

Ancora, l'IA, soprattutto a seguito dello sviluppo di LLM, trova oggi importanti sfere applicative nella produzione automatizzata di testi: peraltro, recenti studi hanno confermato come molti di tali sistemi, a fronte di *prompt* volti alla produzione di contenuto disinformativo, siano capaci e abbiano la tendenza a creare articoli e notizie apparentemente realistici ma non realmente corrispondenti alla verità dei fatti; solo alcuni sistemi, a fronte di tali suggerimenti, tendono a produrre testi che contraddicano questi ultimi o, comunque, solo in taluni casi sono esistenti filtri di sicurezza volti a ridurre la produzione di disinformazione

Chiaramente, la diffusione e l'uso massivo di tali sistemi implica potenziali rischi di particolare importanza nel contesto dell'espletamento di processi democratici.

Oltre che per la produzione di contenuti di disinformazione, l'IA ricopre un ruolo fondamentale nel contesto della (maggiore o minore) diffusione di tali contenuti. Con riferimento a tale aspetto, almeno due profili sembrano essere di particolare rilievo: da un lato, l'utilizzo di IA da parte dei produttori o di soggetti comunque interessati a promuovere la disseminazione di materiali falsi in rete al fine precipuo di aumentarne l'impatto; dall'altro lato, il peculiare sistema di organizzazione dei contenuti in internet.

Per quanto concerne il primo profilo, particolarmente diffusa è infatti la pratica di ricorrere a *social bot*, ovverosia ad account falsi gestiti in modo automatico o semiautomatico (in quest'ultimo caso si parla di "*cyborg*", ovverosia di profili gestiti in parte da persone vere e in parte dall'IA) con il preciso obiettivo di contribuire alla diffusione di materiali "inquinanti" (non solo disinformazione ma anche, per esempio, *hate speech*).

Per quanto concerne il secondo profilo, l'IA veste un ruolo di fondamentale importanza nel contesto della cura dei contenuti operata dalle piattaforme digitali. Di particolare rilievo sono, in questo contesto,

i sistemi di raccomandazione (*recommender* o *recommendation systems*) i quali, partendo dai dati e dalle informazioni raccolte sulle preferenze del singolo utente, sono in grado di predirne l'indice di gradimento con riferimento a nuovi contenuti ed elementi. Così, per esempio, le piattaforme di *video-sharing*, quali per esempio YouTube o Vimeo, processano le informazioni relative ai video visualizzati in precedenza dall'utente per suggerire allo stesso utente ulteriori contenuti ad essi simili; allo stesso modo, le piattaforme di programmi *on demand* (Netflix, Amazon Prime, Disney+ ecc.) utilizzano tali sistemi per rendere più facilmente fruibili i loro servizi ai clienti.

I sistemi di raccomandazione costituiscono peraltro strumenti essenziali anche nel contesto dei *social network*, permettendo ai fornitori di tali piattaforme di garantire agli utenti di essere raggiunti da quei contenuti maggiormente rispondenti non solo ai loro gusti e interessi personali, ma anche al loro sistema di valori e al loro orientamento ideologico-politico: sono, in altre parole, tasselli fondamentali per la costruzione di quello che è stato definito da Cass Sunstein come il "*Daily Me*"[23]. Appare pertanto evidente il ruolo centrale che tali sistemi di raccomandazione giocano in generale nella diffusione delle informazioni e, pertanto, nella formazione della stessa coscienza pubblica: essi hanno la capacità di influenzare e strutturare le stesse preferenze degli utenti e di guidarne le scelte sia a livello individuale sia a livello sociale e collettivo

L'utilizzo di sistemi automatizzati di cura dei contenuti in rete rappresenta un'arma a doppio taglio nel contesto del contrasto alla disinformazione e della promozione del pluralismo informativo. Sotto il profilo del loro possibile uso strumentale in favore del miglioramento del dibattito pubblico si è infatti suggerito di utilizzare proattivamente tali strumenti al fine precipuo di massimizzare il valore della diversità e del pluralismo informativo nella fase di sviluppo e di *design* dei sistemi di raccomandazione (cosiddetta "*diversity by design*"[24]), in modo tale da favorire un arricchimento dell'ecosistema informazionale e da ridurre gli effetti collaterali connessi alla formazione di camere dell'eco nell'ambiente digitale.

Al tempo stesso, sotto il profilo dei possibili effetti collaterali del ricorso ai sistemi di raccomandazione, si è sottolineato da più parti come lo sviluppo di questi ultimi sia in realtà per lo più orientato all'interesse – economico – della massimizzazione dell'*engagement* degli utenti e pertanto, dei profitti stessi. Poiché, peraltro, contenuti altamente divisivi e polarizzanti – come sono del resto sono sovente le notizie false – tendono frequentemente ad attrarre e catalizzare intorno a sé l'attenzione del pubblico, si pone il significativo rischio che l'algoritmo, pur di suscitare l'interesse degli utenti, sia disincentivato a ridurre la diffusione di disinformazione (se non, addirittura, incoraggiato a promuoverne la visibilità)

[23] C.R. SUNSTEIN, *#Republic.com. La democrazia nell'epoca dei social media*, Il Mulino, Bologna, 2017.
[24] N. HELBERGER, K. KARPPINEN, L. D'ACUNTO, *Exposure diversity as a design principle for recommender systems*, in *Information, Communication & Society*, n. 21(2), 2018, pp. 191-207.

Tale rischio è particolarmente acuto alla luce del fenomeno delle camere dell'eco e delle bolle-filtro, in quanto – soprattutto all'interno di gruppi sociali legati a teorie cospirazionistiche – l'interazione con contenuti di disinformazione apre alla concreta possibilità di ulteriore esposizione ad altra disinformazione. Inoltre, numerosi studi hanno posto in luce come l'IA applicata ai sistemi di moderazione e cura dei contenuti sia essa stessa esposta a errori e *bias* a danno, soprattutto, di minoranze ovvero di categorie vulnerabili o discriminate, con significativi impatti sull'effettiva promozione di un ecosistema informazionale pienamente pluralistico.

Lo strumento dell'intelligenza artificiale, peraltro, non deve essere percepito, e si tratta di un passaggio importante per non cadere nella trappola "distopica" o catastrofista sempre in agguato, solo quale mezzo di produzione e diffusione dei contenuti disinformativi, in quanto ne può invece rappresentare altresì un'importante arma per contrastarne gli effetti. Invero, come accade anche con riferimento alla moderazione di altri fenomeni di "inquinamento" dell'informazione, l'IA costituisce oggi un fattore essenziale e necessario nella riduzione della disinformazione in internet.

Sotto un primo profilo, l'IA viene correntemente utilizzata dalle piattaforme online anche al fine specifico di individuare non tanto i contenuti falsi in sé quanto, piuttosto, i profili falsi (ad esempio i cosiddetti "*bot*" o "*troll*" della rete). Tali *account* falsi, in effetti, presentano alcuni specifici e osservabili *pattern* comportamentali, quale per esempio un incremento abnorme delle proprie attività in rete negli stadi iniziali della diffusione di disinformazione

Sotto un secondo profilo, l'IA può essere utilizzata per individuare e identificare contenuti falsi e, in particolare, contenuti sintetici o comunque manipolati attraverso l'IA stessa. A tal fine, particolarmente diffusa è la pratica di allenare modelli di *deep learning*, fondati su reti neurali "convoluzionali" (*convolutional neural networks*, CNN) attraverso l'utilizzo di *dataset* largamente affermatisi in letteratura. Studi empirici hanno in effetti mostrato che immagini e contenuti audio, video o testuali tendono a presentare alcune caratteristiche che ne rivelano l'impronta sintetica.

Allo stato attuale, tali sistemi sono particolarmente efficaci quando siano stati allenati a identificare specifiche forme di manipolazione, anche se non mancano tentativi di sviluppare modelli aventi portata più generale: capaci, cioè, di astrarre i *pattern* invisibili caratterizzanti generalmente l'architettura dei contenuti disinformativi. Peraltro, l'utilizzo di tali tecniche di rilevazione non è esente da problemi significativi in termini di accuratezza. I limiti in termini di correttezza dei risultati ottenuti attraverso l'utilizzo dei sistemi di IA per l'individuazione e rimozione dei contenuti disinformativi in rete sono determinati da diversi fattori. Per esempio, tali sistemi non solo sono in molti casi incapaci di cogliere le sfumature di significato e i sottintesi caratterizzanti il linguaggio umano, ma sono altresì limitati nella loro capacità di cogliere gli aspetti di contesto e le allusioni culturali connessi a una specifica affermazione.

5. La reazione dei poteri pubblici in una prospettiva transatlantica

Appare a questo punto opportuno interrogarsi su come, sulle due sponde dell'Oceano Atlantico, si sia cercato di provare ad arginare i rischi che si sono evidenziati fino ad ora, specialmente alla luce della stagione elettorale e senza mai dimenticare che qualsiasi tentativo di regolazione relativo alla lotta contro la disinformazione deve sempre muoversi all'interno del principio di proporzionalità per non oltrepassare quella sottile linea rossa che fa di una restrizione giustificata e ragionevole alla libertà di espressione un'odiosa censura[25].

5.1 La sponda statunitense

Per quanto riguarda la sponda americana, in tutte le discussioni pubbliche e le audizioni governative a cui si è potuto assistere durante la permanenza di 9 mesi, ormai agli sgoccioli, quale *Fulbright fellow* presso la New York University, si è sempre fatto riferimento alla necessità di imparare dagli errori del passato. In primo luogo, si è spesso sostenuto come una lezione che dovrebbe essere ormai chiara è quella di non commettere quella che anche negli USA è oggi considerata da molti un'ingenuità: ovverosia, l'aver rinunciato a qualsiasi tentativo regolatorio nel periodo di genesi dell'allora nuova tecnologia di internet, di fatto delegando al settore privato, in una stagione di pieno liberismo digitale, operazioni di *enforcement* e di bilanciamento in caso di conflitto tra diritti fondamentali.

La seconda lezione a cui si è fatto riferimento è quella legata all'inerzia normativa e, in senso più ampio, regolatoria che ha preceduto le elezioni presidenziali americane del 2016, in cui l'impatto, in termini di (tentativi di) manipolazione del dibattito pubblico online e, conseguentemente, del risultato delle urne, è emerso in modo inequivocabile a valle delle elezioni stesse. Ed ancora si afferma quasi unanimemente che sarebbe un errore fatale, otto anni dopo, quando il professionismo della disinformazione è tutt'altro che un fenomeno sconosciuto e può, del resto, essere amplificato all'ennesima potenza grazie ad un perverso utilizzo dei modelli di intelligenza artificiale non agire a monte, cercando di attenuarne l'impatto attraverso un tentativo, da parte dello Stato federale, di prevedere degli strumenti più incisivi di lotta alla disinformazione.

Il problema è che, nonostante tali convinzioni, negli Stati Uniti vi è una situazione abbastanza paradossale quanto al (non) contrasto alla disinformazione, che fa sì che il dibattito tenda ad accartocciarsi su sé stesso e che va, se pure brevemente, affrontata.

[25] Si vedano, in tal senso, O. POLLICINO, *General Report: Freedom of Speech and the Regulation of Fake News*, in O. POLLICINO (a cura di), *Freedom of Speech and the Regulation of Fake News*, Intersentia, Cambridge, 2023, pp. 2-38; G. PITRUZZELLA, O. POLLICINO, *Disinformation and Hate Speech: A European Constitutional Perspective*, Bocconi University Press, Milano, 2020, pp. 97-143.

Da una parte, come si diceva, si avverte un certo terrore – potremmo dire fondato, visto quanto è stato provato nelle elezioni politiche Trump *versus* Clinton in merito ai professionisti della disinformazione che hanno inquinato il dibattito e provato ad incidere sull'esito delle votazioni – delle possibilità di interferenze esterne in occasione delle elezioni del 2024. L'obiettivo è salvaguardare la sicurezza nazionale da attacchi esterni (Cina e Russia sono visti come gli indiziati principali, anche per quanto riguarda le differenti, ma altrettanto efficaci, tecniche di disinformazione).

Nelle discussioni accademiche e istituzionali a cui si è avuto la possibilità di partecipare è emerso che ci sarebbero gli spazi per, come è stato espressamente proposto, *a narrow law prohibiting the use of AI to deceptively undermine US elections through fake speech*». A detta di molti, tale normativa potrebbe trovare una copertura costituzionale. Cosa vuol dire? Semplicemente che è vero che il Primo emendamento, con la famosa metafora del libero mercato delle idee, può anche estendere la sua protezione a quelle espressioni ed opinioni che possono risultare non veritiere, ma certamente non c'è (né ci può essere) alcuna protezione costituzionale per chi intenzionalmente decide di frodare gli elettori (e indirettamente, ovviamente, gli USA).

Da molte parti si è fatto notare come la *Federal Election Commission* (FEC) negli Stati Uniti avrebbe, volendo, il potere di *enforcement* di un'eventuale normativa che, in termini precisi e dettagliati, per superare il temuto *strict scrutiny* test della Corte suprema, proibisca l'uso di strumenti di intelligenza artificiale in grado di dare una rappresentazione del tutto falsa della realtà e di condizionare dunque discorso pubblico e votazione finale.

Si aggiunga che il Congresso dovrebbe avere un interesse (ed anche un dovere costituzionalmente rilevante) a proteggere l'integrità del processo elettorale. E, d'altro canto, se parliamo di tutela della libertà di espressione, il Primo emendamento dovrebbe anche riconoscere a chi è legittimato ad esercitare il diritto di voto di poterlo fare in modo informato e consapevole. Il che, ovviamente, include anche la capacità di non essere tratti volutamente in inganno da messaggi politici veicolati attraverso meccanismi di intelligenza artificiale. E, specialmente, di avere il diritto di sapere se l'autore del messaggio in questione è una persona umana o un algoritmo.

Se tutto ciò appare forse condivisibile in teoria, in pratica appare quasi impossibile pensare che una normativa del genere possa essere adottata prima delle elezioni presidenziali di novembre. Vi è infatti, come si accennava, un altro terrore che aleggia e costituisce quasi un tabù per ogni tentativo di regolazione. Quello, anche in questo caso culturalmente prima che costituzionalmente comprensibile, di poter minare in qualsiasi modo le fondamenta di "sua maestà "Primo emendamento

Un esempio può fare emergere plasticamente il paradosso cui si accennava. L'amministrazione Biden è stata accusata di aver indebitamente interferito con il *free speech* dei cittadini quando si è rivolta

direttamente agli operatori delle piattaforme online al fine di spingerli a porre in essere azioni di moderazione di contenuti disinformativi e misinformativi relativi, soprattutto alla pandemia di COVID-19: secondo gli attori, infatti, sotto la più o meno velata minaccia di interventi regolatori, l'amministrazione Biden avrebbe fatto pressioni su tali operatori in modo tale da sollecitare un silenziamento delle voci degli oppositori politici di destra. Nella sua sentenza resa l'8 settembre 2023, la Corte d'Appello per il Quinto Circuito degli Stati Uniti confermava l'incompatibilità di gran parte delle azioni governative in tal senso con il Primo emendamento, sottolineando in particolare che esse costituivano misure volte a costringere e incoraggiare le piattaforme a rimuovere i contenuti degli utenti[26]. Nel momento in cui si scrive, la causa è stata peraltro discussa, come *Murthy c. Missouri*, dalla stessa Corte Suprema federale, la quale, tuttavia, non ha tuttavia ancora reso la propria decisione: in ogni caso, durante l'udienza, la Corte, e in particolare la giudice Elena Kagan, ha espresso alcuni dubbi relativi all'efficacia coercitiva delle azioni intraprese dall'amministrazione Biden[27].

Quale che sia il risultato innanzi la Corte Suprema, e francamente appare assai improbabile – ma mai dire mai – che il "*free speech expansionism*" *trend* che sta caratterizzando questa stagione della giurisprudenza della stessa Corte si spinga fino al punto di identificare, in questa vicenda, una violazione della libertà di espressione il caso è emblematico dell'ancora assai attuale ostilità statunitense nei confronti di qualsiasi forma di intervento statale finalizzato alla limitazione del *free speech*, anche a costo di consentire la diffusione e disseminazione di contenuti e idee potenzialmente dannosi per la vita sociale (e democratica) del paese.

5.2 La sponda europea

Concentrandosi adesso sulla sponda europea, è noto come l'arsenale del costituzionalismo del vecchio continente abbia più frecce nel proprio arco rispetto alla controparte statunitense per provare a combattere efficacemente il fenomeno della disinformazione. Pur volendo andare al di là dell'abusato antagonismo dignità *versus* libertà quali rispettive stelle polare del costituzionalismo europeo e statunitense è un dato di fatto che la libertà di espressione non riveste nelle tradizioni costituzionali comuni in Europaquel valore sacrale proprio del Primo emendamento ed è soggetta a limitazioni esplicitamente codificate – il che sarebbe considerato quasi blasfemo negli Stati uniti.

Anche con riguardo al contesto europeo, come si è notato in precedenza per l'esperienza americana, si è provato recentemente a imparare dagli errori del passato. In particolare, in apertura, si è detto come

[26] Corte d'Appello per il Quinto Circuito degli Stati Uniti, *Missouri c. Biden*, n. 23-30445 (5 Cir., 8 settembre 2023). Si veda, tra gli altri, C. CALVERT, *Missouri v. Biden and the crossroads of politics, censorship and free speech*, in *The Hill*, 13 settembre 2023.

[27] G. MILLER, B. LENNETT, *Supreme Court Justices Question Standing, Evidence in Murthy v. Missouri*, in *Tech Policy Press*, 21 marzo 2024.

l'aver voluto fare propria, da parte della Commissione europea, nella sua prima strategia contro la disinformazione, la metafora del libero mercato delle idee di matrice statunitense abbia portato a opzioni di politica del diritto, con particolare riferimento ad un ricorso eccessivo alla *self-regulation*, che si sono rilevate di un'efficacia assai limitata. Basti pensare al primo *Code of practice on disinformation* del 2018 che, come si è detto, si è stato assai deludente.

Già due anni dopo, nel 2020, con specifico riferimento ai rischi della disinformazione nella stagione elettorale, la Comunicazione della Commissione europea sul piano d'azione per la democrazia europea riconosceva, specificatamente, che «la rapida crescita delle campagne elettorali online e delle piattaforme online ha portato allo scoperto nuove vulnerabilità e ha reso più difficile mantenere l'integrità delle elezioni, garantire la libertà e il pluralismo dei media e proteggere il processo democratico dalla disinformazione e altre forme di manipolazione», osservando nel contempo che l'impatto della disinformazione «è amplificato dall'uso di algoritmi opachi controllati da piattaforme di comunicazione ampiamente utilizzate»[28]. Inoltre, la stessa Comunicazione sottolineava come l'efficacia degli strumenti di propaganda online siano «tanto più efficaci in quanto combinano i dati personali e l'intelligenza artificiale con la profilazione psicologica e tecniche complesse di *microtargeting*»[29].

Il riferimento al ruolo di sistemi di profilazione e tecniche di *microtargeting* pone in luce la consapevolezza dell'Unione della centralità dei sistemi automatizzati di cura dei contenuti nel contesto dell'informazione in rete. Con riferimento a tale aspetto, appare opportuno sottolineare come la necessità di tenere in considerazione il ruolo preminente dei sistemi di raccomandazione, soprattutto ai fini del contrasto alla disinformazione, sia stato posto in luce dallo stesso *Digital Services Act* (DSA) il quale, con riferimento agli obblighi di valutazione e attenuazione dei rischi sistemici di cui agli articoli 34-35, prevede che che i fornitori di piattaforme online e motori di ricerca di dimensioni molto grandi debbano «concentrarsi sui sistemi o su altri elementi che possano contribuire ai rischi, compresi tutti i sistemi algoritmici che possano essere pertinenti, in particolare i loro sistemi di raccomandazione e i loro sistemi pubblicitari» e, precisamente in questa ottica, dovrebbero «prestare particolare attenzione al modo in cui i loro servizi sono utilizzati per diffondere o amplificare contenuti fuorvianti o ingannevoli, compresa la

[28] Comunicazione COM/2020/790 della Commissione del 3 dicembre 2020 al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni sul piano d'azione per la democrazia europea, p. 2. Similmente, la precedente Comunicazione COM/2018/236 della Commissione del 26 aprile 2018 al Parlamento europeo, al Consiglio, al comitato economico e sociale europeo e al Comitato delle Regioni, Contrastare la disinformazione online: un approccio europeo, p. 12 sottolineava che, nel contesto dei processi elettorali, costituenti la base della democrazia europea, «la disinformazione fa ormai parte di una più ampia gamma di strumenti usati per manipolare questi processi, che comprende l'*hacking* o il *defacing* di siti web, o anche l'accesso a informazioni private sul conto di personaggi politici e la loro diffusione».

[29] Comunicazione COM/2020/790, *op. cit.*, p. 3.

disinformazione»[30]. Per di più, il DSA specifica che, tra le altre misure potenzialmente adottabili, i fornitori di tali servizi «potrebbero dover attenuare gli effetti negativi delle raccomandazioni personalizzate e correggere i criteri utilizzati nelle loro raccomandazioni»[31].

Coerentemente, la stessa Commissione, all'interno delle recenti linee-guida relative all'attenuazione dei rischi sistemici concernenti il regolare svolgimento dei processi democratici, sottolinea l'impatto significativo che tali sistemi di raccomandazione hanno nella formazione del panorama informativo e, di conseguenza, nell'orientamento della stessa volontà collettiva[32]. La Commissione suggerisce pertanto, in modo particolare, di: assicurare che i sistemi di raccomandazione siano programmati e adattati in maniera tale da garantire agli utenti effettive possibilità di scelta e controllo sui *feed*, avendo riguardo della diversità e del pluralismo mediatico; stabilire misure per ridurre la diffusività della disinformazione nel contesto di elezioni sulla base di metodologie chiare e trasparenti; stabilire misure per limitare la diffusione, attraverso i sistemi di raccomandazione, di contenuti ingannevoli, falsi o fuorvianti, generati con l'ausilio di IA nel contesto di elezioni; valutare regolarmente la *performance* e l'impatto dei sistemi di raccomandazione e affrontare qualsiasi rischio emergente o problema relativo ai processi elettorali, anche attraverso l'aggiornamento e l'affinamento di *policy*, buone pratiche e algoritmi; stabilire misure per garantire la trasparenza del *design* e del funzionamento dei sistemi di raccomandazione, soprattutto in relazione ai dati e alle informazioni utilizzate; collaborare con enti esterni per condurre test e simulazioni per identificare i potenziali rischi dei sistemi di raccomandazione[33].

D'altro canto, lo stesso Codice rafforzato sulla disinformazione adottato nel 2022,che chi scrive ha avuto la possibilità di coordinare, riconosce il significativo impatto che i sistemi di raccomandazione hanno sulla dieta informazionale degli utenti e, pertanto, riconosce che tali sistemi debbano essere trasparenti e offrire loro la possibilità di scegliere modificare in ogni tempo le impostazioni relative a come i contenuti vengono loro presentati[34]. Per tale motivo, il Codice prevede una serie di impegni assunti dai firmatari i quali includono, tra l'altro, la minimizzazione dei rischi di propagazione virale della disinformazione attraverso l'adozione di pratiche sicure di sviluppo degli algoritmi e l'adeguamento delle proprie *policy*[35] e

[30] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), considerando 84.
[31] *Ibid.*, considerando 88.
[32] Così all'Allegato alla Comunicazione alla Commissione C(2024) 2121 del 26 marzo 2024 sull'approvazione del contenuto di un progetto di Comunicazione della Commissione sul-le linee-guida per i fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi sull'attenuazione dei rischi sistemici per i processi elettorali ai sensi della legge sui servizi digitali, para. 27(d): «*Recommender systems can play a significant role in shaping the information landscape and public opinion*».
[33] *Ibid.*
[34] Strengthened Code of practice against disinformation, 16 giugno 2022, parte V, preambolo, lett. (e).
[35] *Ibid.*, impegno n. 18.

la predisposizione di funzionalità e strumenti a favore degli utenti stessi finalizzati a consentire loro una maggiore autonomia decisionale.

Appare a questo importante ricordare come, inoltre, il nuovo codice del2022 si avvii, attraverso la mediazione del DSA, a rappresentare un modello non tanto di *self-regulation* quanto, piuttosto, di *co-regulation*. L'articolo 45 del Regolamento del 2022, infatti, prevede che la Commissione e il neoistituito comitato europeo per i servizi digitali, alla luce delle «sfide specifiche connesse alla lotta ai diversi tipi di contenuti illegali e ai rischi sistemici», possano incoraggiare e agevolare l'elaborazione di «codici di condotta volontari a livello di Unione per contribuire alla corretta applicazione» del DSA. Benché il rispetto di tali codici non sia di per sé obbligatorio per i *provider*, l'adeguamento o meno alle regole in essi contenuti rappresenta un fondamentale strumento a favore di (o contro) gli stessi fornitori per provare l'effettiva conformità delle proprie azioni alle norme contenute del DSA[36]. È chiaro come, pertanto, il codice del 2022, costruito precisamente con l'obiettivo di fungere da complemento al DSA, abbia tutte le carte in regola per trovare una più ampia e robusta applicazione rispetto al suo antecedente del 2018.

Oltre a tutto ciò, l'Unione è altresì intervenuta attivamente ad affrontare le strette connessioni tra disinformazione, interferenze esterne e processi democratici. Si inserisce, in questo senso, il già menzionato pacchetto sulla difesa della democrazia europea, dichiaratamente in continuità con lo stesso piano d'azione per la democrazia europea[37]. Tale pacchetto include, tra l'altro, una Raccomandazione volta specificatamente alla tutela dei processi elettorali nell'Unione[38], la quale, al fine di promuovere «norme democratiche rigorose in materia di elezioni nell'Unione» e di sostenere «il rafforzamento della natura europea e dell'efficienza nello svolgimento delle elezioni del Parlamento europeo»[39], prevede raccomandazioni rivolte agli Stati membri, ai partiti politici, alle fondazioni politiche e agli organizzatori di campagne elettorali europei e nazionali dirette, tra l'altro a, incoraggiare l'integrità delle elezioni e la correttezza delle campagne elettorali, adottare misure di trasparenza per le affiliazioni e la pubblicità politica, proteggere le infrastrutture connesse alle elezioni e garantire la resilienza nei confronti delle minacce informatiche e di altre minacce ibride e proteggere le informazioni relative alle elezioni.

Inoltre, la Raccomandazione riconosce espressamente il crescente ruolo dell'IA all'interno di tale ciclo e, a tal fine, sottolinea in più punti la necessità da parte dei suoi destinatari di adottare misure volte a contrastarne i possibili effetti collaterali. Così, per esempio, essa specifica la necessità che i partiti politici,

[36] Si veda in tal senso il considerando 104 del DSA.
[37] Comunicazione COM/2023/630 della Commissione del 12 dicembre 2023 al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni sulla difesa della democrazia, p. 9.
[38] Raccomandazione (UE) 2023/2829 della Commissione del 12 dicembre 2023 relativa a processi elettorali inclusivi e resilienti nell'Unione e al rafforzamento della natura europea e dell'efficienza nello svolgimento delle elezioni del Parlamento europeo.
[39] *Ibid.*, para. 1.

europei e nazionali, siano maggiormente trasparenti rispetto alle strategie di propaganda adottate, ivi incluso l'eventuale utilizzo di sistemi di IA[40]. Per di più, essa menziona chiaramente la relazione tra tali sistemi e la qualità dell'ecosistema informazionale stesso, sottolineando tra l'altro che «la manipolazione delle informazioni, l'ingerenza e la diffusione della disinformazione, anche con mezzi automatizzati sui social media, possono avere conseguenze negative sulla qualità del dibattito democratici, sull'esercizio dei diritti di voto, sulla percezione e sull'atteggiamento degli elettori, con effetti a lungo termine anche sulla partecipazione a future elezioni»[41].

Proprio nell'ottica di tali crescenti preoccupazioni concernenti il rapporto tra IA, disinformazione e processi elettorali si inserisce, tra l'altro, anche il recentissimo Regolamento 2024/900 sulla trasparenza e sul *targeting* della pubblicità politica[42], il quale riconosce espressamente il rischio che la pubblicità politica, tra l'altro sostenuta dall'avanzamento tecnologico e dalle contemporanee modalità comunicative, si trasformi in un vettore di disinformazione[43]. Di conseguenza, il Regolamento prevede, tra le altre cose, specifici requisiti di trasparenza relativi all'eventuale utilizzo di IA per la profilazione degli utenti e la conseguente targhettizzazione dei messaggi di pubblicità politica[44].

6. Conclusioni

Intelligenza artificiale, spazi digitali, disinformazione e principi, valori e processi democratici sono sempre più strettamente interconnessi nel contesto contemporaneo. Come si è avuto modo di mostrare, l'affermarsi di nuovi e sempre più sofisticati sistemi di IA ha determinato un salto qualitativo e quantitativo per quanto concerne il fenomeno delle comunemente note "*fake news*", conducendo a un incremento soprattutto delle risorse disponibili ai fini della creazione di contenuti falsi – e, tuttavia, particolarmente realistici e capaci di influenzare l'opinione pubblica – nonché dei mezzi per diffondere tali contenuti.

[40] Così *Ibid.*, para. 13: «I partiti politici europei e nazionali dovrebbero fornire sul proprio sito web informazioni in merito al proprio utilizzo della pubblicità politica, compresi gli importi destinati a tale pubblicità e le fonti di finanziamento utilizzate. Essi dovrebbero valutare la possibilità di garantire, su base volontaria, che la propria pubblicità politica possa essere chiaramente identificata come tale, anche quando comporta materiale elaborato internamente per la divulgazione online tramite social media. La pubblicità politica dovrebbe essere corredata di informazioni sull'identità del partito politico che la sponsorizza e, se del caso, di informazioni utili sui destinatari della pubblicità e *sull'impiego di sistemi di intelligenza artificiale* nell'elaborazione del contenuto o nella diffusione della pubblicità» (corsivo nostro).

[41] *Ibid.*, considerando 39.

[42] Regolamento (UE) 2024/900 del Parlamento europeo e del Consiglio, del 13 marzo 2024, relativo alla trasparenza e al targeting della pubblicità politica.

[43] *Ibid.*, considerando 4.

[44] *Ibid.*, art. 19, para. 1, lett. c). Con riferimento alla promozione di una maggiore trasparenza in materia di diffusione della pubblicità politica (anche e soprattutto in periodi di propaganda elettorale), si veda altresì il codice di buone pratiche sulla disinformazione, *cit.*, spec. parte III.

L'impatto della disinformazione nell'ambito della "società algoritmica"[45] è stato particolarmente evidente in anni recenti con riferimento all'influenza che tale disordine informativo ha esercitato nel contesto di alcuni importanti processi democratici a livello europeo e globale. Del resto, una simile influenza ha evidentemente allarmato una fetta significativa degli attori istituzionali, a livello nazionale e sovranazionale, i quali, come si è cercato di porre in evidenza nelle pagine che precedono, hanno conseguentemente adottato – o per meglio dire, soprattutto alla luce del contesto costituzionale considerato, tentato di adottare – strategie più o meno efficaci di contrasto al fenomeno. In particolare, assai emblematica sembra essere l'evoluzione della strategia europea di contrasto alla disinformazione che, nella sua fase attuale, ha subito una drastica accelerazione. D'altro canto, se dal lato europeo si è assistito a un accrescersi della risposta legislativa al fenomeno della disinformazione, il quadro giuridico statunitense in tal senso appare attualmente per lo più invariato: al di là dell'Oceano sembra infatti prevalere ancora il magistero del Primo emendamento.

Occorre d'altro canto rifuggire il rischio di cadere in narrative tecno-distopiche con riguardo al rapporto tra informazione online, intelligenza artificiale e democrazia. Infatti, se l'IA, come si è detto, nel sollevare nuove e importanti sfide per la regolazione degli spazi digitali, richiede un ripensamento delle strategie legislative concernenti la tutela della società da indebite influenze da parte del fenomeno della disinformazione, è importante altresì sottolineare il versante positivo dell'automazione e, in particolare, il crescente avanzamento tecnologico nell'ambito dello sviluppo di sistemi automatici per la riduzione di quello stesso fenomeno.

La sfida più significativa, per il legislatore – europeo ma non solo – sarà insomma quella di riuscire a incanalare i nuovi strumenti e le nuove risorse offerte dall'IA per promuovere i fondamentali valori dello stato democratico e pluralistico, nonché per garantire il corretto funzionamento dei processi elettorali e decisionali pubblici. In tal senso, peraltro, sembra evidente che il futuro della regolazione della disinformazione, soprattutto alla sua intersezione con l'IA, sarà grandemente influenzato, precisamente, dai risultati di quelle consultazioni elettorali del 2024 cui si faceva cenno in apertura. Permane quindi, allo stato attuale, una cifra di imprevedibilità rispetto alla prossima evoluzione delle strategie descritte. Tuttavia, la consapevolezza della complessità crescente dello scenario che si è provato ad abbozzare può essere un primo passo per non farsi trovare dl tutto impreparati.

[45] J.M. BALKIN, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, in *U.C. Davis Law Review*, n. 51, 2018, pp. 1149-1210.