

Disinformazione digitale

**Modelli algoritmici
e valori democratici**

a cura di
Licia Califano
Federica Fabrizi
Giovanni Sartor

FrancoAngeli 

Collana

di Diritto

SAGGI E RICERCHE



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più: [Pubblica con noi](#)

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "[Informatemi](#)" per ricevere via e-mail le segnalazioni delle novità.

Disinformazione digitale

**Modelli algoritmici
e valori democratici**

a cura di
**Licia Califano
Federica Fabrizzi
Giovanni Sartor**

FrancoAngeli 

Collana

di Diritto

SAGGI E RICERCHE



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

Questo volume rappresenta il risultato conclusivo del Progetto PRIN 2022 PNRR “DAFNE” (*Democratic governance of Automated system for Fake News*), finanziato dall'Unione europea – Next Generation EU nell'ambito del PNRR, Missione 4 Componente 2, Investimento 1.1, Bando “PRIN 2022 PNRR” del MUR emanato con Decreto Direttoriale n. 1409 del 14 settembre 2022, Codice MUR P2022R7RS9, CUP H53D23010930001, e raccoglie alcune delle relazioni e degli interventi presentati al Convegno “Poteri pubblici e poteri privati nel governo dell'informazione digitale: la scelta europea dell'accountability” tenutosi presso l'Università degli Studi di Urbino Carlo Bo il 24-25 ottobre 2025.

Isbn e-book Open Access: 9788835191209

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy.

Pubblicato con licenza *Creative Commons*

Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale

(CC-BY-NC-ND 4.0).

Sono riservati i diritti per Text and Data Mining (TDM), AI training e tutte le tecnologie simili.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore.

L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Gli eventuali link attivi e QR code inseriti nel volume sono forniti dall'autore.

L'editore non si assume alcuna responsabilità sui link attivi e QR code ivi contenuti che rimandano a siti non appartenenti a FrancoAngeli.

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835191209

INDICE

<i>Introduzione</i>	pag. 9

Sezione I Disinformazione e democrazia

<i>Verità e disinformazione nella sfera pubblica digitale</i> Massimo Durante	» 15

<i>La cura contro la disinformazione. Rimedi giuridici (e non) per contrastarne la diffusione</i> Federica Fabrizzi	» 35

<i>La qualificabilità della disinformazione quale problema di cybersicurezza</i> Giovanni Barozzi Reggiani	» 59

<i>Oltre la disinformazione: il problema della manipolazione nel quadro della regolazione algoritmica europea</i> Ludovica Durst	» 75

<i>Il disordine informativo online: classificare le condotte per definirle e prevenirle giuridicamente</i> Andrea Ruffo	» 101

<i>La crisi delle democrazie costituzionali al tempo della disinformazione: quale futuro per la liberal-democrazia?</i> Luca Maria Tonelli	» 113

Sezione II

Informazione digitale e tutela dell'individuo

- Poteri pubblici, poteri privati e governo costituzionale
dell'informazione digitale*
Giulio Enea Vigevani pag. 131
- AI e nuovi modelli di garanzia dei diritti fondamentali:
criticità e prospettive applicative*
Giulia Vasino » 141
- Violenza digitale di genere in ambito politico: profili comparati
tra Unione europea e Messico*
Rosa Iannaccone » 165
- L'informazione nell'era dell'algoritmo: intelligenza artificiale
e libertà di espressione nella governance europea*
Giulia Napoli » 179
- Disclosure, Explainability e Human Oversight: nuovi diritti
fondamentali nell'era dell'IA e dei Private Governors*
Andrei-Mihai Pop » 193

Sezione III

Comunicazione elettorale e politica

- Comunicazione istituzionale e intelligenza artificiale. Appunti*
Massimo Cavino » 207
- Il Regolamento europeo sulla pubblicità politica,
ovvero del difficile tentativo di regolamentare
la propaganda online (e non solo)*
Anna Papa » 219
- La comunicazione politica nella digital era*
Giuliaserena Stegher » 237

Fighting Foreign Interference. *Un'analisi comparata delle legislazioni contro le interferenze straniere nei processi elettorali in Australia, Canada, Irlanda, Regno Unito e Nuova Zelanda*
Emanuele Gabriele pag. 269

Il contrasto alla disinformazione nei procedimenti elettorali: spunti comparatistici tra Italia e Spagna
Lorenzo De Carlo » 283

Comunicazione e nuove tecnologie: la propaganda politica online nella Costituzione
Allegra Dominici » 297

Il sistema di governance ed enforcement del Regolamento (UE) 2024/900 alla prova dell'effettività: un'analisi comparativa con il Digital Services Act
Emanuela Palomba » 311

Sezione IV **Soluzioni regolatorie e tecniche** **nel mondo del digitale**

The Governance of "Augmented Disinformation" between the Digital Services Act and the AI Act
Federico Galli, Giuseppe Contissa » 325

Fake news e Intelligenza Artificiale: sfide etiche e soluzioni tecnologiche
Andrea Ongarini, Federico Cerutti, Andrea Loreggia » 357

Misure europee e nazionali contro la disinformazione strategica online. Contributo a partire dal caso Storm-1516
Domenico Bruno » 369

Large Language Models: applicazioni e criticità nel dibattito pubblico-informativo. Spunti di riflessione informatico-giuridici
Casimiro Coniglione » 385

*The state of the art legal and technical of automated fake news
moderation systems*

Camilla Scarpellino

pag. 399

Affrontare la disinformazione online:

il sistema di mitigazione dei rischi nel Digital Services Act

Sara Gallone

» 415

Riflessioni conclusive

*L'attualità del costituzionalismo a garanzia dei diritti
fondamentali nella società digitale*

Licia Califano

» 427

Autrici e Autori

» 439

INTRODUZIONE

Licia Califano, Federica Fabrizzi**, Giovanni Sartor****

Il presente Volume raccoglie gli atti e gli interventi del convegno finale del PRIN 2022 “DAFNE – *Democratic governance of Automated systems for Fake News*”, tenutosi presso l’Università degli Studi di Urbino Carlo Bo il 24 e 25 ottobre 2025.

Costituisce dunque il momento conclusivo di un intenso itinerario di ricerca, di confronto e di dibattito scientifico che ha coinvolto le Unità di Ricerca dell’Università di Urbino, di Roma Sapienza e di Bologna negli ultimi due anni, in un dialogo aperto con studiosi ed esperti anche esterni al progetto, in grado di contribuire ad arricchire la riflessione su una tematica complessa e trasversale, quanto urgente e attuale, quale è quella della lotta alla disinformazione.

Il tema al centro del progetto di ricerca puntava invero verso un obiettivo di per sé piuttosto ambizioso e sfidante per il giurista, trattandosi di indagare il fenomeno sfuggente della disinformazione e dei sistemi di moderazione delle fake news attraverso le linee del loro possibile sviluppo, da orientare secondo la bussola del rispetto dei diritti fondamentali e dello Stato di diritto.

La rapidità di trasformazione del contesto tecnologico di riferimento e la necessità di adattare di conseguenza la risposta normativa, pur nell’arco di tempo di poco più di due anni, ha contribuito a connotare significativamente il percorso di indagine intrapreso, alla luce delle continue sfide di aggiornamento delle coordinate regolatorie: per citare solo i più incisivi interventi normativi, basti pensare, per quanto attiene al quadro sovranazionale, all’approvazione, nel 2024, del Regolamento europeo sull’Intelligenza artificiale e alla piena applicabilità, dallo stesso anno, del Digital Services Act e del Digital Markets Act,

* Professoressa ordinaria di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

** Professoressa associata di Diritto costituzionale e pubblico, Sapienza Università di Roma.

*** Professore ordinario di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

mentre nel 2025 divenivano applicabili lo European Media Freedom Act e il Regolamento sulla trasparenza e il targeting della pubblicità politica ed entrava in vigore, in autunno, la legge italiana sull'Intelligenza artificiale, per quanto riguarda la dimensione domestica.

In tale quadro, non si può non mettere in rilievo innanzitutto un dato metodologico: potersi avvalere di una prospettiva multidisciplinare è in questo contesto un elemento essenziale al fine di poter efficacemente ragionare sull'impatto dell'evoluzione normativa in corso, e di restituire, grazie all'integrazione di *law and technology*, una visione completa delle dinamiche in gioco.

L'impostazione metodologica trasversale del progetto, contrassegnata dal dialogo tra approccio costituzionalistico, gius-privatistico e di informatica giuridica, si è rivelata indispensabile per cogliere, in tutte le sue implicazioni, il cruciale punto di caduta relativo agli effetti della disinformazione perpetrata attraverso gli strumenti algoritmici. Un impatto evidente sotto il profilo dell'incisione della sfera di libertà e dei diritti fondamentali, così come sotto il profilo della riconfigurazione del ruolo dei poteri pubblici e privati: elementi parimenti essenziali per il corretto funzionamento dell'ordinamento democratico nel quadro del rispetto dello Stato di diritto.

Le complesse dinamiche che legano il sistema informativo a quello democratico, su cui si dispiega l'effetto sempre più disarticolante delle *fake news*, sono state il filo rosso del primo convegno intermedio ospitato presso l'Università Sapienza di Roma il 9 dicembre 2024, intitolato "Information disorder e sistema democratico. Principi, regole e tecniche contro la disinformazione", i cui atti sono stati raccolti e pubblicati in un fascicolo speciale della rivista *federalismi.it*.

Era stata quella l'occasione per illustrare approfonditamente i principali nodi problematici nel quadro normativo nazionale e sovranazionale, con importanti interrogativi ancora aperti consegnati alle fasi conclusive del progetto.

Non tutti i profili in discussione hanno trovato ovviamente risposta: in un contesto connotato da una continua trasformazione o transizione, forse è strutturale che gli adattamenti non possano mai essere definitivi.

Cionondimeno è possibile far affiorare alcune risposte, che il gruppo di ricerca ha individuato intorno al principio dell'*accountability*. È questo principio ad aver fatto da guida, sullo sfondo, alle riflessioni proposte nel convegno finale di Urbino, dedicato a "Poteri pubblici e poteri privati nel governo dell'Informazione digitale: la scelta europea dell'*accountability*", che si è voluto impostare prevedendo anche il coinvolgimento di giovani studiosi tramite una *call* che ha avuto una risposta molto significativa.

Si tratta di una problematica invero particolarmente ampia e articolata, come dimostra la varietà dei contributi raccolti in questo Volume suddivisi se-

condo le principali aree tematiche travolte dal fenomeno della disinformazione, pur tutte fra loro interdipendenti: la tenuta generale del sistema democratico e dei principi e diritti che ne sono alla base, il dispiegarsi della normativa in merito alla disciplina del rapporto tra poteri (pubblici e privati) e alle garanzie dei diritti individuali, lo specifico ambito della comunicazione elettorale e politica cardine e infine il ruolo integrato della tecnologia nell'individuazione delle soluzioni secondo un approccio *technology-based*.

Il risultato – che crediamo presenti profili di interesse – è una riflessione corale, condotta secondo le sensibilità e le competenze di ciascuno, ma con il filo conduttore dell'attualità del costituzionalismo a garanzia dei diritti fondamentali nella società digitale.

Al termine di questo lungo percorso, un ringraziamento doveroso va indirizzato a tutti i componenti delle tre Unità di Urbino, Roma Sapienza e Bologna per i confronti, gli stimoli e i contributi offerti alla riflessione; tra loro, un grazie particolare va a Ludovica Durst, Federico Galli e Giulia Vasino, valentissimi giovani studiosi e lavoratori instancabili.

Sezione I

DISINFORMAZIONE E DEMOCRAZIA

VERITÀ E DISINFORMAZIONE NELLA SFERA PUBBLICA DIGITALE

*Massimo Durante**

SOMMARIO: 1. Premessa – 2. La revisione del modello triadico di Balkin – 3. Le nuove forme di disinformazione – 3.1. Parcellizzazione del vero e metamorfosi del falso – 3.2. Gli obiettivi strategici della disinformazione – 3.3. La personalizzazione dei contenuti – 4. Contro-incentivi ed *epistemic agency* – 5. Conclusioni.

1. Premessa

Nel contesto attuale della comunicazione digitale, la disinformazione rappresenta una delle sfide più complesse e pervasive sia per gli ordinamenti giuridici sia per la stessa stabilità delle democrazie liberali. Si tratta di un fenomeno articolato e multilivello, non più riconducibile alle tradizionali forme di produzione del falso, ma oggetto di una radicale trasformazione. Oggi la disinformazione si manifesta tramite modalità differenti, alla luce delle quali è possibile individuare, come si vedrà, tre distinte articolazioni del processo disinformativo. Pertanto, sia la risposta del diritto sia quella della società devono necessariamente confrontarsi con questa pluralità di livelli e con la complessità strutturale del fenomeno. In questa prospettiva, occorre fin dal principio osservare che la disinformazione è di regola il sottoprodotto di una struttura di incentivi, prevalentemente di carattere politico ed economico¹. La disinforma-

* Professore ordinario di Filosofia del diritto, Università degli Studi di Torino.

1. Non vi è lo spazio qui per esaminare quello che rappresenta un nuovo filone di ricerca: l'economia della disinformazione online. Si v., di recente in tema, C.-A. Diaz Ruiz, *Market-Oriented Disinformation Research: Digital Advertising, Disinformation and Fake News on Social Media*, Routledge, 2025. Vedi inoltre in tema, J.-A. List, L.-M. Ramírez, J. Seither, J. Unda, B. Vallejo, *Toward an Understanding of the Economics of Misinformation: Evidence from a Demand Side Field Experiment on Critical Thinking*, NBER Working Paper 32367, 2024, pp. 1-18; T. Harashima, *An Economic Theory of Disinformation*, in *Theoretical and Practical Research in*

zione emerge dunque come l'esito di un processo alimentato primariamente da tre ordini di fattori:

1. interessi economici;
2. interessi politici;
3. la saldatura, o convergenza strategica, tra interessi economici e politici.

Quest'ultimo fattore, spesso sottovalutato, può invece assumere un ruolo di primo piano, tale da rendere necessaria una riconsiderazione della struttura triadica del modello di governance dell'informazione elaborato da Jack Balkin²: il cosiddetto *triangolo della libertà*. Se la disinformazione è il sottoprodotto di una struttura di incentivi, allora la risposta con cui si intende contrastare la disinformazione deve parimenti fondarsi su una articolata struttura di contro-incentivi. Nei momenti di crisi – come quello attuale, che investe il ruolo dell'informazione e, più in generale, della verità nello spazio pubblico – si tende, soprattutto nel contesto europeo, ad attribuire alla regolazione giuridica il compito di fornire soluzioni per una vasta gamma di problemi. Sebbene il diritto svolga indubbiamente una funzione rilevante in questo ambito, una parte sostanziale della questione riguarda processi più ampi di natura economica, politica, epistemologica e sociale, che non possono essere ricondotte solamente alla dimensione giuridica. La responsabilità di confrontarsi con tali processi ricade in larga misura su di noi, in quanto soggetti epistemici che producono, ricevono e utilizzano informazioni capaci di plasmare la sfera pubblica e, in particolare, il discorso politico, evitando al contempo di scivolare verso pericolose forme di paternalismo o verso indebite compressioni della libertà di espressione. Dobbiamo tornare a chiederci, come fa Michael Lynch nel suo ultimo libro³, se la verità (ovvero l'informazione di qualità, rilevante e attendibi-

Economic Fields, vol. 14, n. 1(27), 2023, pp. 16-28; G. Lisi, *Online disinformation: an economic analysis*, in *Economics and Business Letters*, vol. 12, n. 4, 2023, pp. 266-276.

2. J. Balkin, *Free Speech is a Triangle*, in *Columbia Law Review*, vol. 118, n. 7, *Symposium: A First Amendment for All? Free Expression in an Age of Inequality*, 2018, pp. 2011-2056; Id., *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School of Regulation*, in *UC Davis Law Review*, vol. 51, n. 3, 2018, pp. 1149-2010. Ho esaminato le principali implicazioni di tale modello in M. Durante, *Potere computazionale. L'impatto delle Ict su diritto, società, sapere*, Meltemi, 2019, pp. 209-225. In realtà, il modello triadico potrebbe essere considerato, di per sé, di natura plurale, poiché ciascuno dei tre attori può essere ulteriormente articolato in figure diverse che ne rappresentano distinte sfaccettature. In tema vedi di recente, G. Giacomini, *Il trilemma della libertà. Stati, cittadini, compagnie digitali*, La Nave di Teseo, 2025.

3. M.-P. Lynch, *On Truth in Politics: Why Democracy Demands It*, Princeton University Press, 2025. L'interesse di Lynch per il tema della verità investe da tempo l'ambito filosofico-politico (con un approccio pluralista alla verità) e, più di recente, quello del digitale. Si v., dello

le⁴) rivesta ancora un ruolo centrale nella sfera pubblica e nella formazione del discorso politico e, dunque, delle nostre credenze e opinioni. Si tratta di una diversa formulazione di un tema messo in evidenza nei primi anni Duemila, in relazione a Internet e alla produzione di informazioni online, in particolare da Yochai Benkler: la necessità di disporre di meccanismi di filtraggio, sia *ex ante* sia *ex post*, volti ad accreditare le informazioni sulla base della loro rilevanza e attendibilità⁵. Oggi, il dibattito sulla disinformazione si è concentrato al tal punto sul requisito dell'attendibilità epistemica delle informazioni da avere trascurato, in larga misura, il tema della rilevanza. Occorre, invece, tornare a interrogarsi su come, perché e in relazione a quali fini le informazioni sono rilevanti. Per dirlo in breve, ritengo che il principale antidoto alla disinformazione sia la produzione di informazioni di qualità al contempo rilevanti e attendibili, che richiedono naturalmente elevati costi di produzione. Ciò porta a chiedersi se, nelle attuali società digitali dell'informazione, vi siano incentivi economici, politici, culturali e sociali sufficienti a sostenere la loro produzione, diffusione e fruizione sia sul lato dell'offerta sia sul lato della domanda⁶. Al contempo, bi-

stesso autore: *Truth in Context: An Essay on Pluralism and Objectivity*, Bradford, 1998; *Truth as One and Many*, Oxford, 2011; *In Praise of Reason: Why Rationality Matters for Democracy*, The MIT Press, 2012; *The Internet of Us: Knowing More and Understanding Less in the Age*, Liveright, 2016; and *Know-It-All Society: Truth and Arrogance in Political Culture*, Liveright, 2019; *La verità e i suoi nemici*, Raffaello Cortina Editore, 2007.

4. Sulla scorta della teoria dell'informazione di Luciano Floridi, consideriamo che l'informazione semantica è un insieme di dati ben costruiti dal punto di vista sintattico, dotata di significato e veridica. Si v. L. Floridi, *Semantic Information and the Correctness Theory of Truth*, in *Erkenntnis*, vol. 74, n. 2, 2011, pp. 147-175. Per un'accurata disamina della "veridicality thesis", si v. S. Sequoiah-Grayson, *Semantic Conceptions of Information*, in *Stanford Encyclopedia of Philosophy*, 5 October 2005 (rev. 14 January 2025).

5. Vedi. Y. Benkler, *La ricchezza della rete. La produzione sociale trasforma il mercato e aumenta le libertà*, Egea, 2007, p. 214: «La selezione dell'informazione in base alla rilevanza e all'affidabilità è parte integrante di tutti i sistemi di comunicazione. Una comunicazione deve essere rilevante sia per chi la invia sia per chi la riceve. Il processo di accreditamento introduce un filtro ulteriore che seleziona l'informazione pertinente. La selezione dell'informazione rilevante e affidabile viene effettuata in base ai valori della persona che si occupa di filtrarla, non in base a quelli della persona che la riceve». Avevo affrontato il tema in M. Durante, *Il futuro del web: etica, diritto, decentramento. Dalla sussidiarietà digitale all'economia dell'informazione in rete*, Giappichelli, 2007, in particolare par. 4: «Cultura critica e autoriflessiva», pp. 259-265.

6. Ciò vale parimenti per la domanda e offerta di disinformazione, J.-A. List, L.-M. Ramírez, J. Seither, J. Unda, B. Vallejo, *Toward an Understanding of the Economics of Misinformation: Evidence from a Demand Side Field Experiment on Critical Thinking*, cit., p. 15: «In addition, we trust that future work will tackle the interplay of the demand and supply sides of the misinformation market. Until key features are understood about the production and consumption sides, and relevant interactions, efficient policies to combat the creation and spread of misinformation will be difficult, if not impossible».

sogna ricordare che la verità e la qualità dell'informazione non possono essere imposte per legge. Per quanto le norme assolvono a un ruolo importante nella regolazione della sfera pubblica (sanzionando il falso in circoscritte ipotesi) e nel bilanciare interessi, valori e diritti contrapposti, in democrazia nessuno (e tantomeno lo Stato) può detenere alcuna forma di monopolio sul vero⁷. Il *valore democratico della verità*, di cui parla Lynch, o dell'informazione di qualità possono risiedere solo nel processo con cui sono perseguiti nel confronto aperto, inteso come spazio delle ragioni, tipico di una democrazia pluralista⁸ (pur con tutti i limiti che circondano la metafora del *mercato delle idee*)⁹, e mai in un obiettivo fissato per legge, senza rischiare di trasformarsi nell'ideologia del più forte o del potente di turno o in forme di autoritarismo.

7. Ho insistito su questo punto in M. Durante, *Potere computazionale*, cit., in particolare cap. 6: *Verità e fake news*, dove ho sottolineato anche l'importanza di un ritorno al tema della rilevanza delle informazioni nel quadro di una concezione dell'*informatività* come *distanza dal vero*, delineata in L. Floridi, *Understanding Epistemic Relevance*, in *Erkenntnis*, vol. 69, n. 1, 2008, pp. 69-92.

8. M.-P. Lynch, *On Truth in Politics*, cit., pp. 30-31: «Democratic truth, and its value, lies not in the end, but in the process of pursuing that end. Again, I take this to be a fundamentally pragmatist idea. But against some pragmatists, I will argue that we cannot make sense of it unless we fully acknowledge the plurality of political judgements, and hence a *pluralism about truth itself*». Qui, come altrove (cfr. *Potere computazionale*, cit., in particolare cap. 6: *Verità e fake news*, dove critico il tema della post-verità), non affronto la questione metafisica dell'esistenza della verità. Ci si può limitare a osservare che, da un punto di vista epistemologico, pressoché in ogni epoca e società umana il compito – e il potere – di dire o accertare il vero è stato, di volta in volta, delegato a soggetti specifici (profeti, sacerdoti, filosofi, giudici, storici, scienziati, cronisti, ecc.). Ciò che muta nel tempo non è tanto l'esistenza di tale delega, quanto la sua estensione e il novero dei soggetti a cui è attribuita. Tale prospettiva epistemologica mette in gioco la questione politica della delega di potere, senza sollecitare necessariamente la prospettiva metafisica. Per una ricostruzione di tale dinamica nella modernità si veda, in particolare, l'opera di M. Foucault, di cui mi limito qui a operare un solo riferimento: *Truth and Power*, in P. Rabinow (ed.), *The Foucault Reader*, Pantheon Books, 1984, pp. 51-75.

9. Si v. O. Pollicino, *Fake News, Internet and Metaphors (to Be Handled Carefully)*, in *Rivista di diritto dei media*, 2017, pp. 1-3, e più di recente G. Di Gregorio, O. Pollicino, *The European Constitutional way to address disinformation in the age of Artificial Intelligence*, in *German Law Journal*, vol. 26, n. 3, April 2025, pp. 449-470. Per un'analisi del dibattito in tema si veda più ampiamente, D.-E. Ho, F. Schauer, *Testing the Marketplace of Ideas*, in *New York Law Review*, n. 90, 2015, pp. 1160-1228; E. Volokh, *In Defense of the Marketplace of Ideas / Search for Truth as a Theory of Free Speech Protection*, in *Virginia Law Review*, vol. 97, n. 3, 2011, pp. 595-601; A.-I. Goldman, J.-C. Cox, *Speech, Truth, and the Free Market for Ideas, Legal Theory*, n. 2, 1996, pp. 1-32. Da un punto di vista più strettamente epistemologico vedi di recente, L. Herzog, *Citizen Knowledge: Markets, Experts, and the Infrastructure of Democracy*, Oxford University Press, 2023, in particolare cap. 5: *What's Wrong with the "Marketplace of Ideas"?*

2. La revisione del modello triadico di Balkin

Le attuali forme di disinformazione mettono in gioco oggi una molteplicità sempre più ampia e diversificata di attori, che richiede di aggiornare il modello triadico della governance dell'informazione proposta da Jack Balkin. Il modello triadico di Balkin, che tiene in considerazione il ruolo dei soggetti pubblici tradizionali (stati, enti sovranazionali, ecc.), degli intermediari privati (titolari di piattaforme, social media, ecc.) e degli utenti (cittadini, consumatori e produttori di contenuti), aveva, e conserva a mio avviso, un triplice merito:

1. sottolineare la divisione decisiva tra chi produce i contenuti e chi li rende fruibili;
2. segnalare il passaggio dei soggetti privati da intermediari ad attori della governance, che diventano pertanto destinatari di aspettative, pressioni e richieste dall'alto (da parte dei soggetti pubblici) e dal basso (da parte degli utenti);
3. interpretare il tema della governance a partire dalla considerazione non solo degli effetti primari derivanti dall'interazione tra attori ma anche degli effetti secondari, spesso indesiderati o imprevisi, derivanti dalla moltiplicazione e stratificazione delle relazioni in gioco¹⁰.

Oggi, questo modello può e deve essere rivisto alla luce di tre ulteriori fattori di complicazione:

1. siamo confrontati a un processo di progressiva saldatura o compenetrazione tra interessi pubblici e privati di stati e intermediari privati (per cui è possibile parlare di soggetti macro-agenti, ossia di complesse entità pubblico-private in ambito tecnologico¹¹) o quantomeno a un processo di agevolazione dell'innovazione tecnologica dei privati da parte dei soggetti pubblici¹²;

10. Ho esaminato l'emergere di effetti indesiderati o imprevisi come esito della *governance multilivello* del digitale in M. Durante, *Potere computazionale*, cit., cap. 4: *Memoria e oblio* e 6. *Verità e fake news*, e inoltre in *Governance del digitale. Aporie della nuova intermediazione*, in L. Taddio, G. Giacomini (a cura di), *La filosofia del digitale*, Mimesis, 2020, pp. 223-242.

11. Tale strategia è sottolineata e apertamente rivendicata, ad esempio, da Alexander C. Karp, fondatore e amministratore delegato di Palantir Technologies, e Nicholas W. Zamiska, direttore degli affari societari dell'azienda, in A.-C. Karp, N. Zamiska, *The Technological Republic: Hard Power, Soft Belief, and the Future of the West*, Crown Currency, 2025. Si v. al riguardo P. Gerbaudo, *Il nuovo Deep State tecnologico*, in *Jacobin*, 21 luglio 2025, accessibile online: jacobinitalia.it/il-nuovo-deep-state-tecnologico/.

12. Si pensi, ad esempio, all'America's AI action plan (www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf), in cui si sottolinea che, p. 3: «To maintain

2. la lotta contro la verità, la conoscenza e l'informazione (non soltanto la mera disinformazione) riguarda e sempre più spesso prende di mira ciò che sono state definite come *infrastrutture epistemiche* della democrazia¹³ e quanti operano al loro interno. Questo processo coinvolge una molteplicità di entità e attori secondo Lisa Herzog, che vanno dalla scuola ai media, fino alle cosiddette associazioni secondarie (dalle organizzazioni della società civile alle associazioni sindacali)¹⁴. A ciò si deve aggiungere oggi, più ampiamente, l'attacco che è da più parti rivolto contro l'università, la scienza, la storia, il diritto, ecc. Ciò attiene a un livello più profondo della disinformazione, poiché investe direttamente i processi di formazione del sapere e, a uno stadio ancora più radicale, le forme della temporalità come orizzonti di costruzione del senso. In questa prospettiva, l'università contribuisce alla produzione della conoscenza nel *presente*; la storia e il diritto presiedono

global leadership in AI, America's private sector must be unencumbered by bureaucratic red tape. [...] AI is far too important to smother in bureaucracy at this early stage, whether at the state or Federal level. The Federal government should not allow AI-related Federal funding to be directed toward states with burdensome AI regulations that waste these funds, but should also not interfere with states' rights to pass prudent laws that are not unduly restrictive to innovation». Analogamente, dopo anni di intensa attività regolatoria, ai limiti della superfetazione normativa, la commissione europea ha formulato l'intenzione di invertire la rotta, tramite la proposta di un *Omnibus Package*, che vuole adattare «the Union's regulatory framework to a more volatile world: a new drive to simplify, clarify and improve the EU acquis, as a key measure to support the EU's competitiveness. This vision reflects the broader plan laid out by Commission President von der Leyen in her political guidelines for the 2024-2029 term. As also highlighted in the Draghi and Letta reports, the accumulation of rules has sometimes had an adverse effect on competitiveness. Fast and visible improvements are needed for people and businesses, through a more cost-effective and innovation-friendly implementation of our rules» (digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal).

13. L. Herzog, *Citizen Knowledge*, cit., p. 221: «If democracy has epistemic preconditions for functioning well, then democrats have good reasons for endorsing the institutions that ensure that these preconditions are in place. All citizens need to have access to the institutions that enable them to get access to relevant forms of knowledge and to make informed decisions. By using the term "infrastructure", I emphasize the character of these institutions as conditions of the possibility of democracy», e p. 223: «What matters is the right institutional framework, in which all epistemic tasks are taken on by trustworthy institutions on which citizens can rely, and in which citizens are, on the one hand, empowered vis-à-vis fake news and pressures to conform to prejudices, and on the other hand, understand the limits of their own knowledge and understand where to place their epistemic trust».

14. L. Herzog, *Citizen Knowledge*, cit., p. 221: «In this section, I discuss, exemplarily, three sets of institutions that are crucial for making sure that democratic practices are sufficiently truth-conducive: schools, the media, and civil society organizations and unions. Building and maintaining them is an investment in the epistemic functionality of democracy, which provides citizens with social structures in which they can take well-informed decisions, drawing on reliable sources or using reliable heuristics».

all'accertamento del vero nel *passato*; mentre la scienza è impegnata nell'elaborazione di modelli previsionali che incidono sulla capacità di formulare ipotesi sul *futuro*;

3. gli intermediari sono sempre più esposti ad aspettative, pressioni e richieste che provengono, in realtà, da tre direzioni differenti: non soltanto dall'alto (dai soggetti pubblici) o dal basso (dagli utenti); ma anche dall'interno (dai propri dipendenti). Gli intermediari sono primariamente guidati dall'esigenza di massimizzare il profitto e proteggere il loro modello di business: per questo tendono a rispondere a queste forme di pressione nella misura in cui la loro risposta non incide sul profitto e non pregiudica il loro modello di business. Più in generale, rendono conto del loro operato, attuando concrete modalità di *accountability*¹⁵, quanto più sono esposti a forme di pressione pubblica che supera una data soglia di tolleranza, che spesso ha a che vedere con la brand reputation o la pressione interna dei dipendenti¹⁶. In realtà, come osserva acutamente Balkin, sono proprio gli intermediari (e in particolare i titolari di social media) che dovrebbero essere destinatari di contro-incentivi per «diventare organizzazioni affidabili e degne di fiducia, al fine di facilitare, organizzare e curare il dibattito pubblico»¹⁷.

15. Il Dsa va certamente in questa direzione, istituendo un accurato sistema di norme al fine di potenziare gli obblighi di trasparenza. In altro contesto, ho rilevato, però, come l'attitudine a rendere conto del proprio operato, riconoscendo gli altri come interlocutori reali, richiede non soltanto la riduzione di asimmetrie informative ma soprattutto la riduzione di asimmetrie di potere. In tal senso si v. M. Durante, *The Democratic Governance of Information Societies. A Critique to the Theory of Stakeholders*, in *Philosophy & Technology*, vol. 28, n. 1, 2015, pp. 11-32.

16. Si veda il caso "Azure", in cui Microsoft ha sospeso l'accesso ai propri servizi cloud all'intelligence militare israeliana. La decisione – comunicata direttamente dal vicepresidente Brad Smith – è stata presa dopo un'inchiesta del *Guardian* e le proteste dei dipendenti che hanno rivelato come il reparto dell'esercito israeliano specializzato in cyber-intelligence e sorveglianza elettronica, utilizzasse la piattaforma cloud Azure per raccogliere e analizzare milioni di comunicazioni di civili palestinesi (www.theguardian.com/world/2025/sep/25/microsoft-blocks-israels-use-of-its-technology-in-mass-surveillance-of-palestinians). L'inchiesta ha mostrato che circa 8.000 terabyte di dati intercettati erano conservati nei datacenter olandesi di Microsoft, in violazione dei termini di servizio che vietano l'uso della tecnologia per sorveglianza di massa. Sebbene l'azienda sostenesse di non essere stata inizialmente a conoscenza di tale utilizzo, le pressioni del movimento di dipendenti Microsoft *No Azure for apartheid* (noazureforapartheid.com/) e le manifestazioni presso le sedi aziendali hanno accelerato la decisione di disporre il ritiro di servizi tecnologici occidentali all'esercito israeliano.

17. Si v. J. Balkin, *To Reform Social Media, Reform Informational Capitalism*, in L. Bollinger, G.-R. Stone (eds.), *Social Media, Freedom of Speech and the Future of Our Democracy*, Oxford University Press, 2022, pp. 233-254, in cui Balkin sostiene che, per riformare efficacemente i titolari di social media, occorra intervenire sul loro modello economico di base, puntando meno sulla censura dei contenuti e maggiormente su concorrenza, protezione dei dati e strumenti antitrust, così da promuovere una sfera pubblica digitale più diversificata e incentivare

3. Le nuove forme di disinformazione

Sul piano normativo, l'Unione Europea ha cercato di delineare un quadro di governance della disinformazione attraverso strumenti come il Digital Markets Act, il Digital Services Act, l'AI Act, il Codice di condotta sulla disinformazione, ecc. Ci troviamo come di consueto di fronte a uno scenario di intensa produzione normativa¹⁸. Gli strumenti regolatori si confrontano con tensioni di fondo tra valori fondamentali che si intrecciano tra di loro: dalla tutela del pluralismo dell'informazione e dell'integrità del discorso pubblico alla salvaguardia della libertà di espressione, in quanto principi cardine delle democrazie pluraliste. La difficoltà di definire in che cosa consista il "falso" o la "disinformazione" in un ambiente caratterizzato dalla complessità di un fenomeno in parte trasformato dalla natura digitale e da dinamiche di potere algoritmico rende il confine tra legittima espressione e meccanismi di disinformazione un terreno giuridicamente complesso e scivoloso. Un limite che caratterizza, spesso, la governance del digitale consiste nel separare, o nel non valutare con sufficiente attenzione, l'intreccio profondo tra approccio normativo ed epistemologico al complesso di fenomeni che si vuole regolare¹⁹.

comportamenti più responsabili da parte delle grandi piattaforme; in particolare vedi p. 234: «Only a handful of social media companies dominate online discourse. In addition, the business models of social media companies give them *incentives* to act irresponsibly and amplify false and harmful content. The goals of social media regulation should therefore be twofold. The first goal should be to ensure a more diverse ecology of social media so that no single company's construction or governance of the digital public sphere dominates. The second goal should be to give social media companies – or at least the largest and most powerful ones – *incentives* to become trusted and trustworthy organizations for facilitating, organizing, and curating public discourse. Social media companies serve an important social function in the digital public sphere, but their business models prevent them from serving this function adequately and responsibly» (corsivi miei).

18. Per uno sguardo di sintesi si v. P. Dunn, M. Durante, *Online Disinformation: Regulatory Issues and Approaches in the European Legal Landscape*, in M. Durante, U. Pagallo (eds.), *The De Gruyter Handbook on Law and Digital Technologies*, De Gruyter, 2025, pp. 31-51.

19. L'evoluzione tecnologica solleva sfide che possono essere esaminate secondo un approccio *normativo* e uno *epistemologico*. L'approccio normativo, adottato dalle istituzioni europee, individua pro e contro del progresso tecnologico. L'approccio epistemologico esamina, invece, come le tecnologie digitali generano output che creano informazioni, conoscenze e rappresentazioni della realtà. Per affrontare l'impatto delle tecnologie digitali (algoritmi, sistemi e modelli d'intelligenza artificiale, ecc.), l'approccio epistemologico deve essere attentamente considerato: la corretta individuazione dei rischi e delle opportunità create dalle tecnologie digitali dipende infatti anche e forse soprattutto dalla comprensione di come tali tecnologie funzionino dal punto di vista epistemico. In tema si v. L. Paseri, M. Durante, *Examining epistemological challenges of large language models in law*, in *Cambridge Forum on AI: Law and Governance*, 1, e7, 2025, pp. 1-13; L. Paseri, M. Durante, *Normative and Ethical Dimensions of Generative AI*.

Anche nell'ambito della regolazione della disinformazione online, come è ad esempio il caso del Digital Services Act, l'approccio normativo trova fondamento su un approccio epistemologico piuttosto tradizionale, che si concentra sulla veridicità dei contenuti, con riferimento a singoli enunciati, spesso colti in isolamento e privi di sufficienti informazioni di contesto. Da una parte, è certamente plausibile mettere l'accento sul contenuto dei messaggi, al fine di verificarne fattualità e intenzionalità²⁰ (le due coordinate epistemiche che consentono di distinguere tra *disinformazione*, *misinformazione* e *malinformazione*) in base a un approccio proposizionale ai contenuti diffusi. D'altra parte, occorrerebbe interrogarsi se l'analisi della disinformazione e della verità non debba esaminare in modo più rigoroso e articolato le nuove forme e i meccanismi attraverso cui vengono veicolate rappresentazioni distorte della realtà, in particolare attraverso la comunicazione online, ambito nel quale si assiste a una profonda rielaborazione del rapporto tra vero e falso.

Esiste in effetti una consolidata letteratura relativa agli effetti di rete nella produzione e amplificazione della disinformazione²¹, che sono per natura complessi da cogliere per il diritto o non hanno diretta rilevanza giuridica, come ad esempio la polarizzazione delle opinioni; l'ineguale distribuzione delle informazioni; la sovraccedenza informativa, ecc. È comprensibile in tal caso che il diritto si concentri su dati aspetti della disinformazione ma non possa incidere su altri, poiché non costituiscono di per sé violazioni di norme giuridiche ma solamente di norme epistemiche.

Consideriamo, ad esempio, la sovraccedenza informativa, la quale:

1. consente il sistematico ricorso a processi di *cherry-picking*, con cui si consolidano i bias di conferma;

From Epistemological Considerations to Societal Implications, in M. Zou, C. Poncibò, M. Ebers, R. Calo (eds.), *The Cambridge Handbook of Generative AI and the Law*, Cambridge University Press, 2025, pp. 27-42.

20. Per l'analisi di tali coordinate, in una prospettiva che tiene conto degli effetti di rete della comunicazione online, vedi G. Ruffo, *Rintracciando odio e bugie in rete*, in M. Durante, U. Pagallo (a cura di), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*, Mimesis, 2022, pp. 163-182. In prospettiva filosofica, vedi S. Obelitz Soe, *A unified account of information, misinformation and disinformation*, in *Synthese*, vol. 198, n. 6, 2021, pp. 5929-5949. In prospettiva istituzionale, i tre elementi del disordine e dell'inquinamento informativo sono presentati ed esaminati nel report per il Consiglio di Europa curato da C. Wardle, H. Derakhshan, *Information disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe Report, DGI 2017/09.

21. Si tratta di una letteratura troppo ampia per poterne fare menzione in questo contesto in termini esaustivi. Per uno sguardo di sintesi, rinvio di nuovo a G. Ruffo, *Rintracciando odio e bugie in rete*, cit., pp. 163-182.

2. rende possibile offrire una rappresentazione distorta della realtà entro il perimetro del vero²²;
3. produce un'opacità o un rumore di fondo (una sorta di *information pollution*) che non cancella la verità ma la rende influente.

In questa prospettiva, lo scopo della disinformazione attuale non consiste tanto nel produrre il falso, quanto piuttosto nello svuotare di rilevanza il riferimento alla verità. Quest'ultima è infatti soppiantata da un'eccedenza di contenuti di segno opposto o diverso (tra cui messaggi di disprezzo, odio, consenso, amore e simili), la cui ampiezza e capillarità di diffusione sono tali da costituire un bacino costante da cui attingere per formulare enunciati formalmente veridici²³ ma contenutisticamente ambigui (nella disinformazione online c'è sempre un momento in cui la quantità diviene qualità, sebbene in senso deteriore: si tratta di un sapere alchemico rovesciato, in cui è l'oro a essere trasformato in metallo vile).

3.1. *Parcellizzazione del vero e metamorfosi del falso*

Vi è dunque una frammentazione e parcellizzazione del vero e del falso, che non agevola la comprensione complessiva del fenomeno della disinformazione. Tradizionalmente, il "falso" viene inteso come l'alterazione intenzionale e puntuale della verità di un fatto o di un contenuto, suscettibile di accertamento empirico. Nell'ecosistema digitale, tuttavia, il falso diviene sistemico e relazionale: non è più solo una menzogna, ma un prodotto di dinamiche algoritmiche che selezionano, amplificano e personalizzano l'informazione in funzione di interessi economici o politici. In tal senso, il problema del falso si sposta dal piano della menzogna intenzionale (che pur tuttavia non scompare) a quello dell'opacità sistemica, dove la responsabilità non è più riconducibile soltanto a un singolo autore, ma a una complessa rete di soggetti, algoritmi e piattaforme

22. Ho argomentato questo fenomeno in M. Durante, *Perché l'attuale discorso politico-pubblico fa leva sulla paura?*, in *Filosofia politica*, vol. 1, 2010, pp. 49-70.

23. Immaginiamo per semplicità un soggetto pubblico che posti in rete un messaggio che riceva migliaia di commenti di diverso segno, dall'odio al sostegno fino all'ammirazione, prodotti da utenti o anche da bot intesi ad alimentare la discussione in rete: potenzialmente, ogni enunciato con cui il soggetto pubblico commenta le reazioni al proprio post ("molti hanno espresso odio nei miei confronti", "molti hanno espresso sostegno nei miei confronti", ecc.) è, una volta conseguita una quantità significativa di commenti, suscettibile di rispondere al vero. Siamo in tal caso confrontati a un fenomeno emergente di *iperverità discorsiva*, ossia una condizione in cui la sovrabbondanza di commenti, interpretazioni e meta-discorsi su un discorso genera un sistema autoreferenziale tale per cui ogni enunciato che attinge a tale sistema risulta vero non per corrispondenza a qualche fatto, ma per coerenza interna alla selezione del discorso.

che concorrono alla distorsione dell'informazione e alla costruzione di realtà alternative. La metamorfosi del falso alligna su una concezione non strettamente proposizionale della verità, per cui occorre tenere in considerazione la distinzione operata da Lynch tra *significati politici* e *proposizioni* o *credenze politiche*²⁴ sullo sfondo di quella che viene definita come una regola di conformità nella comunicazione politica: il discorso politico-pubblico ha interesse a produrre emozioni che suscitano adesione al contenuto dei messaggi. Il discorso politico-pubblico, ad esempio, trae vantaggio dal fatto che la moderazione dei contenuti funzioni parzialmente o non funzioni correttamente: l'oscuramento (*shadow banning*)²⁵ o la rimozione di un contenuto può suscitare più emozioni della sua formulazione. La moderazione dei contenuti (che opera in modo più o meno trasparente) diviene essa stessa un tema politico nella misura in cui può suscitare e veicolare reazioni ed emozioni. Per tale motivo, la comunicazione politica si basa, per Lynch, su una regola di conformità²⁶, per cui occorre formulare messaggi politici (*claims*) capaci di evocare piacere o dispiacere, messaggi che in molti casi non hanno più un contenuto descrittivo della realtà (rispetto a cui misurare un grado di veridicità o correttezza) ma un contenuto evocativo, in relazione a eventi futuri, aspirazioni, sogni, frustrazioni o promesse, a cui aderire o da rifiutare. In tal senso, la disinformazione non è nemmeno il fine ma lo strumento stesso della comunicazione politica. Il fatto che la verità possa ancora contare in politica ha, dunque, più a che vedere con il nostro modo di trattare le persone come interlocutori reali, ossia come soggetti epistemicamente liberi ed eguali nello spazio delle ragioni, che non con la puntuale verifica del contenuto di verità di una proposizione (che resta, nondimeno, un aspetto centrale)²⁷. La tecnologia e l'ideologia hanno reso l'epistemologia più impor-

24. M.-P. Lynch, *On Truth in Politics*, cit., p. 11: «Call it the Rule of Conformity: only make political claims that are Liked (or at least not Disliked) by your allies and potentially hated by your enemies»; pp. 21-23: «Political meanings are, therefore, the *political* associations and responses a community has in relation to something, where communities are individuated by shared commitments and conceptions of the good, and the conjoint activities they engage in to realize them. [...] social media is a political-meaning machine, generating and sustaining political meaning at the speed of a tweet»; p. 28: «What I call political *propositions* are those we arrive at when explicitly considering what society should do [...] in the face of collective problems».

25. Per un'accurata analisi del tema, vedi di recente P. Leerssen, *An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation*, in *Computer Law & Security Review*, vol. 48, 2023, pp. 1-13.

26. M.-P. Lynch, *On Truth in Politics*, cit., p. 12: «For them, following the Rule of Conformity is what makes a political claim *correct* in the only sense that makes a difference. In short, truth isn't a real value [...]. Despite what they tell themselves, they don't really care if their claims *accord with reality*, as opposed to reflecting partisan preferences».

27. M.-P. Lynch, *On Truth in Politics*, cit., pp. 14: «I will use the term "democratic politics"»

tante e necessaria²⁸ nell'ambito dell'analisi della comunicazione politica, non meno importante e necessaria: l'alternativa è rimpiazzare il riferimento all'informazione, alla conoscenza e alla verità con l'adesione emotiva al contenuto di messaggi creati e veicolati online.

3.2. *Gli obiettivi strategici della disinformazione*

La disinformazione è dunque una strategia che si articola (e si comprende nella sua complessa dinamica odierna) secondo tre diverse modalità che non si limitano a investire i contenuti, ma hanno principalmente a che fare con le intenzioni che abbiamo al riguardo del nostro interlocutore (individuale o collettivo): si passa, cioè, da una disinformazione basata sui contenuti a una disinformazione orientata all'interlocutore. Queste modalità riguardano il tentativo di manipolare, scegliere o escludere il nostro interlocutore. Queste diverse finalità fanno sì che possiamo distinguere tre diversi ordini della disinformazione:

1. quella che mira ad alterare il contenuto del discorso;
2. quella che mira a impoverire il linguaggio;
3. quella che mira ad alterare le condizioni del discorso.

Non si tratta, pertanto, di contrastare oggi la disinformazione intesa soltanto come manipolazione del vero e metamorfosi del falso, ma anche e forse soprattutto la disinformazione intesa, ulteriormente, come progressivo impoverimento del linguaggio (frammentazione del discorso) e alterazione delle condizioni di possibilità del discorso. In tal senso, la verifica della veridicità di un singolo enunciato, o parte di esso, fino all'enfasi sul senso di singole parole o espressioni (*blocco navale, genocidio*, ecc.), in una sorta di frammentazione sempre più radicale del discorso, appartengono, da una parte, all'opera pur meritevole di contrasto alle informazioni false ma, dall'altra, rischiano di farci perdere di vista il quadro più generale del discorso. Un tale processo di *miniaturizzazione* della verità ci consegna un dibattito pubblico che verte sempre più spesso sull'interpretazione e sulla discussione di frammenti che raramente si ricompongono nell'interesse di un discorso articolato, approfondito, sensato

to mean inclusive, representative, and respectful deliberation between free and equal persons about what society ought to do in the face of collective problems», e p. 83: «To give a person basic epistemic respect [...] is to treat them as someone who is potential participant in the democratic space of reasons. [...] Democratic politics, as I've defined it, requires exercising epistemic agency – that is, being responsive to reasons when forming and exchange beliefs».

28. M.-P. Lynch, *On Truth in Politics*, cit., p. 7: «Technology and ideology have made epistemology matter not less but *more*».

e veridico. Nel contesto della crisi delle grandi narrazioni, il discorso politico-pubblico ha preso la forma retorica di una sistematica sineddoche, in cui la parte vale per il tutto e, cosa più importante, il consenso guadagnato su un punto vale per l'intero (di volta in volta, il tema dell'immigrazione vale per le politiche di sicurezza, i dazi per la politica economica, l'intelligenza artificiale per l'innovazione tecnologica, ecc.)²⁹. La semplificazione o l'impovertimento del linguaggio, che passa attraverso la frammentazione del discorso, è una forma, forse meno visibile ma non per questo meno pericolosa, di disinformazione, perché ancora una volta non punta direttamente ai contenuti della comunicazione come tale, quanto piuttosto a segnalare un'adesione emotiva a messaggi attraverso cui si esprime e segnala lealtà a valori, leader o ideologie, rafforzando meccanismi identitari³⁰. Vi è anche, inoltre, una dimensione *infrastrutturale* della disinformazione che investe le condizioni o regole che rendono possibile il discorso. Da questo punto di vista, non solo le piattaforme ma più complessivamente la rete svolge un ruolo di mediazione nell'accesso, fruizione e diffusione delle informazioni, che può definirsi *infrastrutturale* perché non incide solo sui contenuti ma anche sulle condizioni o regole del discorso³¹. Ciò avviene da almeno due punti di vista, uno che concerne il ruolo di mediazione delle piattaforme e l'altro che investe le regole epistemiche che governano la comunicazione e la ricerca della verità. Dal primo punto di vista, vale la pena sottolineare che il Dsa o più in generale la regolazione europea relativa alla disinformazione

29. Analogamente, M.-P. Lynch, *On Truth in Politics*, cit., p. 23, osserva che: «More pointedly, political fights over “facts” can quickly become fights over whose political interpretation – whose political meaning – becomes dominant in a community. In recent history, for example, such debates are happening across the globe over the political meanings of the terms “immigration” and “genocide” and in the United States over “critical race theory” and “antiracism”. The differing associations people make with these terms affect how they frame the facts. And of course, such conflicts also generate *new* political meanings».

30. M.-P. Lynch, *On Truth in Politics*, cit., pp. 44-45: «Political discourse is performative whenever (a) it is used mostly to express only commitments and emotions and (b) those expressions are intended to signal loyalty to a political identity, leader, or ideology».

31. Questo è sempre stato vero anche nel mondo analogico; e spesso vi è un'asimmetria tra il mero rispetto e la violazione delle condizioni di possibilità del discorso. Il rispetto di una condizione procedurale (ad esempio, parlare uno alla volta) non incide necessariamente sul piano sostanziale dei contenuti, nel senso che il rispetto di tale regola produce un discorso ordinato ma non per forza un discorso informato, veridico e sensato. La violazione della stessa condizione procedurale (parlare tutti insieme) incide, invece, anche sul piano sostanziale, poiché produce non solo un discorso disordinato, ma anche insensato. Si tratterebbe, dunque, di esaminare quali sono le condizioni procedurali del discorso, la cui violazione è suscettibile di incidere anche sui contenuti della comunicazione online, alla stregua della distinzione proposta in altro contesto da Luciano Floridi tra etica e infraetica. Vedi L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, 2017, p. 218 ss.; e dello stesso autore, *Infraethics*, in *The Philosophers' Magazine*, n. 60, 2013, pp. 26-27.

prestano grande attenzione al tema delle regole e procedure che governano la moderazione e metamoderazione dei contenuti attraverso cui sono mediati e regolati l'accesso, la fruizione e la disseminazione dei contenuti online. Sebbene tali pratiche siano ispirate a principi di trasparenza e accountability, volti a garantire che le piattaforme rendano conto del proprio *modus operandi*, è necessario rilevare che, a monte dei processi di accesso, fruizione e diffusione dei contenuti, il feed informativo, i sistemi di raccomandazione e personalizzazione dei contenuti risultano, in generale, caratterizzati da una persistente opacità algoritmica di carattere sistemico. Tale opacità rende più complessa la valutazione delle procedure di moderazione e metamoderazione che intervengono a valle. Dal secondo punto di vista, occorre invece menzionare una serie di pratiche che sono illustrate, ancora una volta, molto opportunamente da Lynch. Nella sfera pubblica della comunicazione politica assistiamo ormai spesso a fenomeni di palesi menzogne, di disinteresse o disaccordo circa il riferimento alle fonti epistemiche o di arroganza intellettuale al riguardo della possibilità di apprendere qualcosa dagli altri³². Tali pratiche condividono una certa indifferenza verso i contenuti: non si preoccupano, cioè, della veridicità, attendibilità o verificabilità delle proprie affermazioni. Mirano, piuttosto, a mettere in discussione, in termini più radicali, l'opportunità o necessità di regole epistemiche che plasmino le condizioni del discorso e della comunicazione e governino «la ricerca della verità nel giornalismo, nell'istruzione, nella scienza, nella storia e nel diritto»³³.

3.3. *La personalizzazione dei contenuti*

Una delle pratiche più sospette è proprio quella già evocata della personalizzazione dei contenuti. Nonostante tale fenomeno abbia assunto un ruolo centrale, fino a configurare la possibilità di un nuovo e diverso modello di produzione dei contenuti (non più orientato e guidato dall'offerta ma orientato

32. Si v. M.-P. Lynch, *On Truth in Politics*, cit., cap. 7, e in particolare, rispettivamente, pp. 193-194: «Bald-faced political lies are puzzling because, unlike normal lies, their function isn't to deceive. [...] so the bald-faced politic liar is saying: "I don't care what about your facts, your rules of evidence. I have my own and I will impose them on you if I can"»; p. 199: «Another source of epistemic corruption is real or manufactured disagreement over what the epistemic rules should be. This what happens when we disagree not only over values – which is healthy in a democracy – and not only over the facts – which is inevitable – but over which sources and methods which should trust for figuring out what the facts are»; infine, p. 203: «This is the problem posed by *intellectual arrogance*, the psycho-social attitude that you have nothing to learn from anyone else about some subject or subjects because you know it all already».

33. M.-P. Lynch, *On Truth in Politics*, cit., p. 196.

e guidato dalla domanda)³⁴, presenta diversi aspetti che sollevano dubbi sia in relazione alla possibilità di trovarsi di fronte a un diverso modello di business nella produzione dei contenuti sia in relazione alla sua dimensione democratica. Consideriamo proprio quest'ultima più da vicino. Se aderiamo all'idea di Lynch, per cui trattare le persone come libere e uguali nella partecipazione alla costruzione del discorso politico costituisce il nucleo essenziale della sfera pubblica, allora la retorica della personalizzazione dei contenuti, ovvero di contenuti pensati fin dall'inizio per plasmarsi in relazione alle preferenze, idee, valori, interessi o idiosincrasie degli utenti, sembra promuovere una visione delle persone forse come libere ma certo diseguali. In tale prospettiva, i contenuti sono indirizzati verso gli individui che con maggiore probabilità li sceglierebbero, alimentando la sensazione che ciascuno sia libero di scegliere come essere *diseguale*, ossia diverso dagli altri ma al contempo coerente con la propria identità. In tal senso, non si tratta solo di ribadire i limiti – più volte evidenziati da Cass Sunstein³⁵ – legati alla ridotta esposizione a una pluralità di idee differenti dalle proprie, ma di mettere in luce ulteriori profili problematici connessi al modello di business sotteso alla personalizzazione dei contenuti online. I contenuti personalizzati ci danno l'impressione di avere un ruolo *attivo* nella costruzione del discorso, sebbene restiamo selezionati dai contenuti personalizzati che sono indirizzati verso di noi in modo opaco. Infatti, tali contenuti possono essere di volta in volta il sottoprodotto di un'ulteriore selezione (meno trasparente se non invisibile) a partire da più ampi insiemi di contenuti. Tale selezione permette di veicolare verso di noi, sotto forma di contenuti personalizzati, messaggi che, aderendo sistematicamente al nostro sistema di credenze, principi o valori possono di fatto ridurre la nostra tendenza a valutare negativamente le pratiche di moderazione e metamoderazione dei contenuti.

4. Contro-incentivi ed *epistemic agency*

Abbiamo sempre bisogno di regole per verificare il vero e per denunciare il falso: abbiamo bisogno, cioè, di un articolato sistema di criteri, principi, norme e procedure per rendere conto delle pratiche – umane o artificiali – con cui

34. Si v. L. Floridi, *La differenza fondamentale. Artificial agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, 2025, in particolare cap. 12, la cui analisi si muove nel più ampio quadro della valutazione dell'impatto dell'intelligenza artificiale generativa sulla formazione dei contenuti e dell'esigenza di sviluppare un nuovo e autonomo ambito di studio, denominato "content studies".

35. Vedi C. Sunstein, *#Republic.com. La democrazia nell'epoca dei social media*, il Mulino, 2017.

sono veicolate le informazioni per evitare che il vero sia alterato o manipolato e il falso sia fabbricato e veicolato. Si tratta di una strategia reattiva o difensiva, ma non basta: occorre anche una strategia proattiva. Il processo con cui sono prodotte informazioni rilevanti e attendibili, da cui dipende l'enunciazione del vero per mezzo del discorso pubblico, richiede qualcosa di più dell'osservanza di regole; richiede essenzialmente un investimento di tempo, competenza e risorse; e pertanto, richiede una solida struttura di incentivi economici, politici e sociali. La più essenziale forma di contrasto alla disinformazione è e resta la produzione di informazioni rilevanti e attendibili: quel vero inteso come una forma di sapere o conoscenza che emerge da una trama articolata di domande e risposte³⁶, più ampia della mera fatticità o correttezza del contenuto di singoli enunciati valutati in isolamento. Questa trama articolata di domande e risposte richiede però un' «epistemic agency»³⁷ esercitata attraverso infrastrutture epistemiche (giornalismo, università, storia, diritto, ecc.), che sostengono i costi di produzione delle informazioni rilevanti e attendibili poiché sono mosse e sostenute, a loro volta, da una struttura di incentivi propri di una società che situa ancora il riferimento al vero (al sapere, alla conoscenza, all'istruzione, ecc.) *al centro* della sfera pubblica. La questione non è tanto se il vero esista o sia possibile un discorso veridico nella comunicazione politica, ma piuttosto quanto ciò conti per noi. Quale ruolo siamo disposti ad attribuire al vero nei processi di formazione delle opinioni? In altri termini, fino a che punto siamo pronti a investire, in termini di tempo³⁸, risorse e competenze, nella ricerca del vero?

36. Si v. L. Floridi, *Filosofia dell'informazione*, Raffaello Cortina Editore, 2024, e dello stesso autore, *Pensare l'infosfera. La filosofia come design concettuale*, Raffaello Cortina Editore, 2020. In tema v. anche M.-P. Lynch, *The Internet of Us*, cit., p. 16 ss.

37. M.-P. Lynch, *On Truth in Politics*, cit., p. 14 in cui l'«epistemic agency» è definita come: «our capacity for forming, maintaining, and exchanging beliefs based on epistemic reasons – that is, on reasons connected to truth», e inoltre, pp. 80-81: «These are the kinds of practices we aspire to employ for example, in scientific, historical, educational, journalistic, and legal institutions – they are the basis of what I'll later call our epistemic infrastructure. [...] Responsible epistemic agency consists in using such practices to pursue having true beliefs».

38. Disponiamo di strumenti di intelligenza artificiale sempre più capaci di creare versioni sintetiche dei contenuti, generando ciò che potremmo definire come una vera e propria *cultura dell'epitome*. Tali strumenti entrano inevitabilmente in concorrenza con forme più articolate e complesse di conoscenza e sapere. Ciò costituisce il vero nodo della controversia tra il *New York Times* e OpenAI, che non concerne solo la violazione del copyright ma anche il ruolo centrale dell'informazione nella democrazia. La disinformazione non deriva, infatti, solo dalla veicolazione di rappresentazioni false o distorte della realtà, ma anche dalla diffusione di versioni eccessivamente semplificate della realtà stessa. Si tratta di un impoverimento del linguaggio e della complessità del sapere, qual è quello che si verifica quando si sostituisce la lettura integrale di un articolo di giornale con quella di un suo riassunto, disponibile online in tempo reale, per tacere di saggi o libri.

Vi sono, nelle nostre società digitali dell'informazione, incentivi abbastanza solidi, o contro-incentivi efficaci, capaci di contrastare gli incentivi economici e politici che alimentano la disinformazione? Il problema, infatti, non consiste solo nel chiedersi se esistano norme adeguate a tutelare l'informazione, ma se da parte nostra esista ancora una domanda di verità sufficientemente forte da giustificare l'offerta.

Per Lynch, come per molti di noi, la risposta non è scontata: il problema non è limitato all'impatto, pur considerevole, della disinformazione. Sono le infrastrutture epistemiche a essere oggi principalmente sotto attacco³⁹ e, con esse, l'opportunità stessa del riferimento al vero. In questo quadro, occorre riaffermare la centralità della verità nella sfera pubblica sotto tre profili che Lynch mette bene in evidenza e che richiamerò qui in breve.

In primo luogo, nella misura in cui ci si muove lungo un asse che va dalla disinformazione agli attacchi diretti alle infrastrutture epistemiche (che minacciano la nostra stessa *epistemic agency*) il problema non è più soltanto quello della *censura* delle opinioni, esercitata tramite strumenti algoritmici di moderazione, oscuramento o rimozione dei contenuti. Diventa invece centrale, e a mio avviso in modo ancora più insidioso, quello dell'*autocensura*. È in gioco, per dirlo con Lynch, il potere di «dire la verità al potere»⁴⁰ (*speaking truth to power*): la capacità di affermare il vero di fronte a chi detiene il potere anche quando ciò comporta rischi o conseguenze negative, se non esiziali, per chi parla. La capacità di fungere da limite al potere di disinformare e di tutelare le infrastrutture epistemiche non dipende solo dalla forza o dall'autorevolezza con cui il falso è denunciato, ma anche e forse soprattutto dal fatto che tale denuncia paghi un dividendo nel mercato delle idee, riesca a fare presa sul pubblico e assuma un ruolo chiave nella sfera pubblica e nei processi di accountability politica.

In secondo luogo, occorre mettere in discussione l'idea, largamente diffusa, secondo cui il riferimento al vero costituirebbe un freno al discorso pubblico: qualcosa che ne interrompe il corso, chiude la conversazione⁴¹ e finisce per

39. Si v. M.-P. Lynch, *On Truth in Politics*, cit., cap. 7. Epistemic Infrastructure and Democratic Hope, in particolare p. 197: «Epistemic infrastructure is what allows us to exercise our epistemic agency, and it is what allows for the growth of responsible epistemic communities. It is only by developing such communities and using such practices that we can hope to achieve a democracy that is a space of reasons. From this perspective, it makes absolute sense (if you can call it that) for antidemocratic forces in society to attack epistemic infrastructure».

40. M.-P. Lynch, *On Truth in Politics*, cit., in particolare p. 4: «Functioning democracies require their residents to speak truth to power. Speaking truth to power means saying what is true even when it is uncomfortable or dangerous».

41. M.-P. Lynch, *On Truth in Politics*, cit., cap. 3: *Can Truth be a Democratic Value?*, in cui si sottolineano i rischi di un riferimento autoritario o antidemocratico al vero, p. 60: «As Estlund

minacciare il pluralismo delle opinioni e forme della comunicazione. Al contrario, il riferimento al vero deve intendersi come una condizione essenziale dell'ordine stesso del discorso. Dire che qualcosa è vera non significa invocare un principio dogmatico o sottratto alla discussione, ma piuttosto affermare che essa è, in linea di principio, verificabile e controvertibile, e pertanto esposta al controllo, alla critica e al giudizio dell'altro. In tal senso, il riferimento al vero è parte integrante di un'etica del discorso, che presuppone e riconosce l'altro come interlocutore reale, che partecipa allo scambio discorsivo in quanto soggetto libero ed eguale, capace di contribuire alla determinazione di ciò che è ritenuto vero⁴². Pertanto, lungi dal comprimere il pluralismo, il riferimento al vero ne costituisce una garanzia essenziale, poiché contrasta quella sistematica destituzione di legittimità dell'interlocutore che, negando la possibilità di un terreno comune di valutazione e confronto, finisce per svuotare il pluralismo di significato e ridurlo a mera giustapposizione di posizioni incomunicabili.

In terzo luogo – ed è forse questo il messaggio più incisivo che emerge dalla riflessione di Michael Lynch – occorre prendere fino in fondo sul serio l'idea secondo cui la verità, qualunque cosa sia, non dipende solo da me⁴³. Non si tratta semplicemente di riaffermare e proteggere la capacità degli individui di accedere al vero e di dire il vero, in opposizione ai processi di parcellizzazione del vero e di metamorfosi del falso più volte denunciati. Si tratta piuttosto di ribadire che ogni impresa conoscitiva, ogni pratica del sapere – inclusa la stessa formazione del vero – ha un carattere intrinsecamente collettivo: essa richiede la partecipazione, il confronto e il contributo di una pluralità di soggetti, senza i quali il vero non può né emergere né essere stabilmente riconosciuto. La dimensione collettiva e plurale della verità costituisce il fondamento normativo di una responsabilità epistemica condivisa (fatta di procedure, istituzioni e infrastrutture), senza la quale né la conoscenza né il discorso possono reggere nel tempo. D'altra parte, la cooperazione, la progettualità e la possibilità di ela-

puts it, just because you know more than I do doesn't, by itself, make you the boss of me», e in particolare pp. 65-77. Il testo di David Estlund a cui Lynch si riferisce è il seguente: D. Estlund, *Democratic Authority*, Princeton University Press, 2008.

42. M.-P. Lynch, *On Truth in Politics*, cit., in particolare pp. 77-87. Per Lynch, il nodo del valore politico della verità riguarda il riconoscimento dell'altro all'interno dell'economia del discorso. Il problema non è tanto il disaccordo sul contenuto di verità di una proposizione – che in democrazia è sempre possibile e, in misura variabile, ragionevole – quanto piuttosto le conseguenze derivanti dal mancato riconoscimento dell'interlocutore come soggetto libero ed eguale nel discorso. Ciò equivale a escluderlo dalla legittima partecipazione alla formazione del discorso, riservando tale facoltà a chi detiene il potere di definire (anche) ciò che in tale discorso è in gioco. Si pensi a un politico che possa scegliere – sistematicamente – a quali domande o a quali giornalisti rispondere.

43. Si v. in tal senso, M.-P. Lynch, *The Internet of Us*, cit.

borare piani comuni sarebbero impensabili senza poter fare affidamento su ciò che riconosciamo mutuamente come vero. Come infatti osserva Lynch, «buona parte dell'attività cooperativa umana dipende dall'assunto che le persone si dicano reciprocamente la verità»⁴⁴.

In definitiva, la rimozione della verità dall'ordine del discorso non implica solo il rifiuto di un possibile accordo con la realtà, ma comporta anche la negazione di un accordo condiviso con gli altri circa le modalità con cui procedere all'accertamento del vero. Ciò vale a maggior ragione se si tiene conto del fatto che dire il falso non dipende solamente da un'asimmetria informativa rispetto all'altro, ma anche e talvolta soprattutto da un'asimmetria di potere, che rende possibile accordarsi con alcuni per affermare il falso a detrimento di terzi⁴⁵.

5. Conclusioni

Il rinnovato richiamo alla centralità della verità come valore democratico, come bussola per navigare il processo di disinformazione in corso, è certamente importante, pur nella consapevolezza o forse proprio in ragione della consapevolezza che assistiamo a una diversa articolazione dei meccanismi di disinformazione (che includono la parcellizzazione del vero e la metamorfosi del falso) e a un attacco sempre più deciso e sistematico alle tradizionali infrastrutture epistemiche delle società dell'informazione.

O, per dirlo diversamente, il contrasto alla disinformazione non può essere adeguatamente pensato e concretamente attuato, se non iscrivendolo in questo più ampio contesto relativo al ruolo che hanno per noi l'informazione, la conoscenza e la verità nella formazione della sfera pubblica. Siamo principalmente noi, come soggetti epistemiche, a dover fornire contro-incentivi, in termini di risorse, tempo e competenze, agli incentivi politici ed economici alla disinformazione. Da soli, però, non possiamo fare tutto, per cui non intendo disconoscere l'opportunità di promuovere una governance etico-giuridica dell'informazione, fondata su tre principali direttrici, a cui abbiamo fatto indirettamente riferimento nel corso del testo e che intendono promuovere forme di:

1. trasparenza algoritmica, intesa non solo come disclosure tecnica, ma come intelligibilità democratica dei processi decisionali automatizzati per contra-

44. M.-P. Lynch, *La verità e i suoi nemici*, cit., p. 218.

45. Ho provato a tracciare la distinzione tra una concezione *intersoggettiva* della menzogna, basata sull'asimmetrie informative, e una *sociale*, che fa leva piuttosto sull'asimmetrie di potere, in M. Durante, *The Democratic Governance of Information Societies*, cit.

stare l'opacità sistemica alla base dei processi di moderazione, veicolazione e personalizzazione dei contenuti;

2. educazione digitale critica (in termini di *AI literacy*)⁴⁶, per restituire alle persone la capacità di discernere le nuove modalità e forme di disinformazione e di assumere maggiore responsabilità epistemica nel consumo e nella diffusione di contenuti informativi;
3. responsabilità condivisa, che coinvolga piattaforme, legislatori e utenti in una rete di corresponsabilità orientata dalla consapevolezza che la ricerca del vero, nello spazio delle ragioni, è un'opera collettiva da attuare tra soggetti liberi ed eguali che si riconoscono reciprocamente come interlocutori reali nella sfera pubblica.

In definitiva, la sfida del falso digitale non si esaurisce in una questione tecnologica o giuridica, ma assume una dimensione epistemologica e politica: chiama in causa il rapporto tra verità, potere, libertà, regole e sistemi di incentivi e disincentivi. Senza riportare al centro del dibattito il valore della verità e dell'informazione di qualità – che ieri come oggi richiedono tempo, risorse e competenze – il ricorso esclusivo a soluzioni normative o tecnologiche rischia di rivelarsi inefficace, se non addirittura controproducente. La verità ha questo di buono: costituisce l'antidoto più efficace contro la tentazione di credere che la realtà debba conformarsi alle nostre pretese o aspettative.

46. Si v. di recente in tema, L. Paseri, M. Durante, U. Pagallo, *The Legal Challenges of AI Literacy Between Enforcement and Compliance*, in *MediaLaws*, n. 1/2025, pp. 327-344.

LA CURA CONTRO LA DISINFORMAZIONE. RIMEDI GIURIDICI (E NON) PER CONTRASTARNE LA DIFFUSIONE

*Federica Fabrizzi**

SOMMARIO: 1. La disinformazione come problema di diritto costituzionale – 2. L’approccio normativo: rallentare la circolazione del *morbo* – 3. Il *debunking* professionistico, medicina contro la disinformazione – 3.1. Gli esempi di NewsGuard e Pagella Politica – 3.2. Una medicina poco efficace – 4. Il *prebunking* e l’educazione, vaccini contro la disinformazione – 5. L’autodeterminazione come obiettivo di educazione costituzionale.

1. La disinformazione come problema di diritto costituzionale

Uno dei criteri principali di valutazione dello “stato di salute” di un ordinamento liberal-democratico è che al suo interno sia garantito uno spazio aperto e plurale, nel quale sia cioè assicurato il libero dispiegarsi di tutte le opinioni e di tutte le posizioni, persino – e, anzi, proprio e soprattutto – di quelle più eterodosse, maggiormente anticonformiste e antisistema¹.

La libera circolazione delle idee è diritto individuale fondamentale (è «garanzia delle garanzie», condizione per il mantenimento di ogni altra libertà, come ha scritto Vigevani²), ma è altresì valore fondante dell’ordinamento democratico giacché – come la giurisprudenza costituzionale ha più volte sottolineato – è su di essa che si costruisce un’opinione pubblica «avvertita e consapevole»³. Essendo la democrazia costituzionale quel regime fondato sul

* Professoressa associata di Diritto costituzionale e pubblico, Sapienza Università di Roma.

1. Come richiamato in dottrina, esiste giurisprudenza chiara su questo punto. Scrive M. Plutino, *Verità e opinione pubblica*, Rubbettino, 2024, p. 88, richiamando anche la giurisprudenza della Corte Edu: «la libertà di espressione è chiamata a proteggere soprattutto i temi minoritari e le critiche scomode ed eterodosse, compreso il dissenso con il politicamente corretto del sentire comune e perciò anticonformista e antisistema».

2. G.E. Vigevani, *Libertà di espressione, onore e controllo del potere. Sviluppi del diritto di critica politica, tra giudice nazionale ed europeo*, in *federalismi.it*, n. 3, 2015, p. 3.

3. Sent. 5/1972.

suffragio universale e sull'uguaglianza del voto, e dunque sulla partecipazione di tutti i cittadini all'assunzione delle decisioni politiche fondamentali, la stessa democrazia costituzionale postula che siano posti in essere tutti i meccanismi possibili per garantire che quella decisione sia, appunto, consapevole e informata. È esattamente sulla base di questo assunto che la Corte di Cassazione ha potuto affermare che il popolo può ritenersi costituzionalmente sovrano «solo se pienamente informato» (Civile, III sez. sent. 16236 del 2010)⁴.

Il tema della disinformazione attiene, dunque, strettamente al diritto costituzionale sotto una duplice prospettiva: per le ricadute e i pericoli che comporta l'espressione di un voto "inconsapevole", e per il conseguente ruolo attivo che l'autorità pubblica deve assumere per garantire «l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese», come prescritto dall'art. 3, secondo comma, della Costituzione italiana⁵.

Sotto il primo profilo, il cardine su cui si regge tutta la ricostruzione dommatica sul punto è l'idea che il suffragio universale possa essere realmente democratico solamente «a condizione che i titolari dei diritti di partecipazione siano messi concretamente nelle condizioni di esprimere un voto libero e consapevole, attraverso quelle che sono state definite le precondizioni liberali e le precondizioni sociali della democrazia»⁶: tra queste precondizioni vi è il godimento di una serie di diritti, tra i quali la libertà di espressione, la libertà di riunione e di associazione, il diritto all'informazione e il pluralismo della stessa, il diritto all'istruzione, esattamente preposti a consentire che tutti si formino autonomamente la propria opinione e compiano le proprie scelte politiche con piena consapevolezza.

4. La pronuncia è riportata da M. Plutino, *op. ult. cit.*, p. 90. Sul diritto della collettività a ricevere informazioni veritiere, esaurienti e complete al fine della formazione dell'opinione pubblica e, quindi, del buon funzionamento del circuito democratico-rappresentativo, anche la Corte costituzionale si è più volte espressa. Cfr. il documento a cura di G. Nicastro, *Libertà di manifestazione del pensiero e tutela della personalità nella giurisprudenza della Corte costituzionale*, maggio 2015, disponibile su: www.cortecostituzionale.it/documenti/convegni_seminari/stu_284.pdf.

5. Così G.E. Vigevani, *op. ult. cit.*, p. 4.

6. E. Grosso, *Malattia mentale e suffragio universale: le aporie della democrazia rileggendo "La giornata di uno scrutatore" di Italo Calvino*, in *costituzionalismo.it*, n. 2/2018, p. 11; ma anche E. Spagna Musso, *Lo stato di cultura nella costituzione italiana*, Morano, 1961, p. 47, per cui «la democraticità di un ordinamento statale è garantita dalla effettiva partecipazione dei cittadini alla scelta dei fini ultimi dell'azione statale» e tale effettività di partecipazione può verificarsi in presenza di due condizioni: «sotto un profilo quantitativo ed esteriore, la totalità di partecipazione, cioè l'intervento di tutti i cittadini senza distinzioni di sorta; sotto un profilo quantitativo e interiore, invece, la libertà e la maturità di partecipazione ovvero l'idoneità di ogni cittadino a realizzare una scelta libera e cosciente fra le varie prospettive d'orientamento e di indirizzo sottoposte al suo giudizio».

Per quanto attiene al secondo profilo, secondo quanto riportato da Giovanni Sartori nella voce “Opinione pubblica” dell’enciclopedia Treccani, questa non può dirsi innata, ma si identifica con «un pubblico, o una molteplicità di pubblici, i cui stati mentali diffusi (opinioni) interagiscono con flussi di informazione sullo stato della cosa pubblica»⁷. Questo flusso di informazioni può avvenire in vario modo, dagli incontri *face-to-face* nella vita quotidiana, agli eventi pubblici, ma è fuori di dubbio la rilevanza che la comunicazione pubblica gestita dai *mass media* incarna, essendo quello «l’unico spazio in cui il frastuono comunicativo può compattarsi in opinioni pubbliche rilevanti ed efficaci»⁸.

Per questa ragione, la predisposizione di risposte adeguate da parte della autorità pubblica al possibile inquinamento nel processo di formazione del consenso non è in discussione.

Quale (o, meglio, quali) siano i rimedi è, invece, questione assai più controversa e delicata. In altri termini, a voler inquadrare la disinformazione come una “malattia”, un protocollo medico che certifichi la “cura efficace” ancora non esiste e non è neanche troppo chiaro chi debba essere il “medico”.

Naturalmente, si danno qui per conosciute le diverse accezioni che il termine “disinformazione” incarna⁹: la tassonomia dell’*information disorder* è noto che prenda in considerazione fenomeni tra loro anche molto distanti, dalla limitazione del diritto all’informazione generale, alla manipolazione dell’opinione pubblica da parte di Paesi ostili, passando per la destabilizzazione politico-sociale e l’infiltrazione nel circuito dell’informazione e della sicurezza pubblica. La latitudine coperta già di per sé dice che è difficile, se non addirittura sbagliato, immaginare la medesima cura; taluno, a dire il vero, potrebbe persino mettere in discussione la stessa natura patologica della disinformazione, quanto meno nella sua accezione più generale. È stato, infatti, fatto acutamente notare che, in ragione della diffusione della tradizione politico-culturale liberale, «la dimensione “orizzontale” della tutela del diritto di informazione è stata per lo più circoscritta [...] alla protezione, anche in sede civile, dei diritti della personalità (nome, immagine, integrità morale, onore e reputazione, più avanti identità personale e riservatezza) nei confronti di chi sia individuato come civilmente responsabile dei danni cagionati diffondendo informazioni false o riservate»; ma quella stessa tradizione liberale ha portato a considerare

7. La voce è consultabile su: [www.treccani.it/enciclopedia/opinione-pubblica_\(Enciclopedia-del-Novecento\)/](http://www.treccani.it/enciclopedia/opinione-pubblica_(Enciclopedia-del-Novecento)/).

8. J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Raffaello Cortina Editore, 2023, p. 39.

9. A. Ruffo, *Il disordine informativo e l’Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, numero speciale I-2024, p. 407 ss.

di meno immediata comprensione «l'idea che i diritti fondamentali possano essere lesi, e perciò meritino di essere protetti, anche da altri soggetti privati che tengano comportamenti i quali peraltro, rientrando nella loro sfera di libertà rispetto al potere politico, in prima approssimazione dovrebbero essere inquadrati come atti di esercizio di un loro diritto»¹⁰.

D'altra parte, occorre anche riconoscere che, ad esempio per la Costituzione italiana, «non dire la verità (nel senso di tacerla o di falsificarla), rientra nell'esercizio della libertà di espressione, che proprio in quanto libertà esclude in linea di principio l'obbligo di manifestare il proprio pensiero (art. 21 Cost.)»¹¹.

La questione è certamente di non poco conto e chiama in causa la riflessione non solo sul concetto stesso di "verità", ma altresì sulla esistenza di un diritto alla verità. Come noto, anche chi ha teorizzato l'esistenza dei diritti aletici ha messo correttamente in guardia sulla "vaghezza giuridica" che caratterizza questo tema: «l'imprendibilità di principio del manipolatore (il manipolato è libero di non farsi manipolare) e l'estensione del suo procedimento (dove si colloca, esattamente, la frode?), rendono – è stato detto – la materia estremamente vaga»¹².

Il nodo cruciale che permea tutta la questione è, in ultimo, dove si colloca il confine: sono le zone grigie, come sempre, a creare il problema, perché laddove è evidente cosa è bianco e cosa è nero è fin troppo facile regolare; ma se si ragiona di danno, colpa o dolo epistemici, cioè riguardanti il conoscere e il credere, «le credenze indotte dalla manipolazione sono di solito controverse: coinvolgono teorie religiose, oppure interpretazioni complesse in cui il falso si mescola al vero e al mezzo-vero»¹³, ed è esattamente lì che le certezze vacillano¹⁴.

10. G. Magrì, *Diritto di informazione e dovere di verità al tempo di Covid-19*, in *Giustizia Insieme*, 26 marzo 2019.

11. M.R. Allegri, *Riflessioni sulla verità del diritto e nel diritto: per una società aletica*, in *Nomos. Le attualità nel diritto*, n. 2/2025, p. 19.

12. F. D'Agostini, *Diritti aletici*, in *Biblioteca della libertà*, n. 1/2017, p. 8.

13. *Ibidem*.

14. Molto interessanti sono, a questo proposito, le riflessioni svolte da A. Facchi, *La verità come interesse collettivo*, in *Biblioteca della libertà*, n. 1/2017, p. 64. L'A. riferendosi alla necessità di immaginare una verità che si nutre del discorso scettico non di quello dogmatico, scrive: «Mi chiedo se non ci sia una profonda differenza tra quella concezione di verità e il tipo di verità che può essere veicolato attraverso il diritto. In qualche modo un'intrinseca incompatibilità tra quella verità dialettica, scettica e la verità normativa del diritto. Il diritto non è scettico, il diritto deve decidere, concludere con una scelta e questa scelta è corredata dalla forza. C'è un momento, che sia la norma o la decisione giudiziaria, in cui la discussione si chiude, la verità è quella, va accettata o viene imposta».

A ben vedere, al fondo vi è l'ulteriore interrogativo filosofico di come si possa arginare la proliferazione del falso salvaguardando il pluralismo, atteso che proprio il pluralismo sembra essere l'unico rimedio in grado di evitare che false credenze si incardinino «diventando ideologie deviate e distruttive»¹⁵. Il pericolo sotteso tutte le volte in cui si immagina di operare una selezione sui contenuti è, infatti, quello della potenziale sterilizzazione del discorso pubblico.

È un po' l'idea degli anticorpi: il corpo umano solo nel momento in cui è esposto ad agenti patogeni produce da sé anticorpi che gli consentono di debellare potenziali malattie; senza l'esposizione, il sistema immunitario resta deficitario. Questo meccanismo funziona, peraltro, in situazioni fisiologiche in cui il corpo sia sano e l'agente patogeno da affrontare non particolarmente virulento; nei casi in cui il corpo sia già debilitato o il virus particolarmente aggressivo, occorre intervenire a supporto del sistema immunitario.

La disinformazione rischia di essere la patologia pandemica di questi anni, la malattia che se non adeguatamente affrontata mette a rischio la stessa sopravvivenza in vita delle democrazie, che già non godono di buona salute.

Non sembri allora azzardato il paragone tra la libertà di informazione e la libertà di salute. Con riferimento a quest'ultima, attenta dottrina ha sottolineato come «la libertà di salute non viene attribuita al singolo *in ragione* dell'interesse collettivo alla salute; ben diversamente, essa viene configurata in modo tale che il suo godimento *non pregiudichi* il corrispondente interesse della collettività»¹⁶. Lo stesso si può dire della libertà di informazione, che si configura come diritto individualistico e non funzionale: l'interesse a che l'informazione contribuisca alla creazione di un'opinione pubblica consapevole *non* è la ragion d'essere della libertà, ma si configura come «limite esterno alle

15. F. D'Agostini, *op. ult. cit.*, p. 7. L'A. afferma che «Tipiche società atleticamente sofferenti sono i regimi totalitari, definiti, da vari autori, come situazioni di "menzogna organizzata". I cittadini vengono mantenuti nella persuasione che i provvedimenti governativi siano giusti sulla base di una sistematica distorsione o cancellazione della verità». Interessante, sotto questo profilo, anche la riflessione di A. Ruggeri, *Costituzione e verità (prime notazioni)*, in *Consulta on line*, n. II/2023, p. 412: «Nessun ordinamento d'ispirazione liberal-democratica può essere edificato sulla menzogna e trasmettersi integro nel tempo a mezzo di essa; possono esserlo (e di fatto molte volte lo sono) ordinamenti di connotazione autoritaria che, però, prima o poi si trovano costretti a fare i conti con la storia e con l'aspirazione, profondamente avvertita dalla comunità governata, all'acquisizione ovvero al ripristino della verità e, con essa, al riconoscimento dei valori fondamentali di libertà, eguaglianza, giustizia sociale, democrazia. Condizione, dunque, necessaria, seppur da sola non sufficiente, per la stessa sopravvivenza degli ordinamenti di democrazia occidentale è la tessitura di una trama istituzionale di rapporti con e tra i pubblici poteri le cui maglie risultino saldamente unite tra di loro dal collante della verità, in ciascuna delle sue espressioni e in tutte assieme».

16. D. Morana, *La salute come diritto costituzionale. Lezioni*, Giappichelli, 2025, p. 49.

concrete forme di godimento di essa che il soggetto titolare deciderà di mettere in atto»¹⁷.

In questa chiave è possibile immaginare interventi limitativi della libertà di espressione giustificati dalla necessità di tutelare il bene collettivo a che si crei un'opinione pubblica consapevole.

Il bene tutelato è, dunque, quello dell'«informazione» che, come ha osservato Luciano Floridi, non può che essere incarnata da un contenuto vero o presunto tale («la falsità – ha scritto – non informa»¹⁸).

Non generica diffusione di notizie, dunque, ma informazione di qualità: il tema oggi non è tanto quello della moltiplicazione delle voci presenti nella sfera pubblica, quanto piuttosto quello di mettere i cittadini nelle condizioni di potersi formare un'opinione consapevole attraverso un'informazione di buona qualità. Nella recente sent. 44/2025 della Corte costituzionale, non a caso si dice che, nel «sovraccarico mediatico» in cui è immerso il cittadino, occorre «tutelare e promuovere la qualità della comunicazione», dove l'utilizzo della coppia di verbi non è certamente casuale, richiamando l'endiadi di cui all'art. 2 (ma anche all'art. 5) della Costituzione.

Mantenere una struttura mediatica che permetta il carattere inclusivo della sfera pubblica e un carattere deliberativo per la formazione dell'opinione e della volontà pubblica è dunque «un imperativo costituzionale»¹⁹, giacché in un contesto di comunicazione *incontrollata*, assai elevato diventa il rischio dell'inconsapevolezza, condizione che nega l'idea stessa di partecipazione attiva e di democrazia.

Nei paragrafi che seguono si prenderanno in considerazione tre modalità di intervento – la normazione, il *debunking* e il *prebunking* – che sono stati e sono utilizzati per contrastare la disinformazione. Per tutti e tre gli approcci si cercherà di evidenziare i vantaggi, ma altresì i limiti, nella convinzione che la strategia di contrasto alla disinformazione comunque non possa che essere «multi-dimensionale»²⁰, capace cioè di includere tanto misure giuridico-sanzionatorie quanto interventi per la valorizzazione e il potenziamento del ruolo degli utenti e dei giornalisti, quanto ancora misure volte a incentivare l'alfabetizzazione mediatica e informazionale dei cittadini. Volendo continuare a stare sulla metafora della disinformazione come malattia, anzi come pandemia da estirpare, l'esperienza ci ha insegnato che serve rallentare la circolazione del

17. *Ibidem*.

18. L. Floridi, *Information. A Very Short Introduction*, Oxford University Press, 2010, p. 2 ss.

19. J. Habermas, *op. ult. cit.*, p. 71.

20. O. Pollicino, P. Dunn, *Intelligenza artificiale e democrazia*, Bocconi University Press, 2024, p. 128.

virus con divieti, curare chi si è già ammalato e vaccinare chi non si è ancora ammalato, con l'obiettivo, ambizioso certamente, di arrivare a una sorte di immunità di gregge cognitiva nei confronti della menzogna.

2. L'approccio normativo: rallentare la circolazione del *morbo*

Il primo intervento che va valutato è quello della regolazione giuridica e della normazione.

È, come noto, esattamente questa la cifra che ha caratterizzato l'approccio dell'Unione europea, segnando peraltro in questo una netta presa di distanza dalle politiche d'Oltreoceano, specialmente dopo l'avvento della seconda amministrazione Trump, che in nome di un'interpretazione molto dilatata del I emendamento ha optato per lasciare campo al libero dispiegamento dei mezzi di comunicazione, inclusi quelli in ambiente digitale²¹.

Lo sforzo regolatorio profuso dall'UE nel corso dell'ultimo triennio in materia di nuove tecnologie è davvero eccezionale; a volere indicare una parola chiave che attraversa tutti gli atti adottati, si potrebbe dire che essa è "*accountability*": il concetto di responsabilità, e quello di trasparenza che a questa si accompagna, è quello che guida, fin dal Gdpr, il legislatore europeo in questo suo tentativo che è, al contempo, di regolare il mercato digitale e di farlo mantenendo fermo il sistema di valori e la protezione dei diritti fondamentali a tutela della dignità umana che si pone a fondamento del costituzionalismo continentale.

La complessità di un'operazione di questo tipo è certificata dalle valutazioni, non sempre univoche, rispetto a tale interventismo.

Se da un lato, infatti, l'approccio europeo viene salutato con assoluto favore e si enfatizzano le prospettive di quello che è stato definito da Anu Bradford il «*Brussels effect*»²², dall'altro resta viva la preoccupazione che un eccesso di regole possa frenare lo sviluppo tecnologico e l'innovazione, relegando l'Europa stessa ai margini (Rapporto Draghi)²³. E se, come detto, l'approccio antropo-

21. Ci si riferisce all'ordine esecutivo del 23 gennaio 2025 che ha introdotte nuove regole che disciplinano lo sviluppo dell'AI negli Stati Uniti, all'insegna del libero mercato e dell'assenza di qualsiasi controllo, superando così l'ordine esecutivo, più restrittivo, emesso nel 2023 da Joe Biden.

22. A. Bradford, *Effetto Bruxelles. Come l'Unione Europea regola il mondo*, FrancoAngeli, 2020.

23. *The future of European competitiveness: Report by Mario Draghi*, disponibile su: commission.europa.eu/topics/competitiveness/draghi-report_en. Nel momento in cui si scrive, peraltro, l'Unione europea ha avviato una riflessione sul cosiddetto Digital Omnibus, primo grande in-

centrico, fondato sulla tutela dei diritti fondamentali, costituisce senza ombra di dubbio la cifra distintiva che connota la via europea alla regolazione²⁴, non sono tuttavia pochi coloro che parlano di iperattivismo dell'Unione, di una ipertrofia del quadro regolatorio che rischia di generare sovrapposizioni, confusioni e caos normativo, rendendo di fatto vano lo sforzo.

Con specifico riferimento al contrasto alla disinformazione, al linguaggio d'odio o ad altre forme di violenza e manipolazione dell'opinione pubblica, l'Unione europea ha seriamente preso in carico il problema adottando diversi atti di *hard law* che, da prospettive diverse ma in modo sinergico, hanno tutti come oggetto il tema del contrasto alla disinformazione.

Gli atti cui ci si riferisce sono assolutamente noti, così come noto è il loro contenuto; sarà dunque sufficiente, in questa sede, semplicemente richiamarli per dare anche plastica visibilità di come lo strumento dell'intervento normativo sia stato pesantemente utilizzato dal legislatore eurounitario per provare a curare il morbo della disinformazione.

Dopo l'approvazione del "Codice rafforzato di buone pratiche sulla disinformazione", che aveva impegnato le 34 società aderenti a una forma di collaborazione istituzionale per arginare fenomeni di disinformazione (in particolare attraverso azioni quali demonetizzare la diffusione della disinformazione, garantire la trasparenza della pubblicità politica, responsabilizzare gli utenti per una consapevole fruizione di Internet, incrementare la cooperazione con i verificatori dei fatti e fornire maggiore accesso ai dati delle piattaforme), il primo Regolamento che ha affrontato la tematica dell'effettiva ed efficace moderazione dei contenuti *online*, introducendo forme di responsabilità per le piattaforme digitali, è stato il *Digital Services Act*. Il Dsa ha adottato un approccio volto al coinvolgimento degli operatori privati, sia con riferimento al controllo sui contenuti diffusi sia anche con riferimento alla veridicità degli account, im-

tervento di semplificazione su regole digitali europee, con modifiche mirate al Gdpr che toccano dato personale, ricerca scientifica, intelligenza artificiale e dati pseudonimizzati, nel tentativo di ridurre oneri e incertezze senza abbassare il livello di tutela dei diritti. La legislazione digitale, infatti, secondo quanto detto nella proposta stessa, appare frammentata, con duplicazioni e complessità amministrative. L'obiettivo dichiarato, pertanto, è quello di ridurre la complessità, armonizzare le disposizioni non allineate e alleggerire i costi di conformità per imprese, amministrazioni pubbliche e cittadini.

24. Scrive F. Pizzetti, *Il Regolamento europeo sulla IA come parte integrante della normativa UE per il decennio digitale europeo 2030*, in S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino, F. Pizzetti, *La regolazione europea dell'intelligenza artificiale nella società digitale*, Giappichelli, 2025, p. 3: «Grazie al principio basilico dell'antropocentrismo, costantemente riaffermato dall'Unione, diventa evidente che tutta la regolazione unionale in materia digitale deve essere interpretata in modo da renderla conforme e coerente con i diritti e le esigenze degli esseri umani».

nendo alle piattaforme precisi obblighi con conseguenti sanzioni pecuniarie, anche molto significative, in caso di omesso controllo.

Come noto, peraltro, il Dsa attribuisce alle cosiddette Vlop (*very large online platforms*) e ai cosiddetti Vlose (*very large online search engines*) importanti funzioni di autoregolamentazione rispetto all'adozione di misure volte a mitigare i rischi sistemici derivanti dai loro servizi – quali la circolazione di contenuti illeciti o pregiudizievoli in rete – attribuendo a essi un ampio margine di discrezionalità. Ed è assai significativo che questa circostanza sia, al contempo, criticata da quanti ritengono che in questo modo si possa creare una falla nella normativa e da quanti, al contrario, guardano con sospetto il fatto che il potere di rimuovere contenuti ritenuti pregiudizievoli sia ricondotto, in buona sostanza, a soggetti privati²⁵.

Sulla stessa linea dell'*accountability* si è posto anche il regolamento (UE) 2024/1689, noto come *AI Act*. Anche in questo caso, dando per nota l'impostazione adottata dal legislatore europeo di calibrare l'intervento sulla base del rischio (*risk based approach*), la disinformazione entra a pieno titolo nella materia oggetto di regolazione: basti pensare al tema dei *deep fake*²⁶, ma anche a fenomeni possibili solo attraverso sistemi di intelligenza artificiale quali *trolls*, *sock puppets* e *sealioner*, tutte pratiche che concorrono come canali più o meno affidabili a diffondere informazioni false o parziali e a carpirne altre, anche riservate, produttive potenzialmente pure di condotte di mala informazione.

La responsabilizzazione passa anche attraverso gli obblighi previsti dall'art. 50 dell'*AI Act*: l'obbligo per i fornitori di sistemi di garantire che i contenuti audio, immagini, video o testi generati con AI siano marcati in un formato leggibile meccanicamente e dunque rilevabili come generati o manipolati artificialmente e l'obbligo per i *deployers* di rendere noto che contenuti generati con AI sono stati generati o manipolati artificialmente.

Sulla scia di quanto stabilito nell'*AI Act*, la legge italiana n. 132/2025 introduce figure di reato e aggravanti volte a contrastare nuove forme di aggressione a beni giuridici realizzate con lo strumento dell'intelligenza artificiale. Con la fattispecie di reato di *Illecita diffusione di contenuti generati o manipolati artificialmente* (art. 612-*quater* c.p.), che punisce «chiunque cagiona un danno ingiusto a una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di

25. Nel momento in cui si chiudevano queste pagine è arrivata la prima multa comminata dalla Commissione europea alla piattaforma X di Elon Musk per violazione delle previsioni contenute negli artt. 29 e 40 del Dsa.

26. Sul fenomeno dei *deepfakes* si veda M.E. Bucalo, *Intelligenza Artificiale e deepfakes: le nuove frontiere della disinformazione e i possibili rimedi giuridici*, in *MediaLaws*, n. 3/2024.

sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità», il legislatore ha chiaramente voluto lasciare un margine di indeterminazione tale da poter immaginare l'applicazione del reato in relazione all'utilizzo di *deepfakes* a scopo di propaganda politica o di disinformazione.

Altro atto fondamentale ai fini del tema che qui ci occupa è lo *European media freedom Act* (UE 2024/1083) che annovera tra i propri obiettivi primari la protezione dei servizi di comunicazione da indebite interferenze e l'incremento della loro trasparenza. Estendendo il campo di applicazione ai fornitori di servizi media digitali e di contenuti online, il Regolamento stabilisce, infatti, un regime di trasparenza sulla proprietà dei media, promuove l'armonizzazione delle procedure di controllo sulle concentrazioni, introduce un sistema trasparente di rilevazione dell'audience esteso ai media digitali, stabilisce il divieto di utilizzare *software* spia nei confronti dei giornalisti e prevede altresì una disciplina *ad hoc* sulla moderazione dei contenuti di informazione dei media da parte delle piattaforme online. Tutti obblighi volti a declinare il principio di trasparenza per promuovere la creazione di un dibattito fondato sul pluralismo e sulla qualità dell'informazione.

Da ultimo, anche il Regolamento sulla pubblicità politica (UE 2024/900) si occupa di contrasto alla disinformazione puntando esattamente sulla trasparenza. Per riassumere la *ratio* del provvedimento, sia consentito riportare quanto scritto nel considerando 74, nel quale si trova la spiegazione del perché la pubblicità politica, e in particolare le tecniche di *targeting* usate per rivolgere messaggi di pubblicità politica, debba essere attentamente controllata e soggetta a verifiche:

Trattando dati personali, segnatamente categorie particolari di dati personali ai sensi del regolamento (UE) 2016/679 e del regolamento (UE) 2018/1725, è possibile segmentare diversi gruppi elettorali o di privati cittadini e sfruttarne le caratteristiche o vulnerabilità, ad esempio diffondendo messaggi di pubblicità politica in momenti e luoghi ad hoc per trarre vantaggio da situazioni in cui potrebbe essere più acuta la sensibilità a un certo tipo di informazione/messaggio. Tale trattamento dei dati personali genera un impatto specifico e negativo sui diritti e sulle libertà fondamentali delle persone, come il diritto di essere trattate in modo equo e paritario, non essere manipolate, ricevere informazioni obiettive, farsi un'opinione, prendere decisioni politiche ed esercitare il diritto di voto. Inoltre, incide negativamente sul processo democratico, in quanto conduce a una frammentazione del dibattito pubblico su importanti questioni sociali, una comunicazione selettiva e, in ultima analisi, la manipolazione dell'elettorato. Tutto ciò aumenta anche il rischio di diffusione della manipolazione delle informazioni e delle ingerenze straniere. La pubblicità ingannevole od occulta per scopi politici costituisce un rischio in quanto influenza i meccanismi fondamentali che permettono il funzionamento della nostra società democratica.

A prescindere dalla valutazione sui singoli provvedimenti, sulle loro potenziali sovrapposizioni, sulla loro effettività, quel che preme rimarcare è come la risposta legislativa sia stata pensata in chiave di controllo e protezione. L'obiettivo dell'Unione europea è stato ed è, come detto, quello di accrescere l'*accountability* delle piattaforme promuovendo una maggiore trasparenza, anche con riferimento ai sistemi di raccomandazione che le stesse piattaforme utilizzano al fine, appunto, di non lasciare l'operato di coloro che sono ormai protagonisti del sistema informativo completamente senza controllo e al fine di proteggere gli utenti nei loro diritti fondamentali.

Un intervento volto, dunque, a responsabilizzare, attraverso obblighi e sanzioni, e a incentivare trasparenza e controllo, per creare un ambiente digitale in cui il morbo della disinformazione non circoli più, o quanto meno circoli molto di meno.

3. Il *debunking* professionistico, medicina contro la disinformazione

La pratica di esaminare la rispondenza al vero di una informazione è solitamente chiamata *fact-checking*, termine che è balzato agli onori della cronaca insieme al suo corrispettivo negativo ("*fake news*"), sebbene, esattamente come il suo corrispettivo negativo, indichi una pratica antica quanto l'uomo. Il maggiore proliferare delle *fake news* in ambiente digitale – benché anche su questo non vi sia unanimità di pareri²⁷ – ha portato anche alla emersione e valorizzazione delle attività di controllo, a quella operazione nota con il nome di *debunking*.

Forse in modo un po' approssimativo, i termini *fact-checking* e *debunking* sono spesso considerati sinonimi; questo è vero solo in parte. Mentre, infatti, il termine *debunking* ha a che fare con il verbo "smascherare, smontare", indicando dunque un'azione che viene fatta *dopo* la diffusione della notizia, la pratica informativa volta a esaminare la rispondenza al vero di un fatto o di una notizia può essere svolta anche *prima* che questa venga diffusa. Anzi, il cosiddetto *fact-checking* interno dovrebbe rientrare tra i doveri professionali di chi fa comunicazione: basta rileggere alcuni passi della famosa "sentenza decalogo" della Cassazione²⁸, che ha ormai più di quarant'anni, segnatamente là dove si parla del requisito della verità dei fatti, oggettiva, o anche solo putativa, purché frutto di un diligente lavoro di ricerca. Quella stessa sentenza afferma, altresì, che la ve-

27. Cfr. C. Budak, B. Nyhan, D.M. Rothschild *et al.*, *Misunderstanding the harms of online misinformation*, in *Nature*, 630, 2024 pp. 45-53.

28. Corte cass. I civ. 18 ottobre 1984, n. 5259

rità incompleta deve essere equiparata in tutto alla notizia falsa: «la verità dei fatti, cui *il giornalista ha il preciso dovere di attenersi*, non è rispettata – si legge nella pronuncia – quando, pur essendo veri i singoli fatti riferiti, siano, dolosamente o anche soltanto colposamente, taciuti altri fatti, tanto strettamente ricollegabili ai primi da mutarne completamente il significato. La verità non è più tale se è “mezza verità”» (enfasi nostra).

La cosiddetta “piattaformizzazione” della sfera pubblica ha comportato e comporta almeno due rilevanti problemi con riferimento a quanto appena detto²⁹. Il primo, e principale, è che la narrativa prevale ormai sulla fattualità. Come notato da Han: «L’informazione circola ormai completamente scollegata dalla realtà, in uno spazio iperreale. Si perde la fiducia nella *fattualità*. Viviamo quindi in un universo *de-fatticizzato*»³⁰; non a caso si parla di era della post-verità, termine utilizzato già nel 1992 per indicare un’era in cui «circumstances in which objective facts are less influential in shaping political debate or public opinion than appeals to emotion and personal belief»³¹.

L’altro aspetto attiene al compito dei media tradizionali e alla tendenza alla deprofessionalizzazione del ruolo di giornalista. La “piattaformizzazione”, intesa nel senso di una richiesta di maggiore rapidità dell’informazione, ha infatti condotto a un cambiamento dal punto di vista degli standard professionali, complice anche l’idea che “tutti” possono fare informazioni in ambiente digitale. Le ragioni, di ordine sociale ma anche economico, di questo cambio sono note:

La diversa struttura e organizzazione delle redazioni online, che sono legate a ritmi di produzione più elevati rispetto ai tradizionali media cartacei o televisivi; gli incentivi in termini di visibilità e diffusione verso contenuti che possano raggiungere in fretta il

29. Sulla “piattaformizzazione”, si veda tra gli altri, Tarleton Gillespie, *Custodians of the Internet*, Yale University Press, 2018.

30. B.C. Han, *Infocrazia*, Einaudi, 2023, p. 60.

31. Il processo per il quale le società contemporanee si sono progressivamente trasformate nelle loro rappresentazioni, perdendo quindi di autenticità, è oggetto di studio dunque già da più di un trentennio, e non occorre in questa sede tornare a sottolineare quanto questo elemento influenzi la corretta dinamica della sfera pubblica. Come ricorda M.R. Allegri, *op. ult. cit.*, p. 6, «L’uso dell’espressione post-truth (post-verità) risale almeno al 1992, quando il giornalista e drammaturgo serbo-americano Steve Tesich l’ha utilizzata in un articolo pubblicato il 6 gennaio dalla rivista *The Nation* dal titolo *A Government of Lies*. L’espressione [...] è diventata però di uso comune a partire dal 2016, quando l’*Oxford English Dictionary* la ha eletta “parola dell’anno”». Essa indica, secondo la definizione fornita da A. Besussi, *La doppia natura della verità*, in *Biblioteca della libertà*, n. 1/2017, p. 48, «circostanze in cui asserzioni e credenze non devono rispondere ad alcuna attendibilità fattuale per avanzare pretese di verità. In queste circostanze accade che discorsi falsi esigano di essere trattati come veri e che non-fatti esigano di essere considerati come fatti alternativi».

maggior numero di persone possibili; e infine la crisi economica che investe il settore dei media ormai da alcuni decenni sono tutti elementi che rendono più complesso o meno accurato il lavoro di verifica delle informazioni da parte dei media³².

Nella metamorfosi del ruolo del giornalista, quello che viene in rilievo è, dunque, la ormai quasi completa assenza di controllo sulla provenienza delle notizie e di verifica sulla veridicità delle stesse: il *fact-checking* interno è operazione ormai sempre più rara.

Occorre tuttavia ammettere che, se pure questo aspetto certamente contribuisce allo scadimento nella qualità dell'informazione, la disinformazione, soprattutto quella sistemica e organizzata, passa ormai per altre strade e diventa virale grazie a meccanismi di automazione. In questo senso, l'avvento dell'intelligenza artificiale ha rappresentato un deciso cambio di passo sia dal punto di vista della produzione di falsi che dal punto di vista della loro diffusione.

Al cospetto di questo cambio di passo, lo stesso sistema mediatico ha però ritenuto di dover avviare una serrata attività di verifica delle notizie ed è stato lo stesso mondo del giornalismo a porsi il problema di creare una rete di collaborazione internazionale: a partire dal 2014, la scuola di giornalismo statunitense Poynter Institute, sotto la spinta del fondatore di PolitiFact Bill Adair, è stata la principale organizzatrice di una serie di convegni, chiamati Global Fact, che hanno avuto l'intento di riunire periodicamente i responsabili dei diversi progetti di *fact-checking* a livello globale. Nella prima edizione del Global Fact, i circa cinquanta partecipanti decisero la costituzione di un'organizzazione internazionale dei *fact-checker*, che sarebbe nata l'anno successivo con il nome di International Fact-Checking Network (Ifcn), avente sede presso lo stesso Poynter Institute.

Nel 2016 – l'anno della Brexit e della elezione di Donald Trump – della rete internazionale dei *fact-checker* entra a far parte anche Facebook, ad anticipare quella che sarà poi una previsione del Dsa, ossia che le piattaforme debbano collaborare con i *fact-checkers* per mitigare i rischi della disinformazione e garantire un "contributo finanziario equo" per sostenere il loro lavoro. Peraltro, la decisione adottata da Meta nel gennaio 2025 di ridimensionare il proprio programma di *fact-checking* negli Stati Uniti, pur mantenendolo per il momento nell'Unione Europea proprio per conformarsi alle normative del Dsa, mette seriamente in discussione il "*Brussels effect*", dimostrando come, al contrario, le differenze territoriali continuo e come le stesse piattaforme, laddove lasciate libere, preferiscano non controllare e non collaborare.

32. L. Serafini, G. Zagni, *Fact-checking. Epistemologie e pratiche informative: un'analisi sociologica*, FrancoAngeli, 2024, p. 21

Il diretto coinvolgimento delle piattaforme nelle attività di *fact-checking* è giustificato dalla circostanza che gli stessi sistemi automatizzati possono essere di grande aiuto nell'individuare situazioni anomale di diffusione di notizie false: esistono studi empirici dai quali si evince che immagini e contenuti audio, video o testuali prodotti o manipolati da AI hanno un'impronta sintetica riconoscibile³³ che sistemi automatizzati possono facilmente intercettare.

Naturalmente l'automazione nell'identificazione della disinformazione non è esente da criticità: da un lato la mancata accuratezza, l'incapacità di cogliere le sfumature di significato, la difficoltà di comprendere gli aspetti di contesto e i riferimenti culturali, portano al rischio dei cosiddetti "falsi negativi" (ma anche dei "falsi positivi")³⁴.

Dall'altro, l'aspetto sul quale maggiormente ci si interroga è quello della commistione dei ruoli, ossia di una sovrapposizione tra controllore e controllato che rischierebbe di rendere vano il processo o comunque poco trasparente. Un sistema di *fact-checking* che voglia dirsi trasparente e indipendente dovrebbe, infatti, avere almeno due caratteristiche: in primo luogo dovrebbe essere realizzato da soggetti distinti rispetto ai soggetti che diffondono le informazioni; in secondo luogo, dovrebbe essere realizzato da soggetti qualificati, essendo del tutto evidente che non è sufficiente autopresentarsi come *fact-checkers* per essere attendibili e affidabili. L'affidabilità e l'autorevolezza, data dalla padronanza di un metodo rigoroso, devono poi essere accompagnati anche dalla trasparenza, nel senso che deve essere chiaro quale sia la ragione che spinge a fare attività di *debunking*.

3.1. Gli esempi di NewsGuard e Pagella Politica

Ora, sotto il profilo dei soggetti istituzionali, quale sia la ragione che spinge a fare attività di "smascheramento" è facilmente rinvenibile in quanto si è ricordato in apertura di queste riflessioni. Esistono tuttavia anche delle ragioni altre, di matrice diversa da quella pubblicistica e non per questo meno meritevoli di essere prese in considerazione; ci si riferisce all'apporto che soggetti privati hanno dato e stanno dando al tema del contrasto alla disinformazione, ponendosi non già sull'asse verticale del rapporto tra autorità pubblica e individui, bensì sull'asse orizzontale del rapporto tra individui che vogliono partecipare al cosiddetto "mercato delle idee".

33. Cfr. O. Pollicino, P. Dunn, *op. ult. cit.*, p. 153; ma anche Ongarini, Cerutti, Loreggia, in questo volume.

34. Gli esempi sono ormai numerosi, come quello recente di Meta che su Instagram Reels ha esposto milioni di utenti a contenuti violenti e non sicuri a causa di un malfunzionamento dell'algoritmo di raccomandazione, dimostrando come i sistemi automatizzati possano fallire nell'identificazione corretta dei contenuti da filtrare.

Tra i diversi soggetti privati che nel corso dell'ultimo decennio hanno fatto la loro comparsa sulla scena, si richiameranno qui, a titolo puramente esemplificativo, due realtà, una internazionale e l'altra italiana, accomunate da una esperienza simile di *debunking*: NewsGuard e Pagella Politica.

Per quanto riguarda la prima realtà, come si legge nella pagina di presentazione del loro stesso sito, NewsGuard è un'organizzazione a scopo di lucro. Solo una parte dei loro ricavi deriva da enti governativi occidentali, che prendono in licenza i dati di NewsGuard sulle campagne di disinformazione sponsorizzate da Stati, mentre – sempre sulla base di quanto dichiarato – la maggior parte delle loro entrate proviene dalla concessione in licenza dell'accesso ai database ad aziende, organizzazioni di formazione, sistemi ospedalieri e altri enti del settore privato. A partire dai due Ceo, Steven Brill e Gordon Crovitz, la redazione è composta da tutti giornalisti professionisti (di diversa provenienza geografica), tanto che il loro *White Paper* sulla disinformazione si intitola esattamente «Combattere la disinformazione con il giornalismo, non con gli algoritmi». Oltre alla produzione di Report con focus su diversi settori, l'attività principale è quella di analizzare i siti che pubblicano notizie e informazioni, fornendo valutazioni e punteggi di affidabilità sulla base di nove criteri giornalistici.

Dunque, molto semplicemente: giornalisti professionisti che certificano l'affidabilità di siti di informazione online non per una "missione istituzionale", ma perché è un'esigenza dello stesso mercato delle informazioni quella di poter eventualmente contare sul parere di professionisti nell'individuare siti e fonti inaffidabili³⁵.

Esperienza simile, nata e sviluppata in Italia, è quella di Pagella Politica (Tfcf) Srl. Pagella Politica nasce nel 2012 come progetto editoriale che si oc-

35. Naturalmente, non sono mancate e non mancano le voci che mettono in discussione l'indipendenza di NewsGuard. Critiche che attengono proprio al tema dei finanziamenti sono venute, ad esempio, da *L'Indipendente* che in un articolo del 30 ottobre 2023 scriveva «NewsGuard è [...] una società privata che vede tra i propri fondatori il colosso francese della pubblicità e del marketing Publicis Groupe, una tra le cinque principali aziende del settore al mondo. Dal 2021, inoltre, la piattaforma ha siglato un contratto da 750 mila dollari con il Dipartimento della Difesa Usa, nell'ambito di un programma denominato "misinformation fingerprints" ("le impronte digitali della disinformazione"). In poche parole, a costituire il filtro per discernere le notizie da bollino verde da quelle da bollino rosso sono il Pentagono e la Casa Bianca». Nel luglio 2024 l'attività di NewsGuard è stata anche oggetto di un'interrogazione parlamentare, a firma dei deputati Ascari e Cherchi, nella quale si chiedeva al Governo quali iniziative intenda assumere nei confronti di «un'agenzia creata negli Stati Uniti da uno dei personaggi più potenti della Borsa di Wall Street, come espressione diretta dei gangli del potere Usa e con riferimenti diretti a Nsa, Cia e Council of Foreign Relation, [che] ha il potere di dare bollini di "verità" a chi fa informazione in Italia».

cupa di *fact-checking* e analisi dell'attualità politica; nel 2020 dalla stessa società Tfcf Srl ha lanciato anche il progetto di *debunking* Facta.news, che si occupa di disinformazione non politica.

Come nel caso di NewsGuard, anche per questa realtà la prima domanda da porsi è "chi paga?". A questo proposito, sul loro sito si legge che «non ricev[ono] denaro da partiti politici o da entità affiliate a partiti politici, né ricev[ono] fondi pubblici dalle autorità italiane» e che i loro progetti sono sostenuti «principalmente attraverso (a) attività di content providing, cioè vendendo contenuti e/o servizi a terzi; (b) partecipazione a bandi e progetti europei e internazionali».

Quest'ultimo punto sembra particolarmente interessante: se si va, infatti, ad analizzare quanto dichiarato, sembra evidente che il sostentamento per questa attività controllata e certificata di *debunking* viene in questo caso fatto attraverso finanziamenti europei e internazionali e con finanziamenti dalle stesse piattaforme³⁶. Siamo, dunque, al cospetto di un soggetto privato che questa volta si avvale di risorse o messe a disposizione da istituzioni pubbliche per il contrasto alla disinformazione o da altri soggetti privati. Una sinergia che, di nuovo, è meritevole di attenzione e che segnala una qualche forma di reazione e di risposta endogena al problema dell'*information disorder*.

3.2. Una medicina poco efficace

Per quanto detto sin qui, si sarebbe portati a dire che il *fact-checking* può davvero essere – se professionalmente svolto e finanziato in modo trasparente – un'arma molto potente. Questo è senz'altro vero e tuttavia, anche in questo caso, accanto agli aspetti positivi esistono anche dei grossi limiti dei quali occorre tener conto. Ci si riferisce, in particolare, agli esiti di alcuni studi, cominciati ormai quasi un decennio fa, che dimostrerebbero la scarsa efficacia del *debunking* soprattutto nei confronti degli utenti molto polarizzati.

Ad esempio, in una ricerca pubblicata nel 2017 sulla rivista scientifica Plos One, un gruppo di ricercatori ha analizzato i post, i "like" e i commenti pubbli-

36. L'elenco riportato nel sito dice: «progetti europei nell'ambito del fact-checking e della disinformazione, come l'European Digital Media Observatory (Edmo), l'Italian Digital Media Observatory (Idmo), EE24 Check e FactCrisis; progetti europei nell'ambito della media & information literacy (mil), come Hatedemics; progetti finanziati attraverso l'International Fact-Checking Network (Ifcn); Meta (Facebook e Thread), all'interno del Third-Party Fact-checking Program e della Health Fellowship: entrambi i programmi sono operati dalla redazione di Facta.news e non da Pagella Politica; Meta (WhatsApp), all'interno di un programma che sostiene il numero per le segnalazioni di Facta.news; TikTok, all'interno di un programma per il contrasto alla disinformazione sulla piattaforma operato da Facta.news».

cati su 83 pagine Facebook di carattere scientifico, 330 pagine “complotte” e 66 pagine dedicate al *debunking* (con oltre 50mila post), fornendo dunque un’analisi che ha coinvolto il comportamento social di oltre 54 milioni di utenti seguiti per 4 anni³⁷.

Il risultato di questa analisi è riassumibile nella emersione di due fenomeni oggi molto familiari a chi si occupa di comunicazione in ambiente digitale: le *echo chambers* e i *confirmation bias*.

Rispetto alle prime, è noto che gli stessi sistemi di raccomandazione delle piattaforme social tendono a spingerci in “bolle” entro le quali incontriamo solo chi la pensa come noi. E questo, molto semplicemente, perché i social network non sono progettati per essere strumenti di informazione, ma sono macchine per l’intrattenimento e premiano ciò che cattura l’attenzione, che emoziona, che divide, perché è questo che genera engagement. Le *echo chambers* e le *filter bubbles* dividono, non creano comunità di discussione, ma al contrario favoriscono il fenomeno del cosiddetto “incastellamento”³⁸, aumentando la polarizzazione.

Anche sui *confirmation bias* oggi sappiamo molto: si tratta di quei pregiudizi cognitivi che ci conducono a privilegiare, in modo inconsapevole, tutte le informazioni che confermano le nostre convinzioni e a scartare, pregiudizialmente, quelle contrarie³⁹. È un meccanismo psicologico innato, che ovviamente presiste ai *social network*, ma che le piattaforme social tendono ad amplificare⁴⁰.

Alla luce di tutti questi fenomeni accertati, il *debunking* online sarebbe fallimentare: perché nei partecipanti più coinvolti ideologicamente la correzione

37. F. Zollo, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli, L. Shekhtman *et al.*, *Debunking in a world of tribes*, in *PLoS ONE*, 2017, vol. 12, n. 7.

38. Scrive B.C. Han, *op. ult. cit.*, p. 77 «Mentre pensiamo di essere liberi, oggi siamo intrappolati in una caverna digitale»; sul problema dell’isolamento in ambiente digitale ho svolto qualche considerazione in F. Fabrizzi, *La formazione del consenso, la rappresentanza, il Parlamento nell’era della digitalizzazione*, in *Italian papers on federalism*, n. 1/2025, p. 124 ss.

39. «Siamo, in altri termini, alla continua ricerca di conferme e non di rettifiche, anche quando le seconde sarebbero più che plausibili», così F. Vella, *Diritto ed economia comportamentale*, il Mulino, 2023, p. 30; sulla stessa linea G. Gometz, *Piattaforme online e crisi della democrazia: unde malum?*, in *Parole chiave*, n. 1/2023;

40. In un’intervista pubblicata il 19 aprile 2024 (*Le piattaforme digitali influenzano le interazioni sociali? Intervista a Walter Quattrociocchi*, disponibile su: www.pandoravivista.it/articoli/le-piattaforme-digitali-influenzano-le-interazioni-sociali-intervista-a-walter-quattrociocchi/), Quattrociocchi ha affermato che «all’interno di queste *echo-chamber* la tendenza alla tossicità aumenta e l’*hate speech* è più probabile, perché si innesca quella che viene chiamata la legge della polarizzazione dei gruppi, come definita da Cass Sunstein. Questa ipotesi, che abbiamo verificato empiricamente, sostiene che se una persona con visioni estremiste si muove in un contesto abbastanza tranquillo avrà interazioni moderate, ma se invece sempre in un ambiente tranquillo abbiamo due persone con le stesse visioni estremiste, queste si daranno manforte e saranno estremamente forti e violente nell’esposizione del loro credo condiviso».

nel migliore dei casi è inutile, ma nel peggiore rischia di diventare persino dannosa, perché finirebbe per rinforzare il credo in una narrazione sbagliata⁴¹. Poiché è la sfera emozionale della disinformazione che viene in gioco⁴², la risposta in termini di discorso razionale non funzionerebbe: «le persone non stanno scegliendo informazioni false rispetto a quelle vere; stanno abitando sistemi di realtà completamente differenti»⁴³ e tentare di smascherare con i “fatti” è come «cercare di confutare una religione con la scienza, la fede con il sapere»⁴⁴.

E occorre fare estremamente attenzione, perché «coloro che sono indifferenti ai fatti e alla realtà rappresentano un pericolo maggiore per la verità rispetto ai bugiardi»⁴⁵.

4. Il *debunking* e l'educazione, vaccini contro la disinformazione

Se, dunque, il *debunking* rischia di arrivare tardi e di non raggiungere tutti – e in particolare coloro che sono maggiormente predisposti alla disinformazione – più efficace potrebbe rivelarsi un'altra strategia che fa leva su meccanismi psicologici e cognitivi, quella del cosiddetto *prebunking*⁴⁶.

41. Accanto ai *bias* di conferma, peraltro, se ne devono considerare anche numerosi altri, alcuni dei quali propri e tipici dell'ambiente digitale: l'*overconfidence bias*, per cui si tende a sopravvalutare la propria capacità valutativa degli eventi, la *complacency* (compiacenza), che gli utenti tendono ad avere sovrastimando l'affidabilità dell'output algoritmo rispetto all'informazione o output umano (cfr. M.T. Dzindolet, S.A. Peterson, R.A. Pomranky, L.G. Pierce, H.P. Beck, *The role of trust in automation reliance*, in *International Journal of Human-Computer Studies*, 2003, pp. 697-718; R. Parasuraman, V. Riley, *Humans and automation: Use, misuse, disuse, abuse*, in *Human Factors*, 1997, pp. 230-253; J.M. Logg, J.A. Minson, D.A. Moore, *Algorithm appreciation: People prefer algorithmic to human judgment*, in *Organizational Behavior and Human Decision Processes*, 2019, pp. 90-103); l'*automation bias* che consiste nella «tendenza a sovrastimare la qualità del prodotto dell'elaborazione algoritmica e l'appropriatezza della decisione alle esigenze dell'utente-utilizzatore. L'effetto è una tensione verso l'accoglimento automatico dell'output algoritmico tale per cui si finisce per farlo inconsapevolmente proprio, spesso senza averne colto il senso né valutato, riflettendoci, l'opportunità». C. Arruzzoli, *L'influenza dell'Automation Bias sulle decisioni amministrative*, 3, Edizioni scientifiche italiane – Collana del Dottorato di ricerca in diritto pubblico, comparato e internazionale (in corso di pubblicazione).

42. Sulla difficoltà di abbandonare preconcinzioni errate (c.d. *backfire effect*), B. Nyhan, J. Reifler, *The Backfire Effect*, in *Political Behavior*, 2010.

43. J. Xun, *Ipnocrazia. Trump, Musk e la nuova architettura della realtà*, Edizioni Tlon, 2025, p. 73.

44. *Ibidem*.

45. B.C. Han, *op. ult. cit.*, p. 63.

46. Uno studio risalente, ma interessante: W.J. McGuire, *Inducing resistance to persuasion. Some contemporary approaches*, 1961.

Si tratta di un metodo che gli studiosi definiscono di “vaccinazione attitudinale” in quanto agisce in via preventiva, esponendo le persone a esempi di disinformazione o tecniche di disinformazione, così da aiutarli e abituarli a riconoscerli e a respingerli in futuro. Diversi esperimenti basati sul principio della inoculazione psicologica sono stati compiuti e dimostrerebbero un effetto positivo, registrato, peraltro, su persone di età diverse, con diversi orientamenti politici, differenti livelli di istruzione e differenti tratti di personalità.

Il *prebunking* può essere attivo o passivo a seconda della tecnica psicologica su cui si vuole fare leva.

Sono esempi di *prebunking* attivo, che hanno dato buoni risultati sulla fascia di età dei minori, i videogiochi nei quali i giocatori assumo il ruolo di produttori di *fake news* e imparano a padroneggiare le tecniche comunemente utilizzate nella produzione di disinformazione; uno di quelli più citati è “*Bad Vaxx*”, gioco gratuito per *browser* che simula l’ambiente dei social media, permettendo ai giocatori di sperimentare in prima persona le tattiche usate per diffondere notizie false sui vaccini, sviluppato da un gruppo di ricercatori dell’Università di Cambridge nell’ambito del progetto *Inoculation Science*⁴⁷.

L’attivazione dell’immunità cognitiva può avvenire anche tramite forme di *prebunking* passivo, consistente nel somministrare video per informare circa le più comuni tecniche di manipolazione utilizzate per far passare notizie false, quali ad esempio la manipolazione emotiva, le false dicotomie, lo *scape-goating*. Un esperimento condotto sempre dai ricercatori dell’Università di Cambridge insieme alla piattaforma YouTube, prevedeva, ad esempio, l’inserimento di brevi video di novanta secondi, simili ad annunci del servizio pubblico, negli slot pubblicitari della piattaforma; il risultato ottenuto dimostrerebbe che i 30mila partecipanti all’esperimento hanno registrato una capacità di identificare informazioni false del 5% più elevata.

Quello che emerge, insomma, dagli studi delle neuroscienze cognitive, soprattutto con riferimento ai bambini, non solo è molto interessante, ma va seriamente tenuto in considerazione anche sotto il profilo giuridico: i dati dicono, infatti, che “difendere” i minori dall’esposizione a informazioni false, provando a filtrarle a monte (con un approccio, dunque, di controllo e protezione) potrebbe non essere la strategia corretta, suggerendo invece di fare esattamente il contrario, perché solo in questo modo la capacità di vaglio critico del bambino migliora⁴⁸.

47. Nel sito inoculation.science/ è consultabile tutta la documentazione relativa a questo progetto di ricerca, inclusa una molto ricca *Practical Guide to Prebunking Misinformation*.

48. Cfr. E. Orticio, M. Meyer, C. Kidd, *Exposure to detectable inaccuracies makes children more diligent fact-checkers of novel claims*, in *Nature Human Behaviour*, 2024, n. 8, pp. 2322-2329.

Non si tratterebbe dunque (solo) di proteggere, ma di consolidare le competenze del soggetto in modo che si protegga da sé⁴⁹.

Come per i due rimedi descritti in precedenza (l'apparato giuridico-sanzionatorio e il *debunking*), anche con riferimento al *prebunking*, accanto alle potenzialità virtuose del metodo, si è costretti a evidenziare anche i forti dubbi di tipo etico-giuridico che esso solleva.

Si tratta, infatti, di un sistema in cui in maniera più o meno “subdola” si vanno a toccare leve nel processo mentale con tecniche di influenza psicologica che inevitabilmente sollevano una serie di quesiti, stante la potenza insita nello strumento: chi decide cosa rientra nella categoria di “manipolazione” o “distorsione della realtà”? Chi sarà abilitato a introdurre e utilizzare queste tecniche, i privati o le autorità pubbliche? Come si assicura un controllo democratico su questi processi?

Il problema in misura simile si è posto e si pone anche nel caso del cosiddetto *nudging*, espressione con cui si designano politiche pubbliche che, facendo leva su meccanismi psicologico-cognitivi, mirano a orientare le scelte dei cittadini senza imporre obblighi né divieti. La *nudge theory* – la “spinta gentile” o “pungolo” elaborata da Thaler e Sunstein⁵⁰ – agisce sul piano prerazionale, modificando il contesto decisionale in modo tale che l'opzione desiderata dall'autorità pubblica risulti la più agevole o comunque intuitiva per il cittadino, il quale pur restando libero di scegliere (nel senso di non obbligato a) è indotto tuttavia a compiere esattamente la scelta verso cui lo spinge il *nudger*⁵¹.

Anche in questo caso, dunque, ci si muove su di un piano che è pre-giuridico, ma che ha evidentissime ricadute di tipo giuridico, trattandosi di una forma di esercizio del potere che, seppur “gentile” e privo di carattere coercitivo, influenza le scelte individuali e chiama quindi in causa il principio di autodeterminazione.

D'altra parte, a ben vedere, tutte le volte in cui si pongono in essere attività con finalità educative – per quanto benemerite ed encomiabili – viene in rilievo il tema di come svolgere dette attività senza violare la libertà delle persone, a partire proprio dalla libertà di pensiero⁵².

49. Sposa in pieno questa visione, V. Gallese, S. Moriggi, P.C. Rivoltella, *Oltre la tecnofobia. Il digitale dalle neuroscienze all'educazione*, Raffaello Cortina Editore, 2025.

50. Sull'opera di Sunstein e Thaler, con interessanti valutazioni sul tema dell'autodeterminazione, vedi C. Arruzzoli, *Recensione a C. Sunstein – R.H. Thaler, Nudge. La spinta gentile. L'edizione definitiva*, Milano, Feltrinelli, 2022, pp. 312, in *Nomos. Le attualità nel diritto*, n. 1/2023.

51. Per una ricostruzione del tema si veda C. Arruzzoli, *Intelligenza artificiale e nudging*, in F. Fabrizzi, L. Durst, *Controllo e predittività*, Editoriale Scientifica, 2024, p. 75 ss.

52. G. Sobrino, *L'insegnamento della Costituzione nella scuola oggi, strumento istituzionale per la promozione di una cittadinanza “piena” e consapevole*, in *federalismi.it*, n. 18, 2022, p. 261.

5. L'autodeterminazione come obiettivo di educazione costituzionale

Accanto all'apparato giuridico-sanzionatorio e insieme ai rimedi che provengono dallo stesso sviluppo del "mercato delle idee", occorre, dunque, una seria ed efficace educazione digitale per aggredire il problema della disinformazione.

Si tratta di un compito istituzionale assai delicato e niente affatto semplice da realizzare, che deve essere improntato alla più stretta applicazione dell'imperativo della salvaguardia del pluralismo, quale fondamento del costituzionalismo contemporaneo. La sfida, propria dello stesso concetto di educazione costituzionale, di «"educare" i cittadini alla democrazia e al rispetto dei principi costituzionali senza al contempo "tradire" la Costituzione stessa e violare un suo principio fondamentale, come appunto la libertà»⁵³, si applica tanto più nel caso dell'educazione digitale.

Se da un lato non convince l'atteggiamento paternalistico dell'istituzione pubblica che dice ai cittadini "Io so cosa è bene per te e ti preservo da contenuti pericolosi"⁵⁴; allo stesso modo non convince l'atteggiamento catechetico e assertivo, volto a presentare alcuni valori e posizioni come "verità evangeliche", intrinsecamente e necessariamente "buone"⁵⁵, alle quali prestare fede e obbedienza in una sorta di "catechismo costituzionale" che indichi di volta in volta i valori minacciati e da difendere⁵⁶.

Non esiste, in altri termini un "obbligo generale alla verità", ma deve esistere, invece, un diritto aletico declinato in termini costituzionalistici come diritto a non essere disinformati e diritto a una maggiore educazione civica e mediatica. Non dunque una verità ultimativa, ma «una cultura della verità che implica dialogo, dubbio, disaccordo, revisione dei giudizi e discussione dei fatti»⁵⁷.

La traduzione giuridica di questo approccio non può che ritrovarsi nel riconoscimento di «situazioni giuridiche attive e passive, libertà negative come libertà positive, formule organizzative che lo Stato dovrebbe garantire o pro-

53. Ivi, p. 262.

54. Sul paternalismo libertario cfr. C.R. Sunstein, *Why Nudge? The Politics of Libertarian Paternalism*, Yale University Press, 2014; C.R. Sunstein, *The Ethics of Influence. Government in the Age of Behavioral Science*, Cambridge University Press, 2016; C.R. Sunstein, R.H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, in *University of Chicago Law Review*, 2003, vol. 70, n. 4, pp. 1159-1202.

55. Cfr. G. Sobrino, *op. ult. cit.*, p. 263.

56. Cfr. N. Zanon, *Su alcune questioni e tendenze attuali intorno alla motivazione delle decisioni della Corte costituzionale, tra forma e sostanza*, in M. Losana, V. Marcenò (a cura di), *Come decide la Corte dinanzi a questioni "tecniche". Incontri sulla giurisprudenza costituzionale*, Università degli Studi di Torino, 2020, p. 23 ss.; citato da G. Sobrino, *op. ult. cit.*, p. 263.

57. A. Facchi, *op. ult. cit.*, p. 64.

muovere sia per promuovere un *bisogno* di verità che, più in generale, far circolare la moneta-verità»⁵⁸.

Peraltro, anche la circolazione della moneta-verità si colloca nella stessa linea di problematico rapporto tra libertà e autorità, giacché la stessa esposizione a informazioni, argomenti e posizioni che le persone non avrebbero magari scelto di propria iniziativa potrebbe essere interpretato come un indebito intervento eterodiretto.

In verità, il lavoro di educazione – tanto più oggi – è più complesso rispetto alla “semplice” proposta di più punti di vista e riguarda l’educazione all’auto-determinazione e alla consapevolezza.

L’idea del pluralismo e della esposizione a quante più opinioni e posizioni possibili deve infatti condurre non già alla “conoscenza”, ma alla *curiosità* riguardo alla verità.

La differenza è sottile, ma centrale.

Edgard Morin, citando Michel de Montaigne, ha scritto che «È meglio una testa ben fatta che una testa piena»⁵⁹, volendo sottolineare come l’intento dell’educatore non deve essere quello di riempire di nozioni, ma di abituare al ragionamento critico. Tornano alla mente le parole di Sartori, laddove ricorda che «quando asseriamo che la democrazia si fonda sulla sovranità popolare ne indichiamo soltanto, o comunque soprattutto, il principio di legittimazione». E poi ammonisce: «Resta che un sovrano vuoto, un sovrano che non sa e non dice, è un sovrano da nulla, un re di coppe»⁶⁰.

Oggi la domanda ulteriore è, brutalmente, se sia meglio una testa vuota o una testa riempita da qualcun altro, come può accadere in caso di manipolazioni da disinformazione.

Come ci ha ricordato la Corte costituzionale nella già citata sent. 44/2025, il problema non è più l’accesso alle informazioni (o anche alle opinioni), atteso il «sovraccarico mediatico» cui il cittadino è sottoposto. Il problema oggi è piuttosto come fare in modo che quel cittadino mantenga viva tanto la capacità critica quanto la volontà di rimanere informato, proprio perché per continuare a essere sovrano il popolo deve possedere ed esprimere un “contenuto suo proprio”.

Maurizio Ferrara, citando alcuni studi di psicologi politici dell’Università di Yale, ha messo in luce come la risorsa capacitante per la ricerca della verità e il pensiero critico sia la “curiosità scientifica”, ossia l’interesse e la disponibilità

58. M. Plutino, *op. ult. cit.*, p. 193.

59. E. Morin, *La testa ben fatta*, Raffaello Cortina Editore, 2000.

60. G. Sartori, *Opinione pubblica*, in *Enciclopedia Treccani*, disponibile su: [www.treccani.it/enciclopedia/opinione-pubblica_\(Enciclopedia-del-Novecento\)/](http://www.treccani.it/enciclopedia/opinione-pubblica_(Enciclopedia-del-Novecento)/).

a informarsi per gusto personale su un ampio spettro di temi. Essa è fondamentale perché «chi è curioso ha meno probabilità di farsi ingannare, di indulgere in ragionamenti “motivati”, quelli che selezionano le informazioni che rafforzano i propri assunti e convinzioni di partenza, scartando o screditando quelle scomode»⁶¹.

Mantenere attiva la curiosità scientifica e alimentare le competenze di pensiero critico per formarsi un’opinione consapevole e autonoma, potrebbe essere oggi, paradossalmente, più difficile che in passato. Gli studi più recenti sull’intelligenza artificiale generativa e sugli Llm accendono su questo un faro davvero illuminante e preoccupante.

Come noto, i *Large Language Models* come ChatGpt o Gemini o simili, fondano la loro potenza sulla possibilità di inferenze tra un numero di dati di addestramento senza precedenti; attraverso tecniche avanzate di Deep Learning, questi modelli analizzano immense quantità di testo per identificare *pattern* statistici e relazioni semantiche. Il risultato è che ricombinano bene testo e generano sequenze plausibili in base a correlazioni imparate, non avendone tuttavia alcuna “contezza”, né avendo consapevolezza degli errori. La macchina restituisce, cioè, risultati plausibili, e lo fa con tono autorevole e credibile.

Giustamente fino a oggi, ragionando di Deep Learning, si è posto l’accento sul problema degli errori, dei *bias*, delle discriminazioni algoritmiche, delle allucinazioni, della qualità dei dati. Tutte questioni assolutamente fondamentali che, tuttavia, possono essere affrontate facendo un attento lavoro in fase di progettazione dei sistemi di AI e di controllo sugli stessi (quella strategia *by design* e *by default* che è esattamente la strategia adottata dall’AI Act).

C’è però un pericolo decisamente maggiore che si annida dietro l’utilizzo di questi sistemi e che è stato descritto come un rischio cognitivo “di seconda generazione”, quella che Walter Quattrociocchi nei suoi studi chiama “epistemica”: «Quando scambiamo coerenza per conoscenza – ha scritto Quattrociocchi – stiamo scambiando una simulazione per un giudizio fondato»⁶².

Il pericolo, allora, risiede nella tentazione di delegare alla macchina il processo cognitivo, di essere indotti a scambiare il plausibile per vero e di smettere di imparare – e di essere curiosi – «perché si è convinti di aver già imparato abbastanza e tanto c’è l’AI a colmare le lacune»⁶³.

Il potenziale di diffusione di disinformazione, ma soprattutto di manipola-

61. M. Ferrara, *L’anima atletica della democrazia liberale*, in *Biblioteca della libertà*, n. 1/2017, p. 76.

62. W. Quattrociocchi, *Epistemica*, in *La nostra scuola*, Disponibile su: nostrascuola.blog/2025/10/29/epistemica/.

63. Così, sempre W. Quattrociocchi in un post sul suo profilo LinkedIn.

zione e di esonero cognitivo insito in un meccanismo informativo come quello al quale stiamo assistendo – in cui Google, prima dei risultati del motore di ricerca, offre oggi la risposta di Gemini – è massimo.

L'educazione digitale di cui c'è bisogno è allora un'educazione che insegni a non delegare il processo cognitivo, che scongiuri l'impoverimento intellettuale e che faccia comprendere l'importanza di non addivenire a un esonero della mente, a un'atrofia cognitiva, perché, – è bene tornare a ribadirlo – se è vero che una testa *buona* è meglio di una testa *piena*, una testa *vuota* rischia di essere riempita da chiunque, con evidenti conseguenze che possono divenire nefaste.

L'imperativo costituzionale è allora quello di avere "teste consapevolmente piene", perché come ha scritto Enrico Grosso:

l'isocrazia presuppone l'uguaglianza del "valore" di ogni singolo voto [...], ma ciò a sua volta presuppone l'uguaglianza del "valore" di ogni singola opinione che a quel voto è sottesa. E dunque *presuppone un cittadino non solo educato e informato, ma consapevole e desideroso di essere e restare educato e informato*, e di rendersi attivo, di formarsi – appunto – una "opinione" degna di essere paragonata e resa del tutto equivalente, sul piano del "valore", alle altre⁶⁴ [enfasi nostra].

Quello dell'educazione all'autodeterminazione è, dunque, un imperativo costituzionale anche perché questo è un meccanismo che rischia seriamente di amplificare le disuguaglianze cognitive, dal momento che chi ha capitale culturale alto riconosce incongruenze, verifica, rielabora; chi, invece, non ha strumenti culturali adeguati, prende il plausibile per vero, e può essere indotto a credere acriticamente a qualsiasi cosa. Per i primi – coloro che avranno un solido apparato critico – l'Intelligenza artificiale potrà essere realmente "strumento"; i secondi sono destinati a essere schiavi, del mezzo ed eventualmente anche di chi quel mezzo lo sa usare.

Da un lato, dunque, occorre far crescere e coltivare la volontà di rimanere informato e dell'altro occorre lavorare per educare al pensiero critico e alla consapevolezza. D'altra parte,

perché il voto sia davvero lo strumento di una "democrazia emancipante" [...] non è sufficiente garantirlo a tutti. Esso deve essere accompagnato da una *continua azione di educazione e di "costruzione della consapevolezza"*, affinché l'edificio democratico non poggia su fondamenta troppo fragili e destinate alla lunga a essere spazzate via [enfasi nostra]⁶⁵.

64. E. Grosso, *op. ult. cit.*, p. 21.

65. *Ibidem*.

LA QUALIFICABILITÀ DELLA DISINFORMAZIONE QUALE PROBLEMA DI CYBERSICUREZZA

*Giovanni Barozzi Reggiani**

SOMMARIO: 1. Brevi considerazioni introduttive – 2. La “funzione pubblica” di cybersicurezza – 3. Caratteristiche della disinformazione e strumenti (giuridici) per contrastarla – 4. Considerazioni conclusive. La disinformazione quale problema di cybersicurezza: una realtà già attuale o una semplice prospettiva (*de iure condendo*)?

1. Brevi considerazioni introduttive

Il presente contributo intende indagare se esista un nesso, sul piano giuridico, tra disinformazione e cybersicurezza.

Invero, fornire tale risposta – che, come emergerà, è positiva – non ha richiesto a chi scrive un particolare sforzo: la connessione tra i descritti elementi è infatti evidente e sottolineata da più parti, e anzi i punti di contatto tra cybersicurezza e disinformazione sono plurimi.

Ciò che, piuttosto, impone allo studioso un impegno particolare è comprendere la natura dei citati punti di contatto, e quali effetti giuridici dai medesimi scaturiscano.

In effetti, tanto la disinformazione quanto i cyberattacchi possono danneggiare o mettere in pericolo beni giuridici tutelati dalle Carte costituzionali degli ordinamenti moderni, ciò che costituisce elemento di comunanza e connessione tra i due fenomeni.

Al contempo, è nella natura dei beni minacciati, e nelle forme che le minacce possono concretamente assumere, che trova fondamento l'esigenza di differenziare, sul piano giuridico, il modo in cui disinformazione e attacchi informatici vengono considerati e disciplinati giuridicamente, anche in ragione

* Ricercatore di Diritto costituzionale e pubblico nell'Università di Sassari.

dei diritti, dei valori e delle libertà che, a seconda dei casi, la messa a punto di discipline specifiche viene a intercettare (e limitare).

Pur a fronte della sussistenza di quadri giuridici distinti, non è tuttavia affatto da escludersi, come si proverà a dimostrare, che già all'attualità, e in base a una certa lettura delle disposizioni vigenti, possa sostenersi, con qualche valida argomentazione giuridica, che quantomeno alcune delle declinazioni del fenomeno della disinformazione possano (debbano) essere considerate attività rispetto alle quali si manifestano esigenze di *cybersecurity*, e che dunque tra i due ambiti sussistano zone di intersezione.

2. La “funzione pubblica” di cybersicurezza

È stato recentemente scritto che quella di cybersicurezza costituisce una “funzione pubblica”¹.

Si tratta di affermazione che si ritiene di condividere, specie se con la medesima si intende indicare un'attività (o un insieme di attività) che viene presa in carico da apparati pubblici di un determinato ordinamento al fine di tutelare uno o più beni giuridici che quest'ultimo considera per sé rilevanti.

In effetti, come si accennava, è agevole osservare che tutte le attività riconducibili al concetto di cybersicurezza sono funzionali a garantire la protezione di beni, diritti e valori di sicuro rilievo ordinamentale, e che le medesime per tale ragione vengono affidate (anche) ad apparati pubblici², con definizione per via legislativa di regole e assetti organizzativi, nell'ambito dei quali si prevedono specifici obblighi per soggetti tanto pubblici che privati, nonché l'attribuzione a Enti e Autorità del potere di adottare atti e provvedimenti suscettibili

1. Così si sono recentemente espressi, tra gli altri, M. Macchia, G. Sferrazzo, *Sicurezza e rischio tecnologico. La funzione di cybersecurity*, in *Dir. Amm.*, n. 1/2025, pp. 109 ss.; E. Longo, *Il diritto costituzionale e la cybersicurezza. Analisi di un volto nuovo del potere*, in *Rassegna Parlamentare*, n. 2/2024, pp. 311 ss.; R. Ursi, *La sicurezza cibernetica come funzione pubblica*, in R. Ursi (a cura di), *La sicurezza nel cyberspazio*, FrancoAngeli, 2023, pp. 7 ss.

2. Il ricorso alla congiunzione “anche” si rende necessario in quanto nel settore della *cybersecurity* il coinvolgimento del settore privato costituisce elemento imprescindibile, per ragioni legate non tanto alla “democratizzazione” della funzione quanto piuttosto alla garanzia dell'efficacia della stessa. Si parla, in questo senso, di una «regolazione ibrida pubblico-privata», che appare da prediligersi in quanto, per il tramite della stessa, i destinatari delle regole «diventano parte integrante del processo decisionale iniziale contribuendo con risorse e standard che sono considerati l'elemento principale da prendere in considerazione nella valutazione del rischio» (così E. Longo, *La disciplina del “rischio digitale”*, in Aa.Vv., *La regolazione europea della società digitale*, Giappichelli, 2025, p. 79), ciò che produce ricadute positive sul grado di effettività delle regole stesse.

di determinare unilateralmente incisioni sulle sfere giuridico/economiche di singoli individui o enti³.

Peraltro, se osservato tramite il prisma dei beni giuridici da tutelare, l'ambito di operatività della "funzione di cybersicurezza" appare piuttosto esteso.

Nel cyberspazio (luogo al contempo fisico e virtuale⁴) vengono infatti svolte quotidianamente numerose attività – sul piano economico, delle relazioni interpersonali, dell'istruzione e della ricerca, dell'erogazione (e della fruizione) di servizi e di molto altro⁵ – che costituiscono il cuore pulsante della vita delle società moderne.

Già solo tale constatazione fa apparire evidente quanto la protezione del cyberspazio, di parti di esso o delle attività che nel medesimo vengono svolte, rispetto a possibili aggressioni, costituisca un obiettivo di primario rilievo, per gli ordinamenti, oltre a tema che «interessa tutti» e che non deve ritenersi confinato «all'intelligence o alla messa in sicurezza di alcuni dispositivi che possono essere infettati da "malattie" informatiche»⁶.

Tali affermazioni appaiono ulteriormente rafforzate dalla considerazione secondo la quale il regolare svolgimento della "vita" del cyberspazio presuppone, e determina, l'uso, la raccolta e la diffusione di una enorme quantità di dati (di utenti, imprese, Amministrazioni Pubbliche...) che se non adeguatamente protetti possono venire sottratti o danneggiati (per i più diversi fini), con possibili pregiudizi per le sfere economico-giuridiche dei singoli, danneggiamento di *software* o reti e conseguente impossibilità di svolgimento di determinate attività o di erogazione di servizi, e finanche messa in pericolo di *asset* strategici dello Stato (si pensi alla sottrazione di dati coperti da segreto o al danneggiamento di *software* o reti che governano e presidiano attività di primario rilievo, comprese quelle afferenti al settore della difesa).

3. Si tratta, dunque, di funzioni amministrative, secondo l'accezione comunemente condivisa (in tema, per tutti, vedi il recente saggio di A. Zito, *La funzione amministrativa nella riflessione giuridica: una nozione meramente descrittiva o connotativa?*, in *Rivista CERIDAP*, n. 2/2024, pp. 228 ss.).

4. Secondo la relativa voce dell'Enciclopedia Treccani (*online*), il cyberspazio costituisce «lo spazio virtuale nel quale utenti (e programmi) connessi fra loro attraverso una rete telematica (per es., *internet*) possono muoversi e interagire per gli scopi più diversi». Per alcune interessanti considerazioni sulle caratteristiche del cyberspazio v. K. Fedorov *et al.*, *Cyberspace: Key Properties and Traits*, in *Journal of Physics: Conference Series*, 2021(2096), pp. 1 ss.

5. In ragione di ciò, come è stato osservato, il cyberspazio costituisce «uno spazio integrato, in quanto costituito dall'interazione tra le diverse tecnologie dell'informazione, e pervasivo, data la rilevanza di esse in pressoché tutti gli ambiti della vita sociale» (così M. Mirti, *Struttura e caratteri del Cyberspace*, in *OpinioJuris*, n. 12/2020, p. 6).

6. Per entrambe le citazioni v. E. Longo, *La disciplina della cybersicurezza nell'Unione europea e in Italia*, in Aa.Vv., *La regolazione europea della società digitale*, cit., p. 203.

Da tutto ciò consegue che la presenza e la circolazione massiva in Rete dei dati costituisce un sicuro elemento di vulnerabilità di individui, attività e strutture, cui gli ordinamenti non possono restare indifferenti⁷.

La rilevanza di tale vulnerabilità (e delle altre connesse alla vita del cyberspazio), può naturalmente venire a configurare un problema di sicurezza, volendo con ciò riferirci (anche in questo caso) a un bene giuridico che assume rilevanza costituzionale e della cui protezione gli apparati pubblici devono in qualche modo farsi carico⁸.

Si tratta, peraltro, di un bene non unitario e, anzi, composito; o, diversamente detto, di un insieme, cui è possibile ricondurre diversi elementi costituenti beni giuridici tra loro distinti⁹. Tra questi, quelli che vengono in rilievo a riguardo della cybersicurezza concernono da un lato la salvaguardia di singoli beni, libertà o diritti (come il diritto alla *privacy* e/o alla tutela dei dati personali, ovvero la protezione dei singoli da comportamenti dannosi e/o criminosi, come truffe o furti, o ancora la garanzia della fruibilità di determinati servizi o prestazioni¹⁰), dall'altro esigenze di profilo più "ampio",

7. Sul rapporto tra "presenza" dei dati e vulnerabilità del sistema (cyberspazio) v. S. Calzolaio, *Vulnerabilità della società digitale e ordinamento costituzionale dei dati*, in *Rivista italiana di informatica e diritto*, n. 2/2023, pp. 14 ss.

8. Come osservato recentemente da G. Stegher, voce *Sicurezza* nel volume curato da T. Groppi, C. Bassu, T.E. Frosini, R. Tarchi, *Lessico in-costituzionale*, Editoriale Scientifica, 2025, p. 538, «sin dalle prime elaborazioni dottrinarie (Hobbes) e dalle prime codificazioni (Dichiarazione dei diritti dell'uomo e del cittadino del 1789), la sicurezza ha rappresentato una delle funzioni principali dello Stato, il quale svolge il ruolo di custode rispetto a eventuali ingerenze esterne e a problematiche interne. La stessa, impiegata per mantenere l'integrità degli elementi distintivi dello Stato (popolo, territorio e sovranità), presuppone l'esistenza ed è posta a garanzia di altri diritti. A lungo, perciò, è stata intesa come uno dei compiti dell'azione statale e per questo sono stati predisposti i necessari strumenti sia a tutela dell'ordine interno, sia a difesa nei rapporti esterni».

9. Nel presente scritto si esamina la funzione di cybersicurezza essenzialmente quale attività di "difesa" e "protezione". Appare tuttavia opportuno precisare che, come è stato evidenziato, la cybersicurezza per come è oggi concepita ha un ambito di applicazione più ampio, «sia perché si trova all'incrocio tra *security* e *safety*, sia perché essa si arricchisce di un diverso e più innovativo concetto rivolto a descrivere i metodi – messi in atto tanto da soggetti pubblici quanto privati – per garantire tanto il lavoro volto a rilevare, rintracciare e punire i responsabili delle condotte dolose quanto il ritorno alla funzionalità dei sistemi informatici una volta che essi siano stati compromessi da attacchi malevoli» (così E. Longo, *Il diritto costituzionale e la cybersicurezza*, cit., p. 318).

10. Alcune esigenze di sicurezza (e cybersicurezza) attengono dunque alla protezione di diritti (e libertà). Ciò, tuttavia, a giudizio di chi scrive non consente di individuare, rispetto alla cybersicurezza, un diritto soggettivo riconducibile a singoli individui (e ciò in linea con la lettura a suo tempo fornita da A. Pace, in riferimento alla "sicurezza", in particolare nel saggio *Libertà e sicurezza. Cinquant'anni dopo*, in *DeS*, n. 2/2013, pp. 177 ss.), configurando piuttosto un obiettivo e un fine delle politiche e delle azioni dei pubblici poteri (ritengo che

che possono attenere finanche alla stessa difesa dell'integrità territoriale dello Stato (comunità) e dunque a una delle prerogative essenziali e costitutive di quest'ultimo, venendo a collocarsi sul piano della sfera dei rapporti, anche basati sull'uso della forza, tra diversi ordinamenti sovrani, e quindi nella dimensione geopolitico-militare¹¹.

Tutto ciò emerge in modo piuttosto evidente da un'analisi della definizione di cybersicurezza posta dall'art. 1, comma 1, lett. a), del d.l. n. 82 del 2021¹² (e richiamata da quella – sempre di cybersicurezza – che si rinviene nell'art. 1, comma 2, lett. s, del d.lgs. n. 4 settembre 2024, n. 138, attuativo della Direttiva (UE) 2022/2555), che così qualifica «l'insieme delle attività [...] necessarie per proteggere dalle minacce informatiche reti, sistemi informativi, servizi informatici e comunicazioni elettroniche, assicurandone la disponibilità, la confidenzialità e l'integrità e garantendone la resilienza, anche ai fini della tutela della sicurezza nazionale e dell'interesse nazionale nello spazio cibernetico»; ancora più ampia appare poi la definizione di cybersicurezza posta dal Regola-

nel medesimo senso vada letta Corte Cost. n. 2 del 1956, la quale, dopo aver escluso che la sicurezza si riferisse solo all'incolumità fisica, ha sostenuto come risultasse «razionale e conforme allo spirito della Costituzione dare alla parola "sicurezza" il significato di situazione nella quale sia assicurato ai cittadini, per quanto è possibile, il pacifico esercizio di quei diritti di libertà che la Costituzione garantisce con tanta forza»: l'utilizzo dell'inciso "per quanto possibile" corrobora la tesi secondo la quale a riguardo della sicurezza (intesa nel senso descritto) gli individui non possano vantare situazioni giuridiche soggettive assimilabili ai diritti soggettivi perfetti). In senso analogo, in riferimento alla non configurabilità di un «diritto fondamentale alla cybersicurezza», si è espresso E. Longo, *Il diritto costituzionale e la cybersicurezza*, cit., nt. 42).

11. Scrive in questo senso E. Longo (*op. ult. cit.*, p. 314) che «l'insicurezza del cyberspazio non minaccia solo l'economia, ma addirittura il funzionamento delle nostre democrazie e i valori su cui si fonda la società [dato che] le politiche interne degli Stati che riguardano l'ordine pubblico, la sicurezza nazionale e la stessa protezione delle libertà devono fare i conti con le minacce che provengono o si consumano nel cyberspazio». In senso analogo si è espressa M. Pietrangelo, *La dimensione plurale della cybersicurezza: da potere invisibile a processo collaborativo*, in *Rivista Italiana di Informatica e Diritto*, n. 2/2024, p. 5, rilevando che «nelle politiche e pratiche della transizione digitale la cybersicurezza è oggettivamente una condizione di esistenza: rendere più sicuri reti, sistemi e dati vuole dire giocoforza rendere più sicuri gli Stati e i loro apparati, le persone e le loro vite. Una specie di protezione delle protezioni, una sicurezza massima da cui dipendono le sorti delle Nazioni come degli individui, tutti inesorabilmente immersi nella dimensione digitale». Sul tema del rapporto fra funzione di cybersicurezza e sovranità sia consentito rinviare a G. Barozzi Reggiani, *La race for the cyberspace degli Stati e il tema della cybersicurezza: tra sovranità e modelli di governance*, in *Rivista Italiana di Informatica e Diritto*, n. 2/2025, pp. 1 ss.

12. Decreto recante «Disposizioni urgenti in materia di cybersicurezza, definizione dell'architettura nazionale di cybersicurezza e istituzione dell'Agenzia per la cybersicurezza nazionale».

mento europeo sull'Enisa¹³ (all'art. 2, n. 1), che riferisce espressamente le esigenze di protezione anche agli *utenti* dei sistemi informativi¹⁴.

Quanto, infine, alle possibili “fonti di aggressione” (dei menzionati beni o interessi), si evidenzia che le medesime si riferiscono alle «minacce informatiche» (suscettibili di determinare, se efficacemente attuate, incidenti informatici); queste si configurano – secondo quanto disposto dall'art. 2, n. 8, del menzionato Regolamento 2019/881/UE – al verificarsi di «qualsiasi circostanza, evento o azione che potrebbe danneggiare, perturbare o avere un impatto negativo di altro tipo sulla rete e sui sistemi informativi, sugli utenti di tali sistemi e altre persone»¹⁵; si tratta, anche in questo caso, di definizione ampia, tesa a intercettare una pluralità di fenomeni.

Così sommariamente definito il reticolo dei principali termini e istituti che afferiscono al tema della cybersicurezza, si può iniziare ad abbozzare qualche considerazione, emergente da una lettura combinata delle disposizioni stesse.

Innanzitutto, il tema della sicurezza, a riguardo delle minacce *cyber*, viene in rilievo in riferimento alla totalità delle possibili declinazioni del tema stesso, e in particolare «sia da un punto di vista della tradizionale protezione della collettività organizzata da pericoli di aggressione, provengano essi dall'esterno o dall'interno; sia dal punto di vista della tutela dei diritti dei singoli individui e dalle concrete dimensioni della sua operatività»¹⁶; si tratta dunque (come si

13. Regolamento (UE) 2019/881 del Parlamento europeo e del Consiglio, del 17 aprile 2019, «relativo all'Enisa, l'Agenzia dell'Unione europea per la cybersicurezza, e alla certificazione della cybersicurezza per le tecnologie dell'informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013».

14. Come si accennava, dalle richiamate definizioni emerge come l'insieme delle attività di cybersicurezza possa (anzi debba) essere volto a tutelare anche beni e interessi che afferiscono alla stessa “essenza” e “integrità” dell'ordinamento, come la sicurezza nazionale e l'interesse nazionale nello spazio cibernetico. Ciò risulta ancor più evidente se si considera la definizione di «pregiudizio per la sicurezza nazionale» fornita dall'art. 1, comma 1, lett. *f*, del Dpcm 30 luglio 2020, n. 131 (recante Regolamento in materia di perimetro di sicurezza nazionale cibernetica) e richiamata nell'ambito della definizione di «resilienza nazionale nello spazio cibernetico» posta dall'art. 1, comma 1, lett. *b*, del d.l. n. 82 del 2021; secondo la definizione in parola, configura «pregiudizio per la sicurezza nazionale» un «danno o pericolo di danno all'indipendenza, all'integrità o alla sicurezza della Repubblica e delle istituzioni democratiche poste dalla Costituzione a suo fondamento, ovvero agli interessi politici, militari, economici, scientifici e industriali dell'Italia, conseguente all'interruzione o alla compromissione di una funzione essenziale dello Stato o di un servizio essenziale» (tra quelli indicati all'art. 2 del medesimo decreto).

15. Anche in questo caso, la definizione in parola è richiamata da quella posta dal d.lgs. n. 4 settembre 2024, n. 138.

16. Così, a riguardo delle “dimensioni” della sicurezza, T. Giupponi, *La sicurezza e le sue “dimensioni” costituzionali*, in S. Vida (a cura di), *Diritti umani: Trasformazioni e reazioni*, Bononia University Press, 2008, p. 276.

accennava in precedenza) di tutelare tanto situazioni giuridiche soggettive, riferibili a singoli individui o Enti¹⁷, quanto quelli che, richiamando autorevolissima dottrina che ha indagato il concetto di sicurezza nel diritto costituzionale, potremmo qualificare interessi diffusi o comunque di natura superindividuale, in quanto riferibili a una pluralità non definibile (e non frammentabile) di soggetti o addirittura all'ordinamento nella sua interezza¹⁸.

Viene poi in rilievo il contenuto dei beni/diritti da salvaguardare e, specularmente, quello della natura delle aggressioni che questi possono subire.

E qui si finisce per evocare l'antico tema del rapporto tra sicurezza (e ordine pubblico, che della prima costituirebbe "l'altra faccia", quando non un'endiadi¹⁹, ancorché intesa non come indicante due sostantivi sinonimi, bensì un rapporto di complementarità tra i due²⁰) materiale e ordine pubblico ideale: la prima intesa quale protezione dei beni giuridici (privati o "pubblici") da possibili aggressioni, il secondo quale «garanzia dell'insieme dei principi e dei valori fondanti caratteristici di un determinato ordinamento giuridico»²¹.

È noto che, pur a fronte di qualche indicazione in senso contrario della

17. Sul punto, si ritiene tuttavia di concordare con chi, parlando in generale del concetto di sicurezza (e riportando le conclusioni alle minacce *cyber*) ha inteso escludere che di esso si possa discorrere «in senso autonomo in quanto essa [la sicurezza, *NdR*], al di là della sua definizione in termini di ordine pubblico (id est, di civile convenienza), coincide con il diritto rispetto al quale è richiesta», costituendo dello stesso una «componente necessaria» (cfr., per entrambi i riferimenti, C. Lotta, *Governance della rete, accesso a internet e cybersicurezza. Profili costituzionali*, Editoriale Scientifica, 2024, p. 181).

18. Di interesse pubblico diffuso alla sicurezza ha parlato in particolare A. Pace, *op. cit.*, p. 203.

19. Sulla nozione di sicurezza pubblica intesa come endiadi di ordine pubblico v. P. Barile, *Il soggetto privato nella Costituzione italiana*, Cedam, 1951, p. 116; A. Cerri, *Ordine pubblico*, II, *Diritto costituzionale*, in *Enciclopedia giuridica Treccani*, 1990, XXII, p. 2; G. Corso, *L'ordine pubblico*, il Mulino, 1979, pp. 117 ss.; A. Pace, *Libertà e sicurezza. Cinquant'anni dopo*, cit.; P. Ridola, *Libertà e diritti nello sviluppo del costituzionalismo*, in P. Ridola, R. Nania (a cura di), *I diritti costituzionali*, vol. 1, Giappichelli, 2006, p. 133. Secondo altra, autorevolissima dottrina (cfr. A. Pace, *Ordine pubblico, ordine pubblico costituzionale, ordine pubblico secondo la Corte costituzionale*, in *Giur. Cost.*, 1971 spec. p. 1783), la sicurezza configurerebbe lo stato soggettivo dell'ordine pubblico materiale.

20. Cfr. F. Bocchini, *Convivenza civile e ordine pubblico*, in *Nomos*, n. 1/2020, p. 6, secondo il quale «l'espressione "ordine pubblico", in endiadi con l'espressione "sicurezza", specifica quest'ultima, nel senso che l'espressione sicurezza non si riferisce né alla sicurezza dello Stato, né alla sicurezza privata dei cittadini, ma unicamente e soltanto alla sicurezza pubblica dei cittadini».

21. Cfr. T. Giupponi, *op. cit.*, p. 275. Secondo altra lettura, l'ordine pubblico ideale/immateriale verrebbe a coincidere con la garanzia di quei principi politici ritenuti indispensabili per la sopravvivenza di un ordinamento (così G. Zanobini, *Ordine pubblico*, in *Enciclopedia Italiana*, 1935, XXV, p. 463). Sul concetto di ordine pubblico ideale/immateriale v. anche C. Mortati, *Istituzioni di diritto pubblico*, I, Cedam, 1975, p. 331.

Consulta, la dottrina costituzionalistica pressoché unanime ha ritenuto compatibili con la Costituzione discipline legislative limitative di diritti e libertà fondamentali volte a salvaguardare il solo ordine pubblico (la sicurezza) materiale.

Ed è proprio tale dimensione che a chi scrive pare venga in rilievo a riguardo dei temi concernenti la cybersicurezza, constatazione che consente di non porsi eccessivi scrupoli in riferimento alla giustificabilità, sul piano costituzionale, di discipline che prevedono, quale risposta a minacce o incidenti informatici, l'esecuzione di attività che si traducano in potenziali limitazioni di diritti e libertà fondamentali; e infatti, una disciplina legislativa che preveda attività di reazione a una minaccia o a un incidente informatico – da parte dei pubblici poteri ovvero di privati in ciò “autorizzati” dalla disciplina stessa – se anche dovesse potersi considerare limitativa di una qualche libertà costituzionale ben difficilmente potrebbe considerarsi in contrasto con la Carta (laddove risultasse, naturalmente, ragionevole e proporzionata).

E ciò nonostante sussistano alcuni margini di incertezza circa l'esatta individuazione del perimetro dei beni che le attività di cybersicurezza mirano a tutelare.

A fronte di tali conclusioni, si osserva che ad approdi in tutto o in parte diversi si deve invece giungere in riferimento alla collocazione, nel quadro costituzionale, del tema della disinformazione, o quantomeno di alcune delle sue manifestazioni concrete.

3. Caratteristiche della disinformazione e strumenti (giuridici) per contrastarla

La disinformazione è stata definita dal *High Level Group on fake news and online disinformation* della Commissione europea come un insieme comprendente tutte quelle forme di comunicazione «che si sostanziano di quelle informazioni false, inaccurate o fuorvianti, artificialmente 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»²².

22. La menzionata definizione è richiamata nel recente saggio di O. Pollicino, P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *federalismi.it*, editoriale 29 maggio 2024, V-VI, al quale si rinvia per ulteriori approfondimenti sul concetto, in uno con il lavoro di C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quad. cost.*, 2023, pp. 543 ss.

Rispetto alle semplici *fake news*, la disinformazione si caratterizza per essere attività voluta, preordinata, spesso “progettata” per conseguire, mediante diffusione di notizie non vere o inaccurate, determinati obiettivi, e «far giungere ai destinatari un preciso messaggio, con l’intento di disinformare, di creare scandalo o, molto spesso, di attirare gli utenti su determinate pagine internet, per assicurare ai relativi gestori ritorni in termini economici e di visibilità»²³.

Considerati la natura e i contenuti delle strategie di disinformazione – che oggi peraltro sfruttano i canali e i servizi della Rete, risultando spesso, proprio in ragione di ciò, in grado di conseguire i propri non nobili obiettivi con particolare efficacia – rispetto ad attività volte a contrastare le medesime vengono astrattamente in rilievo temi concernenti la restrizione, oltre che eventualmente della libertà di iniziativa economica (pensiamo a quella dei fornitori dei servizi della società dell’informazione, e in specie dei servizi Ott), della libertà di manifestazione del pensiero (in particolare degli utenti di quei servizi), ovvero di quella libertà che sappiamo essere stata definita, dalla Consulta (tra le altre cose) «pietra angolare dell’ordine democratico» (sent. 84 del 1969), «cardine di democrazia nell’ordinamento generale» (sent. 126 del 1985), elemento «coessenziale al regime di libertà garantito dalla Costituzione» (sent. 11 del 1968).

Invero, che l’autore dell’attività di disinformazione, in quanto propalatore consapevole di notizie false o inaccurate, possa invocare le disposizioni costituzionali a tutela della libertà di manifestazione del pensiero, è circostanza tutt’altro che pacifica.

E infatti, una lettura diffusa e condivisa dell’art. 21 Cost. afferma che tale disposizione non coprirebbe diffusioni di notizie o manifestazioni di pensieri il cui contenuto sia *oggettivamente* falso e la cui falsità risulti conosciuta da chi li esprime o anche solo veicola²⁴; considerato che la disinformazione costituisce at-

23. Così C. Colapietro, *Libera manifestazione del pensiero, fake news e privacy, oggi*, in *dirittifondamentali.it*, n. 2/2022, p. 430.

24. Il nucleo di tale tesi è costituito dal fatto che colui che riporta consapevolmente una notizia falsa non sta esprimendo il proprio pensiero. Tale posizione è espressa, tra gli altri (anche in riferimento al tema delle *fake news*), da R. Perrone, *Fake news e libertà di manifestazione del pensiero: brevi coordinate in tema di tutela costituzionale del falso*, in *Nomos. Le attualità del diritto*, n. 2/2018, pp. 6 ss.; M. Bassini, E. Vigevari, *Primi appunti su fake news e dintorni*, in *MediaLaws*, 2017, pp. 14 ss.; O. Pollicino, *La prospettiva costituzionale sulla libertà di espressione nell’era di Internet*, in G. Pitruzzella, O. Pollicino, S. Quintarelli, *Parole e potere*, Egea, pp. 46 ss.; M. Monti, *Fake news e social network: la verità ai tempi di Facebook*, in *MediaLaws*, 2017, p. 82 s.; M. Cuniberti, *Il contrasto alla disinformazione in rete tra logiche del mercato e (vecchie e nuove) velleità di controllo*, in *MediaLaws*, 2017, p. 30; P. Barile, *Diritti dell’uomo e libertà fondamentali*, il Mulino, 1984, p. 238 ss.; C. Esposito, *La libertà di manifestazione del pensie-*

tività voluta (e costruita) per diffondere falsità, l'autore della stessa non potrebbe invocare i presidi costituzionali della libertà di manifestazione del pensiero.

Ciò posto, si evidenzia in primo luogo la difficoltà di individuare esattamente il novero dei soggetti ai quali, per le ragioni sopra indicate, l'invocabilità delle disposizioni costituzionali (nazionali o europee) a tutela della manifestazione del pensiero non potrebbe ritenersi riferibile, oltre a quelle connesse alla effettiva identificazione degli "autori"; secondariamente, si rileva che, come è stato affermato, la Costituzione «non tutela soltanto le manifestazioni oggettivamente veritiere, bensì tutte le manifestazioni del pensiero, compresi dunque quei fatti oggettivamente errati, qualora in buona fede vengano ritenuti veri da chi ne affermi l'esistenza»²⁵.

In questo senso, lo *sharing*, la condivisione, il *repost* in Rete e tramite *social network* di una notizia falsa, da parte di un utente inconsapevole del contenuto non veritiero della stessa – che quindi, di fatto, realizza attività di c.d. *misinformation* – dovrebbero ritenersi attività coperte dalle tutele dell'art. 21 Cost. (posto che riguarderebbero la diffusione di una notizia falsa solo in senso oggettivo e non soggettivo), anche qualora le stesse risultassero essere state effettuate nell'ambito di una strategia di disinformazione.

Ciò imporrebbe dunque di distinguere tra l'autore della strategia di disinformazione e gli altri diffusori inconsapevoli delle *fake news* alla stessa riconducibili: il che, rispetto a modalità disintermediate di circolazione di notizie e informazioni – tipiche dei servizi Ott – appare tutt'altro che semplice.

E anche l'affidare tale compito alle grandi o grandissime piattaforme, come da un po' di tempo l'Unione europea sta provando a fare (il riferimento va in particolare al *Code of Practice on Disinformation*), può risultare una strategia poco efficace, se non addirittura pericolosa per le possibili derive censorie che la stessa può assumere²⁶.

A tutto ciò si aggiunga un'ulteriore considerazione.

A prescindere dalla esatta definizione dell'ambito di applicazione dell'art. 21 Cost., la diffusione di notizie false, come osservava un maestro come Paolo Barile, può essere considerata illecita solo laddove la stessa risulti suscettibile di ledere diritti (o libertà) o valori costituzionalmente rilevanti²⁷.

ro, Giuffrè, 1958, p. 36 ss.; S. Fois, *Principi costituzionali e libera manifestazione del pensiero*, Giuffrè, 1957, p. 209.

25. Così L. Califano, *La libertà di manifestazione del pensiero... in rete; nuove frontiere di esercizio di un diritto antico. Fake news, hate speech e profili di responsabilità dei social network*, in *federalismi.it*, n. 26, 2021, p. 14.

26. Rischi messi bene in luce, da ultimo, da G.E. Vigevari, *Potere politico e mezzi di comunicazione*, in *Rivista AIC*, n. 2/2025, pp. 236 ss. (*passim*).

27. P. Barile, *Diritti dell'uomo e libertà fondamentali*, cit., p. 229.

Nel caso della disinformazione, non è semplice definire quali beni, diritti o valori costituzionali possano risultare, dalla medesima, messi in pericolo o incisi.

La disinformazione indubbiamente inganna, può indurre alcuni soggetti a compiere scelte sbagliate e a credere in una realtà alterata: tuttavia, tale constatazione non basta di per sé a giustificare la messa a punto di una disciplina legislativa limitativa di tale attività; e ciò anche in considerazione delle difficoltà che si riscontrano, sul piano tanto teorico quanto pratico, nel ricondurre all'art. 21 (o ad altra disposizione costituzionale) un diritto a essere correttamente informati dei singoli, avente una specifica fisionomia (e che dovrebbe in quanto tale risultare tutelabile in via diretta in sede giurisdizionale²⁸).

La disinformazione, poi, come osservato da più parti (in ambito tanto scientifico quanto istituzionale), proprio a cagione della sua capacità di distorcere la realtà può inquinare il dibattito pubblico, venendo a incidere concretamente sui processi democratici, compresi quelli elettorali, e di fatto orientando il voto²⁹.

28. Nonostante la menzione del diritto a essere informati si rinvenga in alcune pronunce della Corte (v. per tutte la n. 349 del 1990 e la n. 112 del 1993) la dottrina tende a escludere che un diritto consimile effettivamente sussista e sia riconducibile all'art. 21 Cost.; tale affermazione non porta a escludere, tuttavia, che rispetto a soggetti che professionalmente informano e/o possiedono la capacità di diffondere informazioni (come i *mass media*, ma oggi anche i *social network*) non sia possibile definire discipline che impongano obblighi a taluni operatori di limitare o contrastare la circolazione di *fake news* e disinformazione, e in particolare di quelle declinazioni delle stesse che appaiono idonee a mettere a repentaglio diritti o libertà costituzionalmente protetti (come, ad esempio, quello alla salute): esempi di discipline di tal fatta si rinvencono tanto nella Direttiva Smau del 2018 (Direttiva 2018/1808/UE, che ha modificato la Direttiva 2010/13/UE «relativa al coordinamento di determinate disposizioni legislative, regolamentari e amministrative degli Stati membri concernenti la fornitura di servizi di media audiovisivi») quanto nel Regolamento *Digital Services Act* (Regolamento (UE) 2022/2065, «relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali)». In tema v. anche M. Orofino, *Art. 21 Cost.: le ragioni per un intervento di manutenzione ordinaria*, in *MediaLaws*, 2019, p. 87: «mentre per l'espressione non è possibile sostenere che esista un obbligo generale di dire il vero e dunque che la norma costituzionale tuteli solo le espressioni vere o verosimili; per l'informazione è possibile, almeno (e forse solo) in linea teorica, individuare limiti alla diffusione di fatti falsi al fine di tutelare altri diritti o gli stessi interessi costituzionalmente rilevanti che pure la libertà d'informazione protegge, come appunto il processo politico e democratico».

29. Se è vero, come affermava Costantino Mortati nelle sue *Istituzioni di diritto pubblico* (cit., Tomo II, p. 972), che la garanzia del buon funzionamento del sistema democratico poggia «sulla più ampia libertà di manifestazione del pensiero, essendo essa che alimenta la forza sociale di base, che è la pubblica opinione», occorre effettivamente interrogarsi sulle conseguenze di un'opinione pubblica che si formi su notizie false e teorie prive di alcun fondamento scientifico, ovvero in contesti nei quali le voci troppo forti di alcuni attori finiscano per sovrastare, se non zittire del tutto, quelle degli altri. Di certo, il proliferare di *fake news* e disinformazione in

È un fenomeno, quest'ultimo, che risulta indubbiamente reale, e di cui si sono avute manifestazioni concrete, da ultima quella concernente le elezioni in Romania, in cui l'affermata sussistenza di attività sistematica di disinformazione funzionale ad alterare le elezioni ha spinto la Corte costituzionale di tale ordinamento ad annullare le elezioni stesse³⁰.

Eppure, è tutt'altro che semplice individuare, in riferimento al descritto scenario, uno specifico bene costituzionale le cui esigenze di salvaguardia legittimerebbero la definizione di una disciplina legislativa (limitativa delle attività costituenti disinformazione); e ciò vale in particolare nei confronti delle attività di quei soggetti che, pur diffondendo inconsapevolmente notizie false (e dunque contribuendo ad alimentare le strategie di disinformazione) possono invocare, proprio a cagione della loro buona fede nel contenuto veritiero di quanto diffuso, l'ombrello dell'art. 21.

Si potrebbe sostenere che a fronte del delineato quadro viene in rilievo l'ordine pubblico, la tutela del quale legittimerebbe limiti alla diffusione di notizie false, o sanzioni per la loro diffusione, come del resto prevede l'art. 656 c.p.; accedere a una simile lettura, tuttavia, imporrebbe di prendere a riferimento una declinazione dell'ordine pubblico (e/o della sicurezza) che appare distante dalla dimensione "materiale" dello stesso, e più vicina all'ordine pubblico ideale, con quanto a ciò consegue.

Né pare invocabile il "bene" sicurezza dello Stato (e/o la «sicurezza nazionale»): nonostante l'inquinamento dei processi elettorali attenga, evidentemente, ai meccanismi profondi di funzionamento degli ordinamenti democratici, non ritengo possa sostenersi che le attività di disinformazione si traducano in attività di aggressione violenta e materiale a beni o assetti dello Stato medesimo.

Ciò a differenza delle minacce informatiche che, come si è detto in prece-

Rete ha messo in seria crisi il concetto e lo schema teorico del *free marketplace of ideas* (scaturito originariamente dalla *dissenting opinion* espressa dal Justice Holmes della Corte Suprema nel caso *Abrams v. United States* del 1919) che sostiene che la genuinità del dibattito pubblico possa essere assicurata solo da una garanzia al massimo grado della libertà di espressione, che tramite il confronto tra le idee finirà per marginalizzare quelle false e/o nocive (è stato in questo senso osservato – cfr. G. De Gregorio, *The marketplace of ideas nell'era della post-verità: quali responsabilità per gli attori pubblici e privati online?*, in *MediaLaws*, 2017, p. 95 – che «come in un mercato concorrenziale dove i migliori prodotti o servizi prevalgono, lo stesso avviene per le informazioni vere e corrette destinate a essere predilette dal mercato stesso»). Peraltro, giova osservare che lo schema del *free marketplace of ideas*, a lungo seguito nell'ambito dell'ordinamento statunitense, non ha mai attecchito, quantomeno nella sua formulazione "pura", in ambito Continentale.

30. In tema, F. Rosa, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025, pp. 53 ss.

denza, possono essere suscettibili di ledere, oltre che reti, servizi o utenti (dei servizi,) anche la resilienza dello spazio cibernetico, e più in generale di arrecare pregiudizio alla sicurezza nazionale, connessa alla indipendenza e all'integrità della Repubblica e delle istituzioni democratiche.

Tali considerazioni porterebbero a sostenere che, nonostante la disinformazione “corra sulla Rete” e venga effettuata negli spazi e tramite i servizi del cyberspazio, tra quest'ultima e la cybersicurezza non vi sia alcun collegamento.

Ma si tratterebbe di conclusione, invero, forse troppo frettolosa.

4. Considerazioni conclusive. La disinformazione quale problema di cybersicurezza: una realtà già attuale o una semplice prospettiva (*de iure condendo*)?

Da qualche anno a questa parte, l'Enisa riconduce espressamente (anche se probabilmente in modo “atecnico”) la disinformazione al novero delle minacce informatiche³¹, e da più parti si sente parlare di disinformazione come di un problema di cybersicurezza.

Per tentare di comprendere le ragioni di affermazioni consimili, si prendano le mosse da una considerazione di carattere generale, concernente la constatazione che gli autori degli attacchi informatici fanno usualmente leva, in una logica di massimizzazione della efficacia degli attacchi stessi, principalmente sulle vulnerabilità del cyberspazio (e dei suoi luoghi).

Ciò detto, e ragionando sul concetto di vulnerabilità, occorre sottolineare che nel cyberspazio si “muovono” anche gli esseri umani, che anzi costituiscono componenti costitutive ed essenziali di uno dei *layers* di cui lo stesso è composto³².

31. Il dato è chiaramente sottolineato nel Rapporto (stilato da Enisa ed European Union External Action) *Foreign information manipulation and interference (Fimi) and cybersecurity – threat landscape*, del dicembre 2022, e poi ripreso nei successivi Rapporti *Threat Landscape* dell'Enisa.

32. Secondo le ricostruzioni più diffuse, il cyberspazio si comporrebbe di tre livelli (*layers*) denominati, secondo le diciture più utilizzate, “fisico”, “logico” e “sociale”. Il livello fisico è quello nel quale si rinvencono le infrastrutture e le componenti *hardware* che consentono la circolazione dei dati e lo svolgimento delle attività nel cyberspazio. Al livello logico si collocano i *software* e i codici che consentono alle macchine di dialogare tra loro, mentre il piano sociale è formato dalle interazioni che possono verificarsi fra persone fisiche o, come ormai sempre più spesso accade, tra persone e macchine o tra macchine e macchine. In tema cfr. P. Pătrașcu, *Missions and Actions Specific to Cyberspace Operations*, in *Proceedings of the International Conference Knowledge-based organization*, 2019, vol. 25, n. 3, spec. p. 53; E.D. McCroskey, C.A. Mock, *Operational Graphics for Cyberspace*, in *Joint Forces Quarterly*, 2017, vol. 85, spec. p. 44.

Ne discende che anche gli individui, con le proprie condotte, possono determinare il manifestarsi di vulnerabilità, e quindi di occasioni proficue per i cyberattacchi.

È il concetto di *human vulnerability*³³ (che si affianca alle vulnerabilità che riguardano le reti, i *software* o le infrastrutture), “falla del sistema” sulla quale la disinformazione può agire.

In particolare, mediante sfruttamento delle debolezze o degli errori umani le campagne di disinformazione «can be a means for various types of cyberattacks, such as deceptive online identity theft attempts (phishing attacks) or the malicious spreading of harmful software»³⁴.

Ecco, dunque, il collegamento tra disinformazione e cybersicurezza, che oltre a essere nella realtà delle cose assume anche rilevanza giuridica: la disinformazione può essere veicolo e strumento di minacce e attacchi informatici, proprio nella misura in cui tende ad aumentare la *human vulnerability* e a far leva sulla stessa per creare le condizioni di un effettivo attacco (sia esso costituito da un *data breach*, dalla diffusione di un *virus* o da una delle altre possibili declinazioni di tale tipologia di attacchi).

Alla luce delle esposte considerazioni, può la disinformazione essere effettivamente considerata, sul piano giuridico, una minaccia informatica, e dunque una questione di cybersicurezza?

Se si considera nuovamente la definizione di minaccia informatica di cui si è dato conto in precedenza (quella posta dal Regolamento 2019/881/UE) non si rinvencono reali ragioni per fornire una risposta negativa, posto che dette minacce sono costituite da «qualsiasi circostanza» suscettibile di danneggiare (quindi anche solo in potenza) sistemi e reti (o utenti).

Vero è che la disinformazione risulta solo “strumentale” a garantire l’efficacia dell’attacco – non costituendo attacco “in sé – e dunque contrastare la medesima significa anticipare la soglia della reazione ordinamentale, ma è altrettanto vero che la riportata definizione qualifica come minacce azioni che possano concretizzare anche solo un pericolo di danno, e mi pare pertanto idonea a ricomprendere le attività prodromiche agli attacchi veri e propri (specie se di questi finiscono per costituire componente essenziale, ai fini della loro efficacia).

Mi pare poi che il problema legato alla individuazione dei beni giuridici da salvaguardare non si ponga: le falsità diffuse ad arte tramite le strate-

33. Sul concetto v. M. Alsharif, S. Mishra, M. AlShehri, *Impact of Human Vulnerabilities on Cybersecurity*, in *Computer Systems Science & Engineering*, 2022, vol. 40, n. 3, p. 1153.

34. Si v. data.europa.eu/en/publications/datastories/leveraging-open-data-tackle-cyber-disinformation.

gie di disinformazione per la realizzazione di cyberattacchi possono ledere diritti, beni e valori di sicura rilevanza costituzionale, come si è in precedenza argomentato.

In ragione di quanto esposto, ritengo che le strategie di disinformazione potrebbero, già all'attualità, essere considerate quali minacce informatiche, e dunque essere astrattamente destinatarie delle discipline e delle attività/funzioni (oltre che degli obblighi per i diversi soggetti) che i quadri giuridici nazionale ed europeo riservano alle stesse.

Il discorso, tuttavia, vale solo per alcune declinazioni delle attività di disinformazione, e in particolare quelle volte ad aumentare la *human vulnerability* sul piano della *cybersecurity*, proprio in ragione del fatto che tale vulnerabilità può determinare lesioni concrete e materiali a beni, diritti e valori, causando la sottrazione di dati o il danneggiamento di software o reti.

Allo stato ritengo invece siano da escludere da tale novero quelle attività di disinformazione non aventi i descritti obiettivi specifici, in quanto al più suscettibili di ledere l'ordine pubblico ideale o comunque non uno dei beni-valori cui la disciplina nazionale o europea in materia di cybersicurezza è volta a fornire protezione.

Tutto ciò detto, e in conclusione, chi scrive ritiene che sarebbe auspicabile un intervento legislativo volto a delineare con precisione le caratteristiche di quelle tipologie di attività di disinformazione che possano assumere effettivo rilievo in ambito di *cybersecurity*: se una disciplina legislativa che consenta di effettuare detta operazione di qualificazione potrebbe infatti ritenersi già sussistere (in particolare nella citata definizione di minacce informatiche), un'operazione normativa di individuazione delle specifiche declinazioni della disinformazione che assumono rilievo appare senz'altro opportuna al fine di garantire certezze agli operatori nonché – in relazione a quei soggetti per i quali, anche rispetto alla diffusione di notizie false, risultino riferibili le tutele di cui all'art. 21 Cost. – per definire una limitazione alla libertà di manifestare il pensiero che risulti, oltre che costituzionalmente necessitata (dalla necessità di tutelare diritti, libertà o valori di rango costituzionale) chiara e proporzionata, e quindi, in ultima istanza, ragionevole.

OLTRE LA DISINFORMAZIONE: IL PROBLEMA DELLA MANIPOLAZIONE NEL QUADRO DELLA REGOLAZIONE ALGORITMICA EUROPEA

Ludovica Durst*

SOMMARIO: 1. Introduzione. Disinformazione e manipolazione nella società digitale – 2. L’incerto confine definitorio della manipolazione e l’ambiguo rapporto con i media: l’apporto della prospettiva sociologica – 3. Manipolazione e profilazione negli interventi regolatori dell’UE – 3.1. Consenso, profilazione e manipolazione: profili di tutela nel Dsa e nel Dma – 3.2. La manipolazione delle informazioni tra pluralismo e profilazione nel *Regolamento sulla libertà dei media* e nel *Regolamento sulla pubblicità politica* – 3.3. Il contrasto alla “manipolazione algoritmica” e l’AI Act – 4. La necessità di un quadro di risposte integrato: lo scudo democratico dell’UE e i nuovi rischi cognitivi per l’autodeterminazione individuale.

1. Introduzione. Disinformazione e manipolazione nella società digitale

Il perpetrarsi del fenomeno della disinformazione attraverso il *medium* digitale costituisce una problematica nevralgica per le democrazie contemporanee. Numerosi sono gli studi di sociologia, teorie della comunicazione e di diritto, i report indipendenti e quelli prodotti da istituzioni nazionali e sovranazionali che ne illustrano forme e nuove sfide, con intenti definitori e rimedi regolatori introiettati anche nella legislazione europea e nelle forme di regolazione – o co-regolazione – più recenti¹. È noto come, al riguardo, una delle principali siste-

* Assegnista di ricerca in Diritto costituzionale e pubblico, Sapienza Università di Roma.

1. Sulle trasformazioni degli strumenti di regolazione adottati nel quadro europeo, con riguardo alla moderazione dei contenuti in ambito digitale, e in particolare sulle tendenze relative al passaggio da soft-law a hard-law e da forme di autoregolazione a etero- o co-regolazione, v. *ex multis*, con diverse prospettive, G. De Minico, *Le fonti del diritto: un argine all’intelligenza artificiale?*, in *Rivista AIC*, n. 3/2025; O. Pollicino, *Regolazione e innovazione tecnologica nell’ordinamento della rete*, e G.E. Vigevari, *Potere politico e mezzi di comunicazione*, in *Rivista*

matizzazioni di riferimento nel panorama europeo sia offerta dal *Codice di condotta sulla disinformazione*, originariamente elaborato sotto forma di *Codice di buone pratiche* nel 2018, rafforzato nel 2022, e di recente integrato nel *Digital Services Act* (Dsa)². Pur non potendosi ritenere del tutto superate le difficoltà pratiche che si incontrano nel tentativo di procedere a nette distinzioni tra ciò che costituisce “disinformazione”, “misinformazione”, o “malinformazione” e dunque l’efficacia specifica delle conseguenti azioni di contrasto, proprio nel Codice di condotta rafforzato si rinviene un primo rimando a una nozione particolarmente ampia di “disinformazione”, tale da ricomprendere all’interno del concetto anche quanto ascrivibile a fenomeni affini quali la *misinformazione*, le operazioni di influenza dell’informazione e le ingerenze straniere nello spazio informativo, che trovano già definizione nel *Piano d’azione per la democrazia europea* del 2020 della Commissione europea³.

AIC, n. 2/2025; A. Simoncini, *La co-regolazione delle piattaforme digitali*, in *Rivista trimestrale di diritto pubblico*, n. 4/2022.

2. Ai sensi dell’art. 45 del Dsa, l’adesione a Codici di condotta approvati dalla Commissione e dal Comitato europeo per i servizi digitali costituisce parametro di conformità (*compliance*) alla disciplina e agli obblighi previsti dal Dsa in materia, in particolare quale misura di mitigazione dei rischi sistemici da parte delle piattaforme e dei motori di ricerca di grandi dimensioni (Vlopse) firmatari (fra cui Facebook, Instagram, LinkedIn, Bing, TikTok, YouTube e Google Search), secondo quanto prescritto dall’art. 35 del Dsa. Fra le principali misure di contrasto alla disinformazione, si prevedono: la “demonetizzazione” dei ricavi pubblicitari per i fornitori di disinformazione; misure di trasparenza rafforzate per la pubblicità politica; rafforzamento delle misure per l’integrità dei servizi contro la manipolazione effettuata attraverso tecniche quali *bot*, falsi account e identità, *deepfakes* dannosi; misure di *empowerment* degli utenti (per assumere decisioni informate) e dei ricercatori (nell’accesso ai dati non personali); maggior coinvolgimento di *fact-checkers*; l’istituzione di un Centro per la trasparenza e una *task force* permanente; l’accesso a report periodici e attività di monitoraggio. L’approvazione del Codice è avvenuta il 13 febbraio 2025, con decorrenza a partire dal 10 luglio dello stesso anno. Va inoltre sottolineata come nell’articolato del Dsa non compaia mai il termine “disinformazione”, presente solo nei considerando. A partire da questo rilievo, v. l’analisi di G. Contissa, F. Galli, *La governance della “disinformazione aumentata” tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025 (spec. p. 133 ss.), e più ampiamente l’intero fascicolo speciale *Information disorder e sistema democratico*, curato da L. Califano, F. Fabrizzi, G. Sartor (*federalismi.it*, n. 15, 2025).

3. Così nel Preambolo del Codice, lett. a), con l’esclusione tuttavia dal concetto di “disinformazione” della pubblicità ingannevole, degli errori di cronaca, la satira e la parodia, le notizie e i commenti chiaramente identificati come di parte, e senza pregiudizio degli obblighi giuridici vincolanti, dei codici di autoregolamentazione della pubblicità e delle norme relative alla pubblicità ingannevole. Il Codice rimanda alla Comunicazione della Commissione sul *Piano d’azione per la democrazia europea* (COM(2020)790 final), per le seguenti definizioni: la “misinformazione” (o “cattiva informazione”) è costituita da contenuti falsi o fuorvianti, condivisi senza intenzioni fraudolente, anche se gli effetti possono essere comunque dannosi, ad esempio quando le persone condividono informazioni false con amici e familiari in buona fede; la “disinformazione” è un contenuto falso o fuorviante diffuso con l’intento di ingannare o ottenere

Per coloro che si occupano in maniera specialistica del tema sono ormai piuttosto conosciute le modalità e gli intenti con cui si diffonde la disinformazione attraverso il *medium* digitale, in particolare internet e le piattaforme online, secondo una duplice alleanza che vede, da un lato, pienamente integrarsi la disinformazione con le tecniche e le logiche di funzionamento che si trovano alla base del modello di business delle piattaforme, il cui sviluppo contribuisce ad alimentare proprio quel tipo di contenuti in quanto funzionali ad attrarre maggiormente l'attenzione dell'utente-consumatore e a "ingaggiarlo" a restare sulla piattaforma⁴; dall'altro lato, l'utilizzo di internet e delle piattaforme risulta estremamente vantaggioso ed efficace al fine di veicolare contenuti di disinformazione e amplificarne la circolazione, in maniera strumentale agli intenti di destabilizzazione dei sistemi democratici perseguiti dai soggetti artefici di disinformazione. Questi ultimi infatti possono approfittare non solo delle dinamiche di polarizzazione e radicalizzazione dei contenuti disinformativi diffusi attraverso i sistemi di profilazione e di ridondanza delle *eco chambers* e delle *filter bubbles*, ma delle stesse caratteristiche di "disintermediazione", mancanza di controllo e di responsabilità che connotano il funzionamento delle piattaforme, nonché del loro basso costo al fine di potenziarne ulteriormente gli effetti, grazie anche all'utilizzo dell'IA generativa e dei *bot* per la creazione e propagazione di contenuti artificiali, i famosi *deepfakes*⁵.

Al di sotto di queste dinamiche, altamente problematiche e di difficile soluzione, soprattutto dal punto di vista di una risposta giuridica che risulti efficace e aggiornata, emerge uno specifico aspetto, già peraltro colto dall'attenzione del legislatore europeo: ovvero quello delle tecniche e degli effetti

un vantaggio economico o politico e che può provocare un danno pubblico; l'"operazione di influenza informativa" fa riferimento agli sforzi coordinati da parte di attori nazionali o stranieri per influenzare un pubblico target utilizzando una serie di mezzi ingannevoli, tra cui la soppressione di fonti di informazione indipendenti in combinazione con la disinformazione; l'"ingerenza straniera nello spazio informativo", spesso attuata nell'ambito di un'operazione ibrida più ampia, può essere intesa come un tentativo coercitivo e ingannevole di ostacolare la libera formazione ed espressione della volontà politica degli individui da parte di uno Stato straniero o dei suoi agenti. Si v. anche J. Pamment, *The EU's role in fighting disinformation: Taking back the initiative*, Carnegie Endowment for International Peace Working Paper, July 2020.

4. Il riferimento è alla c.d. "economia dell'attenzione" che muove dagli studi di economia comportamentale (*economic behaviours*) sviluppati soprattutto a partire dagli anni '70, i cui esiti deteriori sulla società odierna sono stati incisivamente descritti in termini di "capitalismo della sorveglianza" da S. Zubhoff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, 2019.

5. Su tali meccanismi, strettamente legati alla questione della "natura" delle piattaforme, con particolare riguardo alla spinta alla radicalizzazione dei contenuti attraverso il superamento della c.d. "finestra di Overton", v. anche R. Déchaux, *The Right to Free Elections in the TikTok Era: Towards Digital Expertise of the Venice Commission?*, in *MediaLaws*, n. 2/2025.

manipolativi sottostanti al fenomeno della disinformazione, i quali incidono, a ben vedere, su una pluralità di livelli e dimensioni, sia di tipo collettivo sia individuale⁶.

Sotto il profilo collettivo, evidenti e piuttosto risaputi sono gli impatti distorsivi prodotti sui processi democratici e sulla formazione dell'opinione pubblica, che costituiscono l'obiettivo principale delle campagne (anche coordinate) di disinformazione. In tale contesto, è piuttosto lineare – e risalente – l'equazione che dalla manipolazione delle informazioni conduce alla manipolazione dell'opinione pubblica, pur con una serie di difficoltà legate alla dimostrazione “quantificabile” dell'effettivo impatto prodotto e ancor più con riguardo alla complessa configurazione dell'opinione pubblica nel contesto attuale, stante la sua frammentazione in una pluralità di “micro-pubblici” – per riprendere sul punto la più recente riflessione habermasiana⁷ – di cui ha ampiamente dato conto la dottrina (non solo giuridica).

Tale equazione, come si vedrà più dettagliatamente di seguito, è ben presente nei recenti interventi regolatori dell'Unione europea in ambito digitale, ove il riferimento alla “manipolazione delle informazioni”, in endiadi con la disinformazione, appare sovente quale *ratio* giustificativa e fondativa delle normative introdotte⁸.

L'effetto che si produce, tuttavia, non investe il solo piano collettivo, come potenziale destabilizzazione del sistema democratico, ma determina significative ricadute anche – e spesso preliminarmente – sulla dimensione individuale, a partire da quel “diritto a essere informato” e a “formare liberamente le proprie opinioni e idee” che non solo costituisce un aspetto essenziale della libertà di manifestazione del pensiero e di informazione, ma incide direttamente sulla stessa “autodeterminazione individuale”, quale capacità di

6. Talché, come si evince dai citati documenti europei, la stessa disinformazione può considerarsi una forma o tecnica di manipolazione. Vengono in particolare considerate, quali “tattiche di manipolazione”, profili falsi o ingaggi falsi, per amplificare artificialmente le narrazioni su specifiche questioni politiche e sfruttare le divisioni esistenti nella società, con lo sviluppo di un'intera industria intorno alla manipolazione dei social media (cfr. il *Piano d'azione per la democrazia europea* della Commissione del 2020). Sull'impatto, anche dal punto di vista metodologico, delle dinamiche disinformative delle piattaforme sul funzionamento della sfera pubblica, v. anche C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, 3/2023.

7. J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Raffaello Cortina Editore, 2023.

8. In tal senso, per un'analisi a partire dalla disciplina consumeristica, ove emerge il problema della manipolazione in relazione alle condizioni di vulnerabilità del soggetto, v. P. Hacker, *Manipulation by algorithms. Exploring the triangle of unfair commercial practice, data protection, and privacy law*, in *European Law Journal*, 2023, vol. 29, n. 1-2.

formare ed esprimere le proprie scelte, e che ad esempio nell'ambito dei processi elettorali trova tutela attraverso la garanzia del requisito della "libertà del voto"⁹.

Al centro di tale problematica si trovano le sempre più pervasive potenzialità derivanti dall'applicazione personalizzata degli algoritmi, ovvero quella "manipolazione algoritmica" che si basa sul trattamento di dati e informazioni personali e che appare in grado di incidere, finanche, sulla autodeterminazione del singolo e la sua "libertà di pensiero", tramite l'azione combinata di profilazione, personalizzazione (targetizzazione) e polarizzazione dei contenuti, sia sul piano individuale che su quello collettivo.

Il presente contributo si propone dunque di ripercorrere gli aspetti qui sopra sinteticamente esposti, a partire da un primo inquadramento concettuale della manipolazione, di cui – nonostante i numerosi riferimenti presenti anche negli atti di diritto derivato dell'Unione europea – manca una definizione giuridica, al fine in particolare di distinguerla da altre nozioni affini (quali persuasione o influenza), per poi soffermarsi sulle disposizioni che la considerano espressamente e sono volte a fornire una prima tutela del singolo contro questo tipo di trattamenti: in particolare, si richiameranno le disposizioni dell'AI Act che collocano alcune di queste "manipolazioni" tra i sistemi a rischio elevato o proibiti, nonché le garanzie di trasparenza che costituiscono l'ossatura portante del Dsa, del Regolamento sulla pubblicità politica e dello *European Media Freedom Act*, pur restando un imprescindibile punto di riferimento innanzitutto l'art. 22 del Gdpr.

Infine, ci si interrogherà sulla efficacia dei rimedi giuridici predisposti e sull'opportunità di delineare in maniera più puntuale un diritto all'autodeterminazione nell'ambito digitale che, muovendo dal nucleo teorico della nozione di "autodeterminazione informativa", possa eventualmente tener conto dei rischi di manipolazione più pervasivi derivanti dalle tecnologie attuali, anche attraverso una possibile riconsiderazione della "libertà di pensiero" quale ambito di protezione del "foro interiore"¹⁰.

9. Per una lettura volta a sottolineare sotto tale profilo la coincidenza della libertà del voto con l'autodeterminazione e la "libertà morale del votante", v. anche R. Nania, *La libertà individuale nell'esperienza costituzionale italiana*, Giappichelli, 1989, p. 169 ss., e, con enfasi anche sui condizionamenti psichici, T. Martinez, *Art. 56-58*, in G. Branca (a cura di), *Commentario alla Costituzione*, Zanichelli, 1984, p. 79 ss.

10. Tale aspetto appare oggi ricevere una rinnovata attenzione da parte della riflessione costituzionalistica, sollecitata dagli stimoli posti dall'avanzamento tecnologico e dalle inedite capacità persuasive e pervasive da questo mostrate, come messo in luce in particolare da N. Farahany, *Difendere il nostro cervello. La libertà di pensiero nell'era delle neurotecnologie*, Bollati Boringhieri, 2024. Si v. di recente, al riguardo, A. Lamberti, *Costituzionalismo digitale, poteri delle piattaforme, intelligenza artificiale e democrazia*, in *Consulta online*, fasc.

2. L'incerto confine definitorio della manipolazione e l'ambiguo rapporto con i media: l'apporto della prospettiva sociologica

Provando dunque ad avviare la riflessione a partire dal concetto di manipolazione, appare evidente come debba tenersi in considerazione una duplice prospettiva di riferimento: quella dell'impatto sul piano collettivo, espressa in termini di destabilizzazione democratica, e quella dell'incidenza sul piano individuale, tanto in relazione al diritto a essere informato e a formare liberamente le proprie opinioni e idee, quale versante passivo della libertà di manifestazione del pensiero e ancor più della libertà di informazione, quanto in termini di "autodeterminazione individuale", ovvero di capacità di formare ed esprimere le proprie scelte¹¹.

Al riguardo, assume particolare rilievo il legame dell'autodeterminazione individuale con l'aspetto informativo, sostanziato nel "consenso informato" quale strumento attraverso cui si manifesta la scelta consapevole e autonoma del soggetto: tale profilo risulta saliente anche nella delineazione del "diritto all'autodeterminazione informativa", con cui si esprime l'istanza di controllo sulla circolazione delle informazioni e dei dati (innanzitutto di carattere personale) che riguardano il soggetto, divenuta componente fondamentale della loro esplicazione nella dimensione digitale¹².

Rispetto alla tensione dialettica che si genera in relazione al fenomeno manipolativo, va peraltro messo in luce come questo si accompagni a una serie di problemi definitori, che trovano soprattutto nel contributo della sociologia della comunicazione il tentativo di tracciarne confini meno incerti, rispetto ad altri concetti affini quali coercizione, persuasione o influenza, tutti accomunati

II/2025; N. Maccabiani, *Technological colonization and constitutional dysfunction*, Bollati Boringhieri, 2025; D. Martire, Human in the loop. *L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale*, in *Rivista italiana di informatica e diritto*, n. 2/2024; e, volendo, L. Durst, *La libertà individuale nel contesto digitale: profili costituzionali*, in M. Belletti, A. Buratti, L. Durst, P. Mezzanotte, M.G. Rodomonte (a cura di), *Diritti e Costituzione all'alba del terzo millennio. Liber Amicorum per Roberto Nania*, Giappichelli, 2025.

11. Sulla problematica possibilità di riconoscere un generale diritto all'autodeterminazione individuale, tuttora ampiamente discussa dalla dottrina costituzionalistica – propensa piuttosto a declinazioni "settoriali", suffragate dalla giurisprudenza costituzionale in materia – cfr. in particolare S. Mangiameli, *Autodeterminazione: diritto di spessore costituzionale?*, in *Forum di Quaderni costituzionali*, 2009; A. Spadaro, *I due volti del costituzionalismo di fronte al principio di autodeterminazione*, in *Politica del diritto*, n. 3/2014.

12. Cfr. S. Rodotà, *Il diritto di avere diritti*, Laterza, 2012; Id., *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Laterza, 2004; A. Papa, *La problematica tutela del diritto all'autodeterminazione informativa nella big data society*, in *Liber amicorum per Pasquale Costanzo*, *Consulta Online*, 17 aprile 2020.

dalla caratteristica di costituire pur diverse strategie per spingere un soggetto verso un certo comportamento o azione¹³.

Più specificamente, infatti, la *coercizione* si connota come intervento *esterno*, che incide sulla possibilità oggettiva di agire del soggetto, diminuendo la possibilità di scelte alternative (ad esempio, la riduzione del pluralismo partitico) anche tramite la minaccia di punizioni o sanzioni, o la revoca di benefici o vantaggi; l'*influenza* costituisce invece un concetto di portata più ampia e varia, che può riferirsi anche a processi di carattere *involontario*, ricomprendendo ad esempio forme imitative o mode; la *persuasione* infine appare una forma di influenza che si caratterizza per l'essere esercitata adducendo "buone ragioni" a sostegno di una posizione: richiede pertanto un'adesione ideale, *razionale*, basata sulla capacità di giudizio e consenso *consapevole* attraverso cui si mantiene la possibilità di critica delle opinioni o della verità, e che tuttavia può risentire di altri fattori quali asimmetrie informative o dimensioni affettive e pressioni latenti¹⁴. Va a ogni modo rilevato come, definita in questi termini, la persuasione appaia pienamente connaturata alla libertà di manifestazione del pensiero, proprio in quanto libertà di persuadere e convincere gli altri¹⁵.

Premesse tali distinzioni idealtipiche, la *manipolazione* si distinguerebbe allora sia dall'intervento *esterno* che caratterizza la coercizione, sia dall'adesione *razionale* e consensuale che caratterizza la persuasione, in quanto mirerebbe, piuttosto, a condizionare comportamenti, idee e atteggiamenti dell'altro soggetto destinatario della comunicazione senza che quest'ultimo ne sia pienamente o effettivamente consapevole. Connotandosi per tale modalità di intervento, di tipo interno e soggettivo, la manipolazione è invero ritenuta dalla psicologia sociale un aspetto costitutivo delle relazioni interpersonali e sociali; viceversa, secondo le teorie della comunicazione di massa, nel "sospetto di manipolazione" si disvelerebbe piuttosto il vero fine e interesse ulteriore – al di là della mera comunicazione – dei mass media, una sorta di "peccato originale" che si pone alla base della comunicazione di massa¹⁶.

Alla luce di quanto detto, emergono due aggiuntivi aspetti di interesse. Pri-

13. Si v. la sistematizzazione di G. Gili, *Il problema della manipolazione. Peccato originale dei media?*, FrancoAngeli, 2009, pp. 11 ss.; C.R. Sunstein, *Fifty Shades of Manipulation* (February 18, 2015), in *Journal of Marketing Behaviour*, n. 1/2016. Con riguardo all'ampiezza semantica del concetto di manipolazione, v. anche *Manipulation*, in *Cahiers internationaux de symbolisme*, nn. 167-168-169/2024.

14. Cfr. G. Gili, *Il problema della manipolazione*, cit.

15. Tale aspetto è tradizionalmente messo in rilievo da risalente dottrina, v. in particolare A.C. Jemolo, *I problemi pratici delle libertà*, Giuffrè, 1972 (p. 47); P. Barile, *La libertà di espressione del pensiero*, Giuffrè, 1975 (p. 227).

16. G. Gili, *Il problema della manipolazione*, cit., pp. 12-14.

mo, a ulteriore conferma delle difficoltà definitorie, può richiamarsi il concetto di *propaganda*, che risulta oscillare tra le due forme della persuasione e della manipolazione (si pensi ad esempio alla “propaganda occulta”), e trova pertanto ulteriore articolazione nell’individuazione di forme lecite e illecite, graduate attraverso la classificazione di propaganda bianca, grigia o nera¹⁷. Il secondo, lo stretto rapporto che lega manipolazione e propaganda politica all’avvento della società di massa e al ruolo e al potere dei mezzi di comunicazione di fronte al problema dell’integrazione sociale, in un contesto non più omogeneo come quello della società liberal-borghese¹⁸.

Preme qui mettere in rilievo, attraverso il richiamo a questa corrente di studi, come, soprattutto da Lippmann in poi¹⁹, si sia mostrato che la rappresentazione della realtà, in ogni società in cui non è possibile la conoscenza diretta di ciò che accade, produca dei «modelli di realtà», ovvero una rappresentazione fabbricata dell’ambiente, che influenza i comportamenti dei consociati a partire dalla loro “premessa cognitiva”. In altri termini, la propaganda agisce su interpretazioni della realtà, determinando un fraintendimento voluto, ottenuto tramite la manipolazione della notizia in vista di un determinato fine, e avvalendosi di consolidate e note tecniche quali la censura o la distorsione e falsificazione delle notizie. In questo aspetto si rinviene dunque uno dei grandi problemi della democrazia contemporanea: ovvero nella possibilità dei mass media di dare forma al mondo («ambiente mentale») tramite la manipolazione strategica delle immagini del mondo stesso («equivoco cognitivo»), in una partita che si gioca in misura rilevante sul piano della dimensione cognitiva e mentale, e nell’affinamento dell’uso della manipolazione come strategia persuasoria che agisce sotto il livello della consapevolezza²⁰. Una manipolazione dunque operata in maniera intenzionale e occulta, che punta alla “eterodirezione” dei soggetti attraverso la conformità e la ricerca del consenso, e che risulta tanto più efficace se l’emittente detiene un forte controllo sulle informazioni

17. C. Giunta, *L’espressione del pensiero politico: comunicazione, propaganda, pubblicità*, in *federalismi.it*, n. 14, 2025.

18. Cfr. ancora G. Gili, *Il problema della manipolazione*, cit. Sul tema della propaganda politico-elettorale, F. Lanchester, *Propaganda elettorale*, in *Enc. Dir.*, 1988.

19. Oltre al contributo di W. Lippmann, *L’opinione pubblica* (1922), Edizioni di Comunità, 1963, si v. ad esempio E.L. Bernays, *Propaganda. Come manipolare l’opinione pubblica* (1928), ShaKe, 2020, che mostrano fra l’altro la maggior efficacia della manipolazione conseguita tramite l’apparente neutralità dei messaggi (occulti); Si v. Packard, *I persuasori occulti* (1957), Einaudi, 2015, sul contributo delle scienze psicologiche alla creazione di nuovi bisogni funzionali alla società capitalistica e all’industria dei consumi di massa; E. Katz, P.F. Lazarsfeld, *L’influenza personale in comunicazione* (1955), Armando Editore, 2012.

20. Cfr. ancora la ricostruzione di G. Gili, *op. ult. cit.*

e il ricevente non è in grado di proteggersi a causa dell'impoverimento dei legami di gruppo, ovvero per la riduzione di quel pluralismo sociale che rischia di tracciare la rotta verso ciò che recenti studi sull'odierno sistema delle piattaforme hanno cominciato a inquadrare come "capitalismo della solitudine"²¹.

Queste capacità di persuasione sono state ampiamente approfondite dalla dottrina già con riguardo all'uso del mezzo televisivo, contribuendo a mettere in evidenza i caratteri del nuovo "*homo videns*" e l'impatto dell'immagine sui comportamenti e sulle emozioni, al di là del filtro del pensiero – come rilevato anche, si ricorderà, dalla giurisprudenza costituzionale, proprio al fine di giustificare la disciplina differenziata tra emittenza privata a livello locale e monopolio statale per la diffusione nazionale²². Problematiche che hanno determinato inoltre una maggior attenzione alla tutela della sfera inconscia della formazione del pensiero del destinatario, tradottasi in specifici divieti di messaggi (pubblicitari) subliminali e di programmi nocivi per lo sviluppo psichico o morale dei minori²³, e che hanno posto in evidenza, al contempo, la più generale questione della estensione della protezione dell'art. 21 Cost. anche alla manifestazione dei contenuti psichici (inclusi quelli destinati a suscitare pura emozione), laddove la distorta rappresentazione dei fatti finisce per rilevare non solo o non tanto quale problema di per sé, ma quale potenziale rischio di una «schiavitù mentale» che incide sul processo di formazione del pensiero del destinatario quale ambito della identità individuale e della cosciente autodeterminazione personale²⁴.

21. Sul circolo vizioso alimentato dal "capitalismo della solitudine" nell'utilizzo dei social network e delle piattaforme, v. lo studio per conto della Fondazione Jean Jaurès condotto da P. Klotz, *Le capitalisme de la solitude à l'assaut des liens humains*, 29/08/2025, disponibile su: www.jean-jaures.com.

22. Cfr. Corte cost., sent. n. 148 del 1981, punto 3 cons. dir., in cui la Corte richiama «quelli che sono i dati caratteristici del mezzo di diffusione del pensiero in esame che, per la sua notoria capacità di immediata e capillare penetrazione nell'ambito sociale attraverso la diffusione nell'interno delle abitazioni e per la forza suggestiva della immagine unita alla parola, dispiega una peculiare capacità di persuasione e di incidenza sulla formazione dell'opinione pubblica nonché sugli indirizzi socio-culturali, di natura ben diversa da quella attribuibile alla stampa» (corsivi nostri).

23. Confluiti dapprima nella legge n. 223/1990, in attuazione della direttiva CEE 89/552, poi nella legge n. 112 del 2004.

24. Su tali profili, v. le riflessioni di C. Pinelli, *Suggestione dei messaggi televisivi, persuasione e manifestazione del pensiero*, in *Dir. Inf.*, 1995, che, riprendendo anche la storica sentenza della Corte costituzionale sulla caducazione del reato di plagio (n. 96 del 1981) – motivata in ragione del contrasto tra la fattispecie impugnata e le esigenze di tassatività e determinatezza della legge penale, data la eccessiva incertezza dell'esatto confine tra suggestione e persuasione – suggerisce come tale margine possa invece essere diversamente preso in considerazione nell'ambito della tutela della libertà di informazione e manifestazione del pensiero.

3. Manipolazione e profilazione negli interventi regolatori dell'UE

Le dinamiche di funzionamento caratterizzanti la manipolazione, qui sopra esposte, trovano dunque, come detto, ulteriori potenzialità di applicazione nel contesto delle nuove tecnologie digitali, del mondo di internet e delle piattaforme, dove l'intervento del legislatore europeo si è distinto proprio in funzione dell'obiettivo di garantire adeguati strumenti di protezione e difesa dei diritti fondamentali dei soggetti anche nell'ecosistema virtuale, finendo per connotare un vero e proprio modello "*fundamental rights-driven*"²⁵, strettamente correlato alla riconfigurazione dello stesso costituzionalismo nei termini di un "costituzionalismo digitale"²⁶.

In tale ambito occorre preliminarmente muovere dal tema della *profilazione*, che gioca un ruolo basilare nelle dinamiche di funzionamento delle piattaforme digitali (*targeting*) e traccia un raccordo essenziale tra "manipolazione informativa" e "manipolazione algoritmica".

Si può infatti constatare come, nell'attuale quadro regolatorio europeo, si assista alla spinta verso una crescente tutela dai rischi derivanti dalla profilazione, rispetto a un approccio in precedenza maggiormente "bloccato" nella tensione, e ricerca di un equilibrio, tra protezione dei diritti fondamentali (soprattutto quelli relativi alla tutela della personalità) ed esigenze e imperativi del mercato, che prosperano sul loro sfruttamento²⁷.

A tale riguardo è innanzitutto necessario prendere in considerazione la disciplina tracciata dal Gdpr, da cui come è stato correttamente osservato si ricavano, in sostanza, tutti i "principi costituzionali" che regolano la società digitale²⁸. Nel Regolamento generale per la protezione dei dati personali non si

25. Per la tripartizione dei modelli di *governance* del digitale fra UE (*rights-driven*), Usa (*market-driven*) e Cina (*State-driven*), v. l'ormai classico riferimento ad A. Bradford, *Digital Empires: The Global Battle to Regulate Technology*, Oxford University Press, 2023.

26. Su tale nozione, pur dibattuta in dottrina, con particolare riguardo alla sua connotazione in senso descrittivo o prescrittivo, v. G. De Gregorio, *Digital constitutionalism in Europe: reframing rights and powers in the algorithmic society*, Cambridge University Press, 2022; O. Pollicino, *Di cosa parliamo quando parliamo di costituzionalismo digitale?*, in *Quaderni costituzionali*, n. 3/2023; Id., *Costituzionalismo digitale. Pensare la democrazia al tempo dell'IA*, il Mulino, 2025.

27. Così osserva N. Maccabiani, *Technological colonization and constitutional dysfunction*, cit., p. 86.

28. Si tratta in particolare dei principi di *accountability* (responsabilizzazione dei soggetti che trattano dati/servizi), dell'approccio *by-design* e *by-default*, del principio di trasparenza, della valutazione di impatto sui diritti fondamentali e del rischio, della supervisione umana (o *human in the loop*) e dei diversi diritti riconosciuti al soggetto. Sia consentito al riguardo il rinvio anche a L. Durst, *Diritto e predittività. Spunti introduttivi in tema di sistemi decisori automa-*

rinvengono invero richiami espliciti a forme di manipolazione, mentre assume cruciale rilievo il riferimento alla profilazione, definita dall'art. 4 quale capacità di fornire analisi "predittive" a partire dall'elaborazione di dati e caratteristiche personali²⁹, che come noto si trovano alla base delle principali tecniche impiegate per lo sviluppo dell'economia delle piattaforme, quali ad esempio il *neuromarketing*, il *microtargeting* politico, l'*hyper-nudging* (cioè le architetture di scelta), l'*affective computing* (informatica affettiva o emozionale)³⁰.

Proprio con riguardo alla profilazione e alla possibilità di produrre decisioni "automatizzate" grazie all'elaborazione algoritmica viene in considerazione una disposizione-chiave del Gdpr, che ha rapidamente trascorso l'ambito di stretta pertinenza della protezione dei dati personali per assumere una portata parametrica per l'allora solo incipiente uso dell'IA e delle tecnologie di *machine learning*, ovvero l'art. 22 (par. 1), ai sensi del quale «l'interessato ha diritto di non essere sottoposto a una decisione basata *unicamente* sul trattamento automatizzato, compresa la *profilazione*, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona» (corsivi nostri). È da questa disposizione che è possibile infatti dedurre il c.d. "principio di non esclusività" della decisione algoritmica, ossia il divieto di essere sottoposto a una decisione automatizzata che produce effetti giuridici nella sfera del singolo, a meno che non siano rispettate specifiche garanzie, fra cui il consenso esplicito dello stesso, che si traduce nel principio di trasparenza e nel diritto alla conoscibilità dell'interazione uomo-macchina, nonché nella c.d. "riserva di umanità", ovvero la possibilità di richiedere sempre un intervento umano in tale tipo di processi decisionali³¹. Proprio il consenso pare costituire un principio cardine per l'eventuale trattamento automatizzato e la profilazione, soprattutto nei casi in cui sia condotta con metodi particolarmente intrusivi (ad esempio, tramite combinazione di più fonti di dati senza che il soggetto ne sia a conoscenza).

Si tratta di un aspetto di rilievo nell'architettura dell'ecosistema digitale e

tizzati, in F. Fabrizzi, L. Durst (a cura di), *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era dell'algoritmo*, Editoriale Scientifica, 2024, p. 29 ss.

29. Più esattamente, la definizione di "profilazione" fornita dall'art. 4 è la seguente: «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica». Cfr. sul tema anche F. Lagioia, G. Sartor, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *federalismi.it*, 11/2020.

30. N. Maccabiani, *Technological colonization and constitutional dysfunction*, cit., p. 86 ss.

31. Su tali profili, v. anche i diversi contributi in F. Fabrizzi, L. Durst (a cura di), *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era dell'algoritmo*, cit.

dei suoi presupposti, non privo di delicate ambiguità. Pur non potendosi in questa sede approfondire in dettaglio la disciplina, non si può non fare cenno ai problematici profili di “fallacia del consenso”, che emergono soprattutto in relazione alle condizioni di validità del *libero* consenso prestato dall’utente. Tale requisito di libertà si può infatti ritenere mancante in tutti quei numerosi casi in cui il soggetto non sia realmente informato e pienamente consapevole della scelta effettuata, ovvero nei casi di “condizionalità” (in cui manchi una “alternativa equivalente” al consenso al trattamento dei dati personali per poter usufruire del servizio), o nei casi in cui le interfacce e gli strumenti tecnici siano disegnati o adoperati in modo da indurre il consenso stesso.

Tali aspetti sono stati ampiamente messi in luce e dibattuti dalla dottrina e dalla giurisprudenza – anche della Corte di Giustizia – nonché divenuti oggetto di pareri e linee guida delle autorità garanti, sia a livello nazionale sia europeo³², e in parte fronteggiati dallo stesso impianto del Gdpr, laddove si prevede il rafforzamento dei principi di trasparenza (informativa dei soggetti) e di *accountability* dei titolari del trattamento, anche attraverso l’impiego di misure tecniche adeguate (in ossequio ai criteri di *privacy by design* e *privacy by default*, ulteriormente comprovati dalla eventuale valutazione di impatto sulla privacy).

Ciononostante, è proprio sul presupposto che il consenso possa costituire efficace espressione di autodeterminazione informativa e di effettiva capacità di scelta del soggetto che si definiscono alcune zone grigie particolarmente permeabili a usi “distorti” della profilazione (e della *targetizzazione* che a questa è correlata), e su cui tendono a focalizzarsi tanto le possibilità di c.d. “manipolazione algoritmica” quanto i rimedi normativi previsti dai nuovi regolamenti UE.

3.1. *Consenso, profilazione e manipolazione: profili di tutela nel Dsa e nel Dma*

Ampliando dunque il quadro ai principali ambiti di applicazione della regolazione del mercato digitale e delle piattaforme, può notarsi come il rapporto fra *consenso* e *profilazione* – e specificamente in riferimento alla manipolazio-

32. Si v. ad esempio nella giurisprudenza della Cgue i noti casi Planet 49 (C-673/17), Fashion ID (C-40/17) e Meta Platform (C-252/21); nonché WP29, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679* (Versione emendata e adottata in data 6 febbraio 2018); Edpb, *Parere 8/2024 sul consenso valido nel contesto dei modelli “consenso o pagamento” attuati dalle piattaforme online di grandi dimensioni*, adottato il 17 aprile 2024; Garante Privacy, *Linee guida cookie e altri strumenti di tracciamento*, 10 giugno 2021. In dottrina, v. già le riflessioni di S. Rodotà, *Tecnologie e diritti*, il Mulino, 1995.

ne – rivista centrale rilievo anche nel Digital Markets Act (Dma) e nel Digital Services Act (Dsa), che in linea con l'intento di costruire un quadro coerente e sistematico della disciplina applicabile al digitale rimandano espressamente al Gdpr per le definizioni e i requisiti di entrambi i termini.

Nel Dma, che come noto detta la normativa finalizzata a garantire l'equa competitività sul mercato digitale, con particolare riguardo a quelle piattaforme di grandi dimensioni che svolgono un ruolo di *gatekeepers* per la fornitura e controllo di servizi di piattaforme di base, vengono prescritti importanti obblighi di trasparenza, fra cui la descrizione delle tecniche di profilazione impiegate da parte dei *gatekeepers* (art. 15 e considerando 72), e la necessità del consenso esplicito dell'utente per l'utilizzo incrociato dei dati personali provenienti dal pertinente servizio di piattaforma di base in altri servizi forniti separatamente, o per la loro combinazione con altri dati personali provenienti da altri servizi dei *gatekeepers* o terzi, altrimenti vietato (art. 5). A tale riguardo si legge anche, nei considerando, con riferimento ai profili manipolatori, che revocare il consenso dovrebbe essere tanto semplice quanto prestare il consenso, e che i «gatekeeper non dovrebbero progettare, organizzare o gestire le loro interfacce online in modo tale da ingannare, *manipolare* ovvero compromettere o falsare in altro modo, in misura rilevante, la capacità degli utenti finali di prestare liberamente il proprio consenso» (considerando n. 37). Tali prescrizioni intendono dunque intercettare uno degli aspetti più problematici del contesto digitale, in relazione all'asimmetria nel rapporto fra utente e grandi piattaforme che si traduce anche nella perdita di controllo dei dati rispetto alla pluralità di contesti e servizi in cui possono intrecciarsi: da un lato si assiste al potenziamento delle capacità di profilazione, targetizzazione e manipolazione dei soggetti attraverso le posizioni dominanti delle grandi piattaforme-*gatekeeper*, e dall'altro si confida che lo strumento consensuale possa pienamente assolvere alle funzioni di espressione di una scelta e volontà libera e consapevole dell'utente anche in questi contesti³³.

Ancor più il regolamento sui servizi digitali, che mira alla maggior sicurezza e trasparenza dell'ambiente delle piattaforme online (anche attraverso il contrasto di contenuti illegali, della disinformazione online e di altri rischi per la società derivanti dai prestatori di servizi intermediari), prescrive alcune misure specifiche improntate alla trasparenza, quali l'obbligo di informazioni

33. Al fine, inoltre, di non compromettere in modo sleale la contendibilità dei servizi di piattaforma di base i *gatekeeper* sono tenuti a consentire agli utenti finali di scegliere liberamente di seguire le pratiche di trattamento dei dati e accesso con registrazione offrendo un'alternativa meno personalizzata ma equivalente, e senza subordinare l'utilizzo del servizio di piattaforma di base o di talune sue funzionalità al consenso dell'utente finale (considerando n. 36 e art. 5).

chiare e identificabili sulle pubblicità online e sull'uso a tali fini della profilazione dei destinatari, comunque vietata se concernente trattamento di dati sensibili (art. 26) e minori (art. 28). Inoltre, per quanto riguarda l'uso di sistemi di raccomandazione, il rispetto della trasparenza impone un linguaggio chiaro e intellegibile, l'indicazione dei principali parametri utilizzati nei sistemi di raccomandazione e delle opzioni a disposizione dei destinatari del servizio per modificare o influenzare tali parametri principali (art. 27), nonché di assicurare una alternativa per l'uso del servizio non basata sulla profilazione nel caso di sistemi di raccomandazione utilizzati da fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi (art. 38).

Come noto, infatti, il Dsa prevede una serie di obblighi aggiuntivi per le piattaforme e i motori di ricerca di grandi dimensioni (Vlopse) in relazione ai «rischi sistemici» derivanti dalla progettazione, dal funzionamento o dall'uso dei loro servizi, rischi quali in particolare la diffusione di contenuti illegali e gli eventuali effetti negativi sia per i diritti fondamentali e la dignità umana (anche in relazione a violenza di genere, minori, salute pubblica e benessere psico-fisico), sia sul dibattito civico, i processi elettorali, la sicurezza pubblica.

In tale ambito, con riguardo alla valutazione dei rischi sistemici (art. 34) si prescrive anche di analizzare se e in che modo siano influenzati dalla “manipolazione intenzionale” del servizio, ad esempio mediante l'uso non autentico o lo sfruttamento automatizzato, o l'amplificazione e la diffusione potenzialmente rapida e ampia di contenuti illegali e di informazioni incompatibili con le condizioni generali, come nel caso di utilizzo di bot o account falsi o altri usi ingannevoli³⁴. Con riguardo poi alle misure di attenuazione degli stessi rischi sistemici (art. 35), che includono anche l'adeguamento della progettazione delle interfacce e delle procedure di moderazione dei contenuti, si prevede la apposizione di un contrassegno ben visibile di contenuti generati o “manipolati”, siano essi un elemento di un'informazione, un'immagine, un contenuto audio o video: ossia specificamente nei confronti dei c.d. *deepfakes*.

Un ulteriore aspetto della manipolazione, di tipo tecnico o progettuale (come richiamato anche nel citato considerando 37 del Dma), è quello esplicitamente proibito dall'art. 25 del Dsa, che vieta ai fornitori di piattaforme online di progettare, organizzare o gestire le loro interfacce online in modo tale da ingannare o *manipolare* i destinatari dei loro servizi o da materialmente falsare o compromettere altrimenti la capacità dei destinatari di prendere decisioni libere e informate: la disposizione è orientata a intercettare i c.d. “*dark patterns*”,

34. Cfr. in particolare quanto specificato nel considerando 84, nonché in generale il considerando 80 e successivi per l'interpretazione dei rischi sistemici.

ovvero il design di interfacce finalizzato intenzionalmente a indurre gli utenti a fare determinate scelte, spesso contrarie ai propri interessi e potenzialmente dannose, ma vantaggiose per le piattaforme di social media (per esempio, abbonarsi a un servizio senza rendersene conto, in analogia con alcune caratteristiche proprie della normativa sulla pubblicità ingannevole): fenomeno cui già l'Edpb aveva dedicato delle linee guida specifiche nel 2022³⁵.

D'altronde, la consapevolezza del legislatore europeo in merito alla problematicità di tali questioni e agli effetti di disinformazione prodotti dalla manipolazione basata su profilazione e *targeting* è espressa chiaramente nel considerando 69, ove si legge: «In alcuni casi, le tecniche di *manipolazione* possono avere un impatto negativo su interi gruppi e amplificare i danni per la società, ad esempio contribuendo a campagne di disinformazione o discriminando determinati gruppi. Le piattaforme online sono ambienti particolarmente sensibili per tali pratiche e presentano un rischio per la società più elevato. Di conseguenza, i fornitori di piattaforme online non dovrebbero presentare inserzioni pubblicitarie basate sulla profilazione».

3.2. *La manipolazione delle informazioni tra pluralismo e profilazione nel Regolamento sulla libertà dei media e nel Regolamento sulla pubblicità politica*

Quanto alla vera e propria “manipolazione delle informazioni”, abbinata in endiadi alla disinformazione, questa costituisce specifico ambito di intervento per due nuovi regolamenti di settore: lo *European Media Freedom Act* (Emfa)³⁶, che mira a proteggere la libertà e il pluralismo dei media con particolare attenzione ai rischi di pressioni indebite e allo spazio mediatico digitale, prevedendo misure per rafforzare e garantire la trasparenza della proprietà dei media, la protezione delle fonti giornalistiche e dell'indipendenza editoriale, la tutela dei

35. Edpb, *Linee guida 3/2022. Modelli di progettazione ingannevoli nelle interfacce delle piattaforme di social media: come riconoscerli ed evitarli* (adottate il 14 febbraio 2023). Sulla differenza tra *dark patterns* – definiti anche “*sludge*” con particolare riguardo alle tecniche volte a creare ostacoli alla scelta – e “*nudge*”, basata proprio sulla diversa finalità, tra il vantaggio che il *nudge* intenderebbe apportare alla collettività e all'individuo senza ridurne la libertà di scelta, e quello che il *dark pattern* invece intende far conseguire al business, attraverso il condizionamento ingannevole della scelta dell'utente, oltre all'ormai classico C. Sunstein, R. Thaler, *Nudge. La spinta gentile*, Feltrinelli, 2009, v. anche R. Thaler, *Nudge not Sludge*, in *Science*, vol. 361, 2018; S. Mills, *Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications*, in *Behavioural Public Policy*, 2023, vol. 7, n. 2; H. Brignull, *Deceptive Patterns. Exposing the tricks tech companies use to control you*, 2023; M. Leiser, *Dark Patterns, Deceptive Design and the Law*, Oxford University Press, 2025.

36. Regolamento UE 2024/2065.

media contro la rimozione ingiustificata di contenuti online da parte di grandi piattaforme; e il *Regolamento sul targeting e la pubblicità politica* (Rtpp)³⁷, che costituisce una sorta di *lex specialis* nel quadro del Dsa, specificamente mirata a contrastare le forme non trasparenti e riconoscibili di pubblicità politica, ancor più laddove implichino l'uso di tecniche di profilazione e targetizzazione dei contenuti (in armonia, anche in questo caso, con la disciplina dettata dal Gdpr).

Partendo dal primo, nel regolamento sulla libertà dei media si evidenzia, tra i considerando iniziali, come il buon funzionamento del mercato interno dei servizi di media sia compromesso da vari fattori, fra cui i «fornitori, compresi quelli controllati da determinati paesi terzi, che si dedicano in modo sistematico ad attività di disinformazione, o *manipolazione* delle informazioni e ingerenze, e sfruttano le libertà del mercato interno a fini abusivi, ostacolando in tal modo il corretto funzionamento delle dinamiche di mercato» (cons. 4), e come proprio a tal fine sia necessario rafforzare la cooperazione normativa (cons. 6) e i servizi di media di qualità, che «costituiscono anche un antidoto alla disinformazione e alla *manipolazione* delle informazioni e alle ingerenze da parte di attori stranieri» (cons. 14). Soprattutto nel caso di Vlopose, il considerando 53 sottolinea l'esigenza di fornire una funzionalità sulla propria interfaccia online per consentire ai fornitori di servizi di media di dichiarare che soddisfano determinati requisiti, con la possibilità per le organizzazioni della società civile riconosciute e quelle di *fact-checking* di segnalare alle piattaforme online di dimensioni molto grandi qualsiasi eventuale violazione, nonché la necessità di orientamenti pubblicati dalla Commissione sulla prevenzione del rischio di abusi di tale funzionalità, in particolare da parte di fornitori che si dedichino in modo sistematico ad attività di disinformazione, *manipolazione* delle informazioni e ingerenze, compresi quelli controllati da determinati paesi terzi.

Si prevedono pertanto, in particolare, due rimedi: lo strumento del c.d. “dialogo strutturato” tra fornitori di piattaforme online di dimensioni molto grandi, rappresentanti di fornitori di servizi di media e rappresentanti della società civile (art. 19), regolarmente organizzato dal Comitato europeo per i servizi di media, al fine di monitorare l'adesione alle iniziative di autoregolamentazione volte a proteggere gli utenti da contenuti nocivi come la disinformazione o la manipolazione delle informazioni e le ingerenze da parte di attori stranieri; nonché l'esercizio di monitoraggio indipendente e costante del mercato interno dei servizi di media (art. 26) anche con specifico riguardo all'analisi dettagliata dei mercati dei media di tutti gli Stati membri, al livello di concentrazione dei media e ai rischi di manipolazione delle informazioni e di ingerenze da parte di attori stranieri.

37. Regolamento UE 2024/900.

Il rischio di manipolazione e di ingerenze straniere è al centro anche del regolamento relativo alla trasparenza e al *targeting* della pubblicità politica, in quanto ritenuto una delle sfide più importanti dell'UE e degli Stati membri: come riferito dal considerando 4, la pubblicità politica può infatti essere un «vettore di disinformazione, specie se non ne è esplicitata la natura politica, se proviene da sponsor esterni all'Unione o se è oggetto di tecniche di *targeting* o tecniche di consegna dei messaggi pubblicitari», dal che discende la necessità di garantire la trasparenza quale principio cardine del regolamento per iniziative a contrasto di tali fenomeni, «anche per sostenere un dibattito politico e campagne politiche equi e aperti, come pure elezioni o referendum liberi e regolari, e per combattere la *manipolazione* dell'informazione e le interferenze, nonché le interferenze illecite anche da paesi terzi» (cons. 4). Al fine specifico di contrastare la manipolazione delle informazioni e le interferenze nella pubblicità politica, le «piattaforme online» sono incoraggiate dunque, anche attraverso il codice di condotta sulla disinformazione e il Dsa, «a stabilire e attuare politiche mirate e altre misure pertinenti, anche mediante la partecipazione a iniziative più vaste di demonetizzazione della disinformazione, per evitare la collocazione di messaggi di pubblicità politica contenente disinformazione» (cons. 20).

Questo aspetto si completa e compenetra inoltre, nel regolamento sulla pubblicità politica, anche con l'utilizzo delle tecniche di targetizzazione, inteso quale trattamento di dati personali cui si imputa «un impatto specifico e negativo sui diritti e sulle libertà fondamentali delle persone, come il diritto di essere trattate in modo equo e paritario, non essere manipolate, ricevere informazioni obiettive, farsi un'opinione, prendere decisioni politiche ed esercitare il diritto di voto», e a cui si attribuisce di incidere negativamente sul processo democratico, conducendo a una «frammentazione del dibattito pubblico su importanti questioni sociali, una comunicazione selettiva e, in ultima analisi, la manipolazione dell'elettorato», con l'ulteriore rischio di diffusione della manipolazione delle informazioni e delle ingerenze straniere (cons. 74). Aspetti cui si accompagna, come già nel Dsa, la necessità di progettare, organizzare e gestire le interfacce in modo tale da non «ingannare, manipolare o altrimenti distorcere o compromettere in materia sostanziale la capacità dell'interessato di prestare liberamente il proprio consenso per tale finalità specifica» (cons. 80).

Proprio in considerazione di tali presupposti, ovvero che «la pubblicità ingannevole od occulta per scopi politici costituisce un rischio, in quanto influenza i meccanismi fondamentali che permettono il funzionamento della nostra società democratica» (cons. 74), si giustifica infatti l'esigenza, che costituisce specifico oggetto del regolamento, di dettare ulteriori limitazioni e condizioni rispetto a quelle stabilite dal Gdpr, in particolare in tema di *targeting*

e profilazione, il c.d. “microtargeting manipolativo”: a questa esigenza sono finalizzate disposizioni nevralgiche del regolamento, in termini di obblighi di trasparenza (fra cui di individuazione, etichettatura, segnalazione dei servizi o dei messaggi di pubblicità politica, di cui agli art. 6 e successivi), nonché di divieto di *targeting* basato sulla profilazione di dati sensibili o laddove manchi il consenso esplicito e separato (al trattamento dei dati personali a fini di pubblicità politica) dell’interessato, cui si aggiungono il divieto di trattamenti di dati di minori almeno un anno al di sotto dell’età per l’esercizio del voto (art. 18), insieme agli obblighi di trasparenza addizionali previsti dall’art. 19 in ambito di pubblicità politica online che comporti il trattamento di dati personali³⁸.

3.3. Il contrasto alla “manipolazione algoritmica” e l’AI Act

Infine, è indubbiamente nell’AI Act³⁹ che i riferimenti alla manipolazione trovano il più ampio richiamo, proprio in quanto tali tecniche ed effetti risultano intrinsecamente legati ai rischi di utilizzo dei sistemi basati sull’impiego di algoritmi di *machine learning* e di sistemi di IA, con impatti potenzialmente significativi sull’autonomia di scelta dei soggetti e sui diritti fondamentali correlati, e pertanto contrari ai valori sui cui si fonda l’UE.

Sin da quanto enunciato nei *considerando*, si pone in particolare evidenza infatti come l’IA presenti, accanto a molti impieghi benefici, «la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per pratiche di *manipolazione*, sfruttamento e controllo sociale» (considerando n. 28)⁴⁰, con ancor più specifico cenno al fatto che «le tecniche di manipolazione basate sull’IA possono essere utilizzate per persuadere le persone ad adottare comportamenti indesiderati o per indurle con l’inganno a prendere decisioni in modo da sovvertirne e pregiudicarne l’autonomia, il processo decisionale e la libera scelta» (considerando n. 29)⁴¹. Questo anche in ragione

38. Quali l’elaborazione del documento di strategia interna, la conservazione di registri sulle tecniche di *targeting*, informazioni supplementari sulla logica e i parametri delle tecniche impiegate.

39. Regolamento UE 2024/1689.

40. Tali pratiche sono considerate «contrarie ai valori dell’Unione relativi al rispetto della dignità umana, alla libertà, all’uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori».

41. Il considerando n. 29 fa riferimento in particolare a sistemi di IA che «impiegano componenti subliminali quali stimoli audio, grafici e video che le persone non sono in grado di percepire poiché tali stimoli vanno al di là della percezione umana o altre tecniche manipolative o ingannevoli che sovvertono o pregiudicano l’autonomia, il processo decisionale o la libera scelta di una persona senza che sia consapevole di tali tecniche o, se ne è consapevole, senza che

degli sviluppi recenti dell'IA generativa, del cui potenziale di rischio il Regolamento, come noto, ha preso cognizione quasi *in extremis*, alla luce dell'immissione sul mercato digitale di ChatGpt nel novembre 2022, nella fase finale di approvazione dello stesso.

A tale riguardo, oltre agli obblighi di rilevamento e divulgazione degli output generati o manipolati artificialmente, in sinergia con quanto previsto dal Dsa per le piattaforme e i motori di ricerca di dimensioni molto grandi ai fini dell'attenuazione dei rischi sistemici (considerando n. 120 e 136), il considerando n. 133 sottolinea puntualmente la possibilità per diversi sistemi di IA di «generare grandi quantità di contenuti sintetici, che per gli esseri umani è divenuto sempre più difficile distinguere dai contenuti autentici e generati da esseri umani. L'ampia disponibilità e l'aumento delle capacità di tali sistemi hanno un impatto significativo sull'integrità e sulla fiducia nell'ecosistema dell'informazione, aumentando i nuovi rischi di cattiva informazione e *manipolazione* su vasta scala, frode, impersonificazione e inganno dei consumatori». Rimedi quali la previsione di obblighi di trasparenza ed etichettatura vanno a ogni modo bilanciati con i casi di espressione artistica e scientifica, fatte salve le tutele adeguate per i diritti e le libertà dei terzi (considerando n. 134).

Coerentemente con tali presupposti, è possibile innanzitutto rinvenire, nell'articolazione basata sul rischio adottata dall'AI Act, il posizionamento al massimo grado – tra i sistemi vietati o tra quelli a rischio elevato – dell'utilizzo di tecniche di tipo manipolativo⁴².

Innanzitutto, all'art. 5, lett. a, si proibiscono

l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo.

sia in grado di controllarle o resistervi o possa evitare l'inganno», incluse le interfacce cervello-computer o la realtà virtuale. Tali divieti sono complementari alla normativa in vigore per le pratiche commerciali sleali. Viceversa, non sono da considerare di per sé manipolatorie o dannose le pratiche lecite in quanto conformi al diritto applicabile nel contesto di trattamenti medici o nel settore della pubblicità commerciale.

42. Sull'AI Act, v. O. Pollicino *et al.*, *La disciplina dell'Intelligenza Artificiale*, Giuffrè, 2025; F. Pizzetti *et al.*, *La regolazione europea dell'intelligenza artificiale nella società digitale*, Giappichelli, 2025.

Come si può notare, è proprio l'aspetto subliminale che il legislatore europeo intende arginare, ovvero la manipolazione derivante dalla mancanza di informazione *consapevole*, caratterizzata dall'assunzione di una *decisione che non avrebbero altrimenti preso* e tale da cagionare, anche solo potenzialmente, un "danno significativo": un danno, in altri termini, di significativa gravità, che tuttavia non poggia su parametri puntualmente circostanziati, per quanto li considerando 29 indichi, in particolare, sistemi «aventi effetti negativi sufficientemente importanti sulla salute fisica, psicologica o sugli interessi finanziari», da apprezzare in primo luogo, secondo la logica di *accountability* che pervade anche l'approccio dell'AI Act, da parte dei fornitori e utilizzatori dei sistemi di IA.

Il problema dell'utilizzo di tecniche manipolatorie e subliminali⁴³ disvela infatti un nodo centrale, consistente nella difficoltà di determinare il grado di influenza accettabile del sistema impiegato, e di verificare unitamente a esso il nesso di causalità che possa dimostrare l'effettivo impatto del sistema sul comportamento del soggetto, tale da causare il danno o pregiudizio.

In proposito, può farsi riferimento alla problematica rappresentata dai «contenuti generati o manipolati artificialmente» (*deepfakes*)⁴⁴, disciplinati dall'art. 50, che prescrive misure di trasparenza e marcatura (come ad esempio il *watermarking*) al fine di identificare e notificare tali contenuti⁴⁵, che purtut-

43. Cfr. per un parallelo il divieto di "pubblicità subliminale" previsto nella Direttiva sui Servizi media audiovisivi, rivista con Direttiva (UE) 2018/1808 (art. 9).

44. Tra le definizioni di cui all'art. 3 del Regolamento sull'IA si intende per "deep fake" «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona».

45. L'art. 50 prevede obblighi di informazione e trasparenza sia per i fornitori sia per i deployer: in particolare, che «i fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici, garantiscono che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente» e che i deployer «di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un "deep fake" rendono noto che il contenuto è stato generato o manipolato artificialmente». Eccezioni a tali obblighi riguardano i casi in cui i sistemi di IA svolgono una funzione di assistenza per l'editing standard o non modificano in modo sostanziale i dati di input forniti dal deployer o la rispettiva semantica, o se gli usi sono autorizzati dalla legge ad accertare, prevenire, indagare o perseguire reati. Inoltre, nel caso di deep fake che siano parte di opere artistiche o creative gli obblighi di trasparenza sono applicati in modo da non ostacolare il godimento dell'opera, mentre nel caso di "manipolazione di testi" sono esclusi «se il contenuto generato dall'IA è stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione del contenuto», secondo una logica analoga all'art. 18 dell'Emfa, in merito al controllo umano anche nell'eventuale impiego di IA per la generazione di contenuti.

tavia non vengono proibiti in quanto non direttamente ricompresi tra i sistemi manipolativi ai sensi dell'art. 5: piuttosto, per quanto attiene all'IA generativa, il regolamento fa riferimento ai modelli di IA a finalità generali (*general purpose AI* – GpAI) che possono rappresentare un “rischio sistemico”, e che, in ragione di tale valutazione, possono eventualmente essere collocati tra i sistemi ad alto o basso rischio, con i conseguenti obblighi a carico dei *deployer* e dei *provider*.

Con particolare riguardo ai rischi sistemici di modelli di GpAI, l'art. 56 del Regolamento incoraggia inoltre l'adozione di codici di buone pratiche per la conformità agli obblighi previsti dal Regolamento (art. 53 e 55): il *Code of Practice for General Purpose AI Models* è stato pubblicato il 10 luglio 2025 e aperto alla firma degli aderenti, in linea con le scadenze previste per il Capo V del Regolamento (applicabile dal 2 agosto 2025, mentre le sanzioni saranno applicabili a partire dal 2026). Proprio nel capitolo del *Codice di buone pratiche* dedicato alla sicurezza (gli altri due capitoli riguardano trasparenza e diritto d'autore) si individua, tra i rischi sistemici specifici, la *harmful manipulation*, che consiste nei rischi derivanti dalla distorsione strategica del comportamento o delle convinzioni umane prendendo di mira ampie popolazioni o decisori di alto livello, attraverso la persuasione, l'inganno o il *targeting* personalizzato, che potrebbero pertanto minare i processi democratici e i diritti fondamentali⁴⁶.

Quanto ai sistemi ad alto rischio, i criteri di classificazione sono introdotti dall'art. 6, pur rimandando ai sistemi elencati nell'allegato III per consentire una maggiore elasticità di aggiornamento degli elenchi e la possibilità di valutazioni derogatorie, con il conseguente onere di provare che nel caso d'uso specifico il rischio sia tollerabile – ad esempio, nel caso in cui il sistema di IA non presenti un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale⁴⁷. Bisogna in proposito segnalare

46. «Harmful manipulation: Risks from enabling the strategic distortion of human behaviour or beliefs by targeting large populations or high-stakes decision-makers through persuasion, deception, or personalised targeting. This includes significantly enhancing capabilities for persuasion, deception, and personalised targeting, particularly through multi-turn interactions and where individuals are unaware of or cannot reasonably detect such influence. Such capabilities could undermine democratic processes and fundamental rights, including exploitation based on protected characteristics». A questa si affiancano, nell'appendice 1.4 del Codice, dedicata ai rischi sistemici specifici, anche il rischio chimico, biologico, radiologico e nucleare; il rischio di perdita di controllo sul modello; e il rischio di attacchi e reati informatici (*cyber offence*). Per una visione di sintesi, v. digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

47. L'art. 6, par. 3, specifica ulteriori condizioni dei sistemi di IA non considerate ad alto rischio, quali l'esecuzione di un compito procedurale limitato, il miglioramento del risultato di un'attività umana precedentemente completata, che il sistema non sia finalizzato a sostituire

come nel caso in cui effettui *profilazione* di persone fisiche il sistema (presente nell'allegato III) vada sempre considerato ad alto rischio: in altri termini, nei sistemi di IA di profilazione si presume sempre che sia "incorporato" un rischio per i diritti e le libertà fondamentali. Inoltre, fra le soluzioni tecniche previste per i sistemi ad alto rischio, al fine di conseguire un adeguato livello di accuratezza, robustezza e cibersecurity, l'art. 15 impone l'implementazione *by design* di misure volte a prevenire e controllare gli attacchi che cercano di manipolare il set di dati⁴⁸.

Appare in tal senso evidente come l'AI Act sia trasversalmente attraversato dalla preoccupazione legata ai problemi di condizionamento derivanti dall'utilizzo dei sistemi di IA nell'ambito dei processi decisionali umani (anche grazie all'ampio impiego di sistemi predittivi) che trovano in alcuni principi etico-giuridici fondamentali, quale quello di trasparenza ma anche di controllo umano, nonché in strumenti quali la valutazione di impatto sui diritti fondamentali (Fria) importanti presidi di garanzia per la mitigazione dei potenziali impatti discriminatori su singoli o gruppi.

Tuttavia, fino a che punto il quadro giuridico sin qui tratteggiato possa risultare efficace al fine di fornire uno strumento di tutela effettivo e sufficiente contro i rischi di manipolazione rimane ancora un interrogativo aperto, non solo in attesa dei tempi per la piena attuazione del regolamento sull'IA, ma anche a fronte dell'incessante innovazione delle applicazioni che si accompagnano allo sviluppo tecnologico: basti pensare all'IA generativa, alla sua incorporazione nei motori di ricerca o nelle varie app e social, e alle nuove sfide poste non solo al pluralismo informativo, ma anche alla stessa articolazione dell'ambiente cognitivo (inteso anche quale capacità di influenza mentale e relazionale)⁴⁹, che suggeriscono per un verso la necessità di intervenire anche

o influenzare la valutazione umana precedentemente completata senza un'adeguata revisione umana, o che il compito sia meramente preparatorio.

48. Più specificamente, «misure volte a prevenire, accertare, rispondere, risolvere e controllare gli attacchi che cercano di manipolare il set di dati di addestramento (*data poisoning* – "avvelenamento dei dati") o i componenti preaddestrati utilizzati nell'addestramento (*model poisoning* – "avvelenamento dei modelli"), gli input progettati in modo da far sì che il modello di IA commetta un errore (*adversarial examples* – "esempi antagonisti", o *model evasion* – "evasione dal modello"), gli attacchi alla riservatezza o i difetti del modello».

49. Desta preoccupazione ad esempio il crescente utilizzo, soprattutto tra i giovani, dell'IA come psicologo, oltre che in sostituzione di alcuni legami relazionali, legati alla "antropomorfizzazione" dei software, così come alcuni casi di suicidio giovanile "indotti" da algoritmi conversazionali: a tale riguardo, è stato notato come l'obbligo di adozione di misure tecniche ("safety-check") risulti spesso più utile all'esonero di responsabilità dei produttori che alla protezione degli utenti. Fra le pubblicazioni recenti sul tema, G. Scorza, *Diario di un chatbot sentimentale. Come le macchine ci imitano e ci manipolano*, Luiss University Press, 2025; A. Monti, *Lost in the*

attraverso azioni di educazione e alfabetizzazione informatica, al fine di accrescere la consapevolezza degli utenti sugli strumenti utilizzati, e per altro verso interrogano sull'opportunità di incidere attraverso normative più stringenti sulle limitazioni all'uso di sistemi di IA (in particolare nel caso di minori) – come in parte si rinviene nella legge italiana sull'IA⁵⁰ – o sulla conformazione stessa degli algoritmi⁵¹.

4. La necessità di un quadro di risposte integrato: lo scudo democratico dell'UE e i nuovi rischi cognitivi per l'autodeterminazione individuale

Nell'incedere quasi parallelo tra sviluppo tecnologico e risposta normativa, le questioni sin qui delineate mostrano come pur nella pluralità di ambiti attraversati dall'impiego degli algoritmi di IA, a fronte dei diversi rischi di manipolazione a questi riconducibili, si riveli necessaria una spinta alla convergenza del quadro regolatorio (oltre che alla effettività delle misure previste), ai fini di contrasto alla disinformazione, di tutela dell'integrità elettorale, di promozione del pensiero critico e della partecipazione civica.

Come è stato infatti correttamente osservato, le piattaforme digitali e gli algoritmi di IA hanno modificato l'ambiente cognitivo in cui si producono le rappresentazioni del mondo, operando non come canali neutri, ma come architetture dell'attenzione, in grado di filtrare, selezionare, amplificare o silenziare contenuti secondo logiche opache e proprietarie. Questo tipo di azione produce un cambiamento “profondo e silenzioso”, che incide sulla stessa democrazia non attraverso un'invasione esterna, ma una sua riconfigurazione inter-

Shell. Mind, Body, Identity and the Technology of Information, Routledge, 2026. Un test applicativo condotto sugli Llm è descritto da V. Ambriola, *Rischio suicidio, abbiamo testato quattro chatbot: ecco come si comportano*, in *Agenda digitale*, 10 ottobre 2025.

50. L. n. 132 del 2025, che include anche il sostegno a iniziative di alfabetizzazione nell'uso IA sulla base dell'assunto, inserito tra i principi generali della legge (art. 3), della centralità del rispetto dell'autonomia e del potere decisionale dell'uomo anche nei sistemi per finalità generali. Proprio le misure di penalizzazione previste dalla legge – una delle prime in Europa a integrazione dell'AI Act – costituiscono una delle principali novità della normativa italiana, che prevede, ad esempio, alcune modifiche al c.p. (art. 612-*quater*), per l'introduzione del nuovo reato di illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale.

51. Non si tratta solo della risposta alla tecnologia tramite la tecnologia stessa (*technology-based solutions* anche tramite Llm), quali i *guardrail* e i sistemi di sicurezza per il controllo di contenuti illegali, ma di vera e propria regolazione dell'architettura algoritmica utilizzata. In questa prospettiva, v. le conclusioni di R. Déchaux, *The Right to Free Elections in the TikTok Era: Towards Digital Expertise of the Venice Commission?*, cit., pp. 119 ss.

na, che cambia la grammatica del discorso, mutando le forme della partecipazione e le condizioni stesse della visibilità pubblica⁵².

È dunque muovendo dalla necessità di un quadro di risposte integrato a questa “trasformazione epocale” che l’UE ha posto al centro dell’iniziativa sullo “scudo democratico”, approvato il 12 novembre 2025, un approccio “whole-of-society”, inteso a coordinare all’interno della stessa strategia (nonostante i rischi di sovrapposizione normativa) i vari strumenti regolatori fin qui presi in esame, quali l’AI Act, il Dsa, il Regolamento sulla pubblicità politica, l’Emfa, unitamente a interventi di *media literacy* e allo sviluppo di un network europeo di *fact-checkers*, con un accresciuto ruolo della società civile⁵³.

Accanto a tale strategia, che si inserisce pienamente anche nel contesto di c.d. “guerra ibrida”, si stagliano tuttavia gli ulteriori interrogativi incentrati sul problema della formazione della scelta, della decisione, del consenso, dell’opinione attraverso l’ipotesi di definire nuove categorie o diritti per la difesa del “foro interiore”. Si avverte infatti la preoccupazione, se non l’urgenza, di fronte alla potenziale influenza delle tecnologie sulla dimensione cognitiva (c.d. “sesta dimensione”) di estendere l’ambito di tutela dei soggetti a quella fase di formazione “interna” delle determinazioni, pur tradizionalmente irrilevante per il fenomeno giuridico, in quanto collocata prima del momento della «espressione» dei comportamenti e delle scelte che formano oggetto del diritto. Si tratta in realtà di riflessioni che hanno destato l’interesse della dottrina costituzionalistica anche in tempi più risalenti – riportando alla mente quella giurisprudenza costituzionale che si era trovata a confrontarsi con questioni quali il reato di plagio o l’utilizzo di strumenti di prova quali la “macchina della verità”⁵⁴.

Alcuni nodi già prospettati nel dibattito di allora possono essere ripresi dunque anche nelle odierne riflessioni, con riguardo al contenuto e al fondamento delle posizioni soggettive considerate, che si uniscono agli interrogativi sulla necessità o meno di delineare “nuovi diritti” per lo spazio digitale o di po-

52. Così W. Quattrociochi, *Scudo europeo per la democrazia? Ma la disinformazione si combatte con la scienza*, in *AgendaDigitale*, 1° aprile 2025.

53. Cfr. CE, Comunicato stampa del 12 ottobre 2025, in cui si evidenziano i tre pilastri principali su cui è articolato lo scudo democratico europeo: 1) salvaguardia dell’integrità dello spazio dell’informazione; 2) rafforzamento delle istituzioni, libertà e regolarità delle elezioni, libertà e indipendenza dei media; 3) aumento della resilienza della società e della partecipazione dei cittadini, anche attraverso l’istituzione di un nuovo centro europeo per la resilienza democratica (ec.europa.eu/commission/presscorner/detail/it/ip_25_2660).

54. Cfr. anche R. Nania, *La libertà individuale nell’esperienza costituzionale italiana*, cit. Con particolare riguardo all’*excursus* della giurisprudenza costituzionale, A. Bonomi, *Le neuroscienze in rapporto alla libertà morale: aspetti di diritto costituzionale*, in *Forum di Quaderni Costituzionali*, 2018.

ter procedere a interpretazioni estensive di quelli contenuti nelle codificazioni tradizionali.

La dottrina al riguardo si è orientata su ipotesi differenti, a partire dalla riconsiderazione di un nuovo *habeas mentem*, secondo una linea di continuità dall'originario *habeas corpus* al più innovativo *habeas data*, e che muovendo dall'esigenza di protezione dell'integrità della sfera psichica sembra riaprire a quelle virtualità espansive della "libertà morale", oggetto dell'insuperato studio di Vassalli⁵⁵. Evidente l'intento di saldatura con un profilo da altri definito di "autonomia cognitiva"⁵⁶ che richiama in questo contesto all'autodeterminazione informativa (o digitale), quale aspetto di una "libertà di pensiero" non meramente coincidente con la sua manifestazione⁵⁷, secondo una prospettiva che trova peraltro traccia nelle distinte formulazioni della Cedu, e che è oggi rafforzata anche alla luce dei possibili impatti derivanti dalle potenzialità di applicazione delle neurotecnologie, e dalla loro combinazione con le scienze comportamentali e l'IA, rispetto a cui sono già stati configurati alcuni nuovi diritti specifici, i c.d. "neurodiritti"⁵⁸.

Si tratta dunque di configurazioni tali da far propendere, quantomeno, per una dimensione prevalentemente difensiva (negativa) dei diritti in gioco (da interferenze e, nello specifico, manipolazioni), che finisce per essere intesa in senso più o meno restrittivo a seconda della determinatezza dell'oggetto di

55. G. Vassalli, *Il diritto alla libertà morale. Contributo alla teoria dei diritti della personalità*, in *Studi giuridici in memoria di F. Vassalli*, Utet, 1960.

56. Cfr. D. Zolo, *Il tramonto della democrazia nell'era della globalizzazione*, in *Jura Gentium*, 2010, «Per autonomia cognitiva, come essenza stessa della libertà individuale, si può intendere la capacità del soggetto di controllare, filtrare e interpretare razionalmente le comunicazioni che riceve. Entro società informatizzate la garanzia giuridica dei diritti di libertà e dei diritti politici rischia di essere un guscio vuoto se non include l'«autonomia cognitiva»: se questa manca, è impensabile che si formi un'opinione pubblica indipendente rispetto ai processi di autolegittimazione promossi dalle élites politiche al potere. In presenza di una crescente efficacia persuasiva dei mezzi di comunicazione di massa il destino delle istituzioni politiche occidentali sembra dipendere dall'esito della battaglia a favore di questo fondamentale «diritto umano», l'«autonomia cognitiva», che potrebbe essere anche chiamato *habeas mentem*».

57. Cfr. N. Farahany, *Difendere il nostro cervello. La libertà di pensiero nell'era delle neurotecnologie*, cit.; J.-C. Bublitz, *My Mind is Mine!? Cognitive Liberty as a Legal Concept*, in E. Hildt, A. Francke (eds.), *Cognitive Enhancement*, Springer, 2013.

58. Fra questi in particolare sono stati individuati da M. Ienca, R. Adorno, *Towards new human rights in the age of neuroscience and neurotechnology*, in *Life Sciences, Society and Policy*, 2017, vol. 13, n. 5, il diritto alla privacy mentale, all'integrità mentale, alla continuità psicologica e alla libertà cognitiva o autodeterminazione mentale; secondo la proposta di R. Yuste et al., *Four ethical priorities for neurotechnologies and AI*, in *Nature*, 2017, si aggiungono alcuni diritti di carattere positivo-prestazionale, quali l'equo accesso al potenziamento mentale, e la protezione da bias algoritmici. Sul tema, v. anche A. D'Aloia, M.C. Errigo (a cura di), *Neuroscienze e diritto. Un punto di vista costituzionale*, Editoriale Scientifica, 2022.

protezione: dalla libertà di costruire un'opinione senza interferenze indebite, a una sorta di "libertà da inganni" che coinvolge la possibilità di definire veri e propri "diritti aletici" oltre che culturali⁵⁹; da una più circoscritta "privacy mentale"⁶⁰ a una più ampia ed estesa "sovranità cognitiva" degli individui⁶¹.

Nonostante la mancanza, per ora, di approdi epistemici definitivi, anche in ragione della difficoltà di procedere a distinzioni su aspetti scientifici ancora non del tutto univoci, tratto comune a tali percorsi di riflessione è quello di muoversi attraverso una nuova costellazione di diritti che combina i più tradizionali profili della privacy, dell'autodeterminazione informativa, dell'identità personale e della libertà di (manifestazione del) pensiero al fine di estendere gli strumenti di tutela giuridica anche ai più recenti rischi incidenti sulla dimensione cognitiva e dunque, di scelta e autodeterminazione individuale: la quale, per sua stessa natura, si ritiene non possa tuttavia prescindere da una strategia integrata di intervento, attraverso strumenti culturali ed educativi, come previsto dallo stesso progetto di scudo democratico europeo.

Se questi sono dunque gli scenari attuali, non può non tenersi presente come ci si trovi, ancora, in una fase di transizione e di comprensione delle potenzialità di queste tecnologie, indubbiamente soggetta a future trasformazioni, che potrebbero richiedere ben più profondi ripensamenti delle categorie giuridiche applicate, dei modelli di responsabilità e di tecnica regolatoria, pur sempre auspicabilmente orientati a non derogare agli irrinunciabili principi di tutela della persona umana approntati dalla tradizione del costituzionalismo, anche nella dimensione digitale.

59. M. Plutino, *Verità e opinione pubblica*, Rubbettino, 2024.

60. F. Cirillo, *Neuroprivacy. Problemi teorici e prospettive costituzionali*, Giappichelli, 2023.

61. Così nella dichiarazione del Consiglio d'Europa (CoE), *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, 13 febbraio 2019.

IL DISORDINE INFORMATIVO ONLINE: CLASSIFICARE LE CONDOTTE PER DEFINIRLE E PREVENIRLE GIURIDICAMENTE

Andrea Ruffo *

SOMMARIO: 1. Il fenomeno del disordine informativo – 2. La pericolosità per lo Stato – 3. La classificazione delle condotte disinformanti – 4. L'importanza della definizione giuridica delle condotte – 5. Il caso della misinformazione “indotta” – 6. La possibilità di prevenire giuridicamente gli effetti delle condotte disinformanti.

1. Il fenomeno del disordine informativo

Il vertiginoso sviluppo dei mezzi di comunicazione di massa, prodottosi nel XXI secolo grazie alla comparsa e alla costante implementazione della rete Internet, nonché delle tecnologie correlate, ha incrementato esponenzialmente le potenzialità e la multipolarità dell'informazione. La capillare diffusione e le prestazioni sempre crescenti della rete hanno comportato, da una parte il progressivo affiancamento (quand'anche non superamento) dei mezzi classici di comunicazione del Novecento (stampa, radio e televisione), dall'altra un'inesauribile spinta allo sviluppo delle tecnologie digitali di funzionamento. A questo deve aggiungersi che la facilità di accesso a Internet, unitamente all'economicità di strumenti scrittura o registrazione audio-visiva, consentono ormai a ciascun individuo, non solo d'informarsi da canali plurimi e variegati (ufficiali e non) ma anche di concorrere direttamente alla creazione, diffusione e sistematizzazione delle notizie.

Le piattaforme di condivisione di contenuti e di incontro/confronto *online*, genericamente definite *social network* (o semplicemente *social*), che, per la particolarità delle funzioni digitali, rappresentano una sorta di ibrido tra un siste-

* Dottorando di ricerca in Diritto e Impresa, Università degli Studi di Roma Unitelma Sapienza. Il presente contributo è stato redatto dall'Autore quando era Assegnista di ricerca (di tipo B), in diritto pubblico, per il progetto PNRR “*Security and Rights in CyberSpace*” – SERICS (ACK: PE00000014), presso il Dipartimento di Studi Internazionali Giuridici Storico-Politici dell'Università degli Studi di Milano.

ma di comunicazione interpersonale chiuso e un mezzo di diffusione di notizie (senza però essere mass media), contribuiscono ad ampliare il volume dell'immissione di informazioni.

Tali implementazioni tecnologiche, pertanto, stanno comportando profondi mutamenti sociali.

Oltre al miglioramento dei mezzi, delle capacità e dei contenuti dell'informazione e della comunicazione (sia interpersonale che di massa), la poliedrica pervasività dei nuovi strumenti *hi-tech* ha amplificato anche i problemi già esistenti nella diffusione delle notizie.

È proprio la moltitudine dei mezzi e degli "operatori" coinvolti a incrementare i problemi del mondo dell'informazione. Se la pluralità delle fonti e la quasi immediatezza delle notizie, garantite dalla rete Internet, possono essere un elemento di forza per una libera informazione, allo stesso modo rappresentano una potenziale insidia per la loro sovrabbondanza e/o imprecisione. In tal senso, la moltitudine delle fonti genera una sovraesposizione alle notizie in un contesto che, alimentato dall'immediatezza delle tecnologie di comunicazione di massa, produce una saturazione informativa, con inevitabile confusione.

La corsa all'esclusiva dei mass media – in competizione tra loro e, contemporaneamente, anche con operatori non ufficiali (come *blogger*, *influencer* e, *generaliter*, divulgatori digitali) – crea una sorta di "bulimia da notizie", in cui quest'ultime sono più uno strumento per imporsi sui concorrenti che l'obiettivo del servizio.

Questo quadro, così variegato, multipolare e – a volte – incerto, costituisce spesso il terreno ideale per condotte contrarie all'intento di informare e/o, comunque, pur avendo il fine di dare notizia, per diversi motivi (dall'imperizia al recentismo), raggiungono l'effetto opposto. Si verifica, pertanto, il fenomeno del c.d. "disordine informativo".

Il termine disordine informativo (apparso prima nella sua forma inglese di *information disorder*¹) sussume sotto la propria macro-categoria una serie di dinamiche legate all'inquinamento della sfera dell'informazione pubblica, non rubricabili solo alla semplice diffusione di notizie false.

Si tratta di un variegato ventaglio di contenuti, potenzialmente anche satirici-parodistici o ludico-cinematografici² (strumentalizzabili in rete), che vanno dalle generiche falsità (c.d. *fake news* giornalistiche) ad argomentazioni totalmente false (fin dalla fonte *ab origine*), con l'obiettivo di ingannare e generare

1. C. Wardle, H. Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policy making*, Report to the Council of Europe, 2017.

2. A tale proposito, si pensi ai video (anche realizzati con IA) per scene di videogiochi o di effetti speciali cinematografici.

conseguenze manipolando la sfera pubblica. Il discrimine per distinguere le condotte pericolose da quelle innocue, ruota attorno all'intento e alla responsabilità dei soggetti coinvolti nella loro creazione e diffusione.

Per quanto riguarda le condotte dannose, in origine, la dottrina britannica riconduceva al *information disorder* tre differenti: la disinformazione, la misinformazione e la malinformazione³.

L'attuale e più aggiornata tassonomia giuridica⁴ sussume sotto lo stesso *genus* del disordine informativo tutte quelle condotte, anche non necessariamente volontarie e dolose, che possono essere – escludendo i “*rumors*”⁵ – nelle cinque macro-categorie di: disinformazione, misinformazione, informazione malevola, malinformazione e informazione improvvida o coperta da recentismo.

2. La pericolosità per lo Stato

Considerando la vastità delle condotte dannose, ascrivibili al disordine informativo, e la sottigliezza con cui possono essere attuate e combinate tra loro (di cui si dirà approfonditamente in seguito, in sede di classificazione), appare evidente come tutti i loro effetti (disinformanti, manipolativi e/o screditanti) rappresentino un rischio per lo Stato e, ancor prima, per la tenuta istituzionale e sociale⁶.

Le insidie possono essere di diversa natura; da aspetti marginali e connessi soltanto al diritto d'informazione dei singoli cittadini (legato alla libertà di manifestazione del pensiero, *ex art.* 21 Cost.) alla destabilizzazione politica e sociale dell'intero Paese colpito⁷.

3. C. Wardle, *Understanding Information Disorder*, in *firstdraftnew.org*, 2019, disponibile su: www.firstdraftnews.org, n. 24/10/2025.

4. Garante per la Protezione dei Dati personali, *Discorso sul disordine informativo*, in *garanteprivacy.it*, 2023.

5. Il termine indica, nel linguaggio giornalistico, una voce o diceria che circola intensamente ma non è confermata in modo ufficiale. Si v. *Rumor*, in *Vocabolario Treccani*. Nell'ambito finanziario, invece, i “*rumors*” sono notizie e informazioni confidenziali non ufficiali che circolano nell'ambiente di settore. Si tratta di informazioni rilevanti ma poco verificabili, riguardanti operazioni e vicende di emittenti con titoli quotati, come ad esempio un aumento di capitale. La diffusione di tali *rumors* nasce dalla presenza di asimmetrie informative tra i soggetti che partecipano al mercato ed è finalizzata a sfruttare eventi ritenuti “*price sensitive*” non ancora ufficiali e destinati a impattare, una volta confermati, sul valore delle azioni. Si v. *Rumors*, in *Glossario Finanziario di Borsa Italiana*.

6. Si veda sull'argomento, il report del 10 gennaio 2024 del World Economic Forum, che le ha inserite tra le minacce più rilevanti per la stabilità dei Paesi e dello stesso sistema democratico-economico e valoriale occidentale, testo disponibile al sito: www3.weforum.org.

7. 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.

La limitazione del diritto all'informazione generale, la manipolazione dell'opinione pubblica, lo scredito sistematico delle Istituzioni, l'infiltrazione nel circuito dell'informazione e della sicurezza pubblica, la destabilizzazione politico-sociale, la compromissione, il furto e il danneggiamento dei dati veicolati nei circuiti informativi portanti e degli stessi sistemi di archiviazione e diffusione, sono solo i principali rischi che uno Stato può dover fronteggiare in caso di azioni dannose di disordine informativo⁸.

La pericolosità del fenomeno non è insita solo nelle diverse azioni o effetti disinformanti ascrivibili ma anche nel fatto che, se attuato con intento nocivo da un'entità statale esterna, qualsiasi disordine informativo può essere a sua volta parte di una più ampia campagna di guerra ibrida o, peggio, anche preparatorio di un'azione bellica "classica"⁹.

Tralasciando i casi in cui il disordine informativo è prodromico di azioni di guerra tradizionale (come fu per le operazioni britanniche "*Mincemeat*" e "*Bodygard*" nella Seconda guerra mondiale), la guerra ibrida può essere considerata come una combinazione di tattiche convenzionali, irregolari e cibernetiche finalizzate a destabilizzare gli avversari. Tra le tattiche di cui può avvalersi in tale processo, appunto *ibrido*, vi è la guerra cognitiva (che attacca la cognizione umana, influenzando le convinzioni e il processo decisionale), che a sua volta può essere propiziata dalla guerra d'informazione (che manipola o disturba le informazioni per condizionare *ex ante* la percezione della realtà). Infine, per aggredire lo spazio digitale di Internet, vi è la c.d. guerra cibernetica che utilizza attacchi informatici per compromettere sistemi, reti o dati per obiettivi militari o strategici.

Il disordine informativo *online*, pertanto, sussistendo nell'ambiente digitale, può essere considerato sia una minaccia all'informazione e, più in generale, una forma di ostilità cognitiva, quanto una potenziale e sussidiaria insidia cibernetica, nel contesto delle condotte di guerra ibrida.

Pertanto, nelle condotte disinformanti veicolate attraverso Internet, possono sussistere contemporaneamente i tre maggiori profili delle azioni belliche ibride. Queste possono comportare, infatti, la manipolazione, il controllo e la contraffazione delle informazioni *online* o la produzione stessa di fonti false (g. d'informazione), essendo elaborate tenendo conto anche dei processi cerebrali umani, al fine di ottenere la massima influenza (g. cognitiva) e – in taluni

8. S. Giusti, E. Piras, *Democracy and Fake News Information Manipulation and Post-Truth Politics*, Routledge, 2020.

9. A. Giannuli, *Come i servizi segreti usano i media*, Ponte alle Grazie, 2012, p. 88; e, ancora, più ampiamente, A. Giannuli, A. Curioni, *Cyberwar: la guerra prossima ventura*, Mimesis, 2019, p. 21 ss.

casi – essere veicolate mediante attacco cyber o, al contrario, essere esse stesse vettore o strumento per la penetrazione di *malware* in un sistema informatico (g. cibernetica).

Considerata l'estrema versatilità delle condotte di disordine informativo (sopradescritte) e la sottigliezza con le quali le stesse si possono combinare tra loro – per occultarsi, non essere rilevate e centrare così l'obiettivo – appare evidente come tutti i fenomeni disinformati siano considerati un rischio per gli Stati, al centro delle agende internazionali di prevenzione e contrasto¹⁰.

I rischi per lo Stato posso essere di diversa natura, andando da aspetti marginali e connessi solo alla mera informazione dei singoli cittadini (comunque sancita dal diritto costituzionale a essere informati, *ex art. 21 Cost.*) alla destabilizzazione politica e sociale dell'intero Paese colpito.

La limitazione del diritto all'informazione generale, la manipolazione dell'opinione pubblica da parte di Paesi ostili, lo scredito sistematico delle Istituzioni, l'infiltrazione nel circuito dell'informazione e della sicurezza pubblica, la destabilizzazione politico-sociale, la compromissione, il furto e il danneggiamento dei dati veicolati nei circuiti informativi portanti e degli stessi sistemi di archiviazione e diffusione, sono solo i principali rischi – posti in un'evidente scala crescente – che lo Stato può correre se non contrasta efficacemente le campagne disinformati¹¹.

Si comprende, perciò, come le condotte di disordine informativo, soprattutto quelle frutto di intenti dannosi, rappresentino un rischio diffuso per il diritto all'informazione dei cittadini, l'integrità sociale e financo la tenuta democratica degli Stati.

3. La classificazione delle condotte disinformati

Seguendo il principio per cui non è possibile argomentare di qualcosa senza averne colto preliminarmente l'essenza e, quindi, la sua sussunzione tassonomica, ancor più per le condotte disinformati del disordine informativo – trattandosi di un fenomeno variegato e composito nell'estrinsecazione – risulta necessaria un'aggiornata definizione giuridica.

Come si è detto l'ultima dottrina giuridica, ricomprende nella nozione di

10. Si veda sull'argomento, il report del 10 gennaio 2024 del World Economic Forum, che le ha inserite tra le minacce più rilevanti per la stabilità dei Paesi e dello stesso sistema democratico-economico e valoriale occidentale.

11. S. Giusti, E. Piras, *Democracy and Fake News Information Manipulation and Post-Truth Politics*, cit., p. 4.

disordine informativo una serie di condotte che, pur non essendo in alcuni casi caratterizzate da un elemento intenzionale o doloso, trovano una propria sistemazione – con l’eccezione dei meri *rumors* – all’interno di cinque distinte categorie: la disinformazione, la misinformazione, l’informazione malevola, la malinformazione e l’informazione improvvida o affetta da recentismo.

Andando con ordine, si proverà a dare una definizione di ciascuna.

Nel caso della prima tipologia, ovvero della disinformazione “semplice” o “classica”, si tratta di una condotta che prevede la volontaria creazione e/o diffusione dolosa di informazioni false, allo scopo di arrecare danni a uno Stato o a un sistema di Paesi, con canali informativi/mediatici integrati o congiuntamente facilmente influenzabili¹² (in quest’ultimo caso, si può considerare che il risultato di una campagna riuscita di disinformazione verso, ad esempio, la Francia può originare una misinformazione in Italia o viceversa). Fino alla prima metà del Novecento le azioni di disinformazione sono state operate quasi esclusivamente dagli apparati nazionali di informazione e sicurezza, che le adottavano per influenzare altri attori statuali o per “preparare il terreno” a eventuali azioni belliche contro Paesi ostili. Con l’evoluzione dei mezzi di comunicazione di massa (la diffusione, per scopi civili, del canale di Internet) e la comparsa di attori globali non nazionali e asimmetrici si sono sviluppate anche condotte disinformati trasversali e non nazionali, che possono perseguire fini criminali “terzi” e contrari rispetto alle politiche degli Stati o agire in loro appoggio ma senza dipenderne direttamente, in modo da mascherare gli obiettivi e di mimetizzarsi tra forme di informazione d’inchiesta e/o filantropiche.

Un’ulteriore tecnica di disinformazione è la propaganda bellica disinformatante, che non tende solo alla pubblicizzazione di notizie edulcorate per influenzare l’opinione pubblica interna ma ha in quest’ultima il canale (e non l’obiettivo) di diffusione della falsa notizia¹³, secondo il principio “se ci credono i nostri cittadini ci crederà anche il nemico (che magari avrà degli informatori all’interno)”.

La misinformazione, invece, consiste in una condotta che prevede la creazione e/o diffusione inconsapevole di informazioni false. In tal caso, colui che diffonde o genera la notizia (magari, ad esempio, dopo aver consultato una fonte audio-visiva o traducendo la stampa estera) è in buona fede, senza accorgersi che, invece, sta contribuendo a propagare disinformazione¹⁴.

12. In tal caso, si può considerare che il risultato di una campagna riuscita di disinformazione verso, ad esempio, la Francia può originare una misinformazione in Italia o viceversa.

13. S. Gigante, *L’arte oscura della disinformazione: come si fa guerra alle fake news*, in *Agenda Digitale*, 16 dicembre 2024.

14. O. Pollicino, P. Dunn, *Disinformazione e intelligenza artificiale nell’anno delle global elections*, in *federalismi.it*, n. 12, 2024, p. 8 ss.

Si tratta, per questo, di un'azione che non configura il dolo per chi diffonde (giornalisti, Istituzioni o privati cittadini) ma che integra, a seconda delle fattispecie, le possibili responsabilità colpose per imperizia, culpa in vigilando e scarsa professionalità. Oltre ai metodi "classici" per indurre la misinformazione (ad es. voci diffuse ad arte, false fonti di notizia fatte trovare a improvvisi funzionari o giornalisti o personaggi mediaticamente influenti) e alla misinformazione indotta "di riflesso" per disinformazione altrui, grazie all'affinamento delle tecniche digitali di "effetti speciali" è nato – nell'ultimo decennio – lo strumento del *deepfake*¹⁵, che consiste nell'alterazione, per rielaborazione informatica artificiale, di un contenuto audiovisivo, con un grado di verosimiglianza tale da essere difficilmente distinguibile dal vero per l'occhio umano.

La misinformazione, inoltre, aumenta la propria efficacia in proporzione al grado di fama e affidabilità della fonte che, inconsapevolmente, la crea (tanto sarà noto e popolare l'autore della notizia tanto risulterà attendibile il contenuto della stessa; secondo il principio per cui "se lo dice lui/lei sarà vero").

L'informazione malevola, da *malicious information* ("informazione dannosa" o "i. maliziosa"), a sua volta, invece, consiste nella diffusione volontaria e dolosa di notizie vere ma coperte da un regime di riservatezza ufficiale (in Italia se ne distinguono di quattro livelli¹⁶), allo scopo di creare conseguenze avverse e/o scredito nelle Istituzioni (principalmente governi) che le hanno segretate. Molto spesso tale condotta è perpetrata a seguito di altri atti di guerra ibrida come attacchi *hacker* ai server governativi e/o banche dati riservate, sottrazione materiale di documenti di importanza strategica o corruzione di funzionari preposti¹⁷.

Quest'ultima non deve essere confusa con la *malinformazione* che è un'informazione accurata ma diffusa con intento malevolo¹⁸. Si tratta di materiale sensibile che viene diffuso per danneggiare qualcuno o la sua reputazione¹⁹. Tra gli esempi vi sono il *doxing*, il *revenge porn* e l'*editing* di video per rimuovere contesti o contenuti importanti²⁰.

Infine, l'informazione improvvida o influenzata da recentismo è una quarta categoria che contribuisce a incrementare il disordine informativo, seppur ben

15. Garante per la Protezione dei Dati personali, *Deepfake. Il falso che ti "ruba" la faccia e la privacy*, 28 dicembre 2020.

16. Questi sono: "Riservato (R)", "Riservatissimo (RR)", "Segreto (S)" e "Segretissimo (SS)".

17. P. Tettamanzi, M. Rijllo, *Cyber Insurance: strategia, gestione e governance*, Ipsos, 2024, p. 62 ss.

18. N. Marquez, *Research Guides: Misinformation – Get the Facts: What is Misinformation?*, in *guides.lib.uci.edu*, 16 marzo 2023.

19. Da *What is disinformation?*, in *Die Bundesregierung informiert, Startseite* (sito istituzionale tedesco), 23 marzo 2023.

20. Si veda il report *Foreign Influence Operations and Disinformation*, in *Cybersecurity and Infrastructure Security Agency – CISA*, disponibile su: www.cisa.gov.

distinta dalle precedenti, è rappresentata dalla diffusione avventata di notizie (c.d. informazioni improvvise o influenzate da recentismo) vere ma parziali o non ancora completamente definite (in tal caso si parla più specificamente di “*midinformazione*”), rispetto ad avvenimenti complessi e molto recenti.

Con l’affermazione delle piattaforme *social* e, in generale, del Web tra i canali di diffusione delle notizie e di contenuti audiovisivi, il mondo dell’informazione è stato progressivamente influenzato, tanto nella metodica della ricerca delle fonti quanto nelle tempistiche narrative, tendendo a privilegiare sempre più la narrazione mediatica ed emozionale dei fatti rispetto al loro approfondimento critico contenutistico. La “corsa all’esclusiva” (ovvero a fornire per primi un’informazione per superare la concorrenza mediatica) origina, in contesti complessi e mutevoli²¹, informazioni parziali o contraddittorie con gli sviluppi successivi dei fatti o con gli stessi approfondimenti tematici, generando confusione e disorientamento nell’opinione pubblica, nonché senso di sfiducia verso le istituzioni e i canali d’informazione accreditati, lasciando così spazio al possibile insinuarsi di eventuali azioni di disinformazione o misinformazione indotta.

Per quanto azioni tecnicamente dissimili, considerata la poliedricità delle tecnologie di canale mediatico e l’interconnessione dei rapporti causa-effetto, spesso si possono riscontrare contemporaneamente le quattro condotte disinformanti, direttamente concatenate tra loro²².

4. L’importanza della definizione giuridica delle condotte

La transizione dall’analisi fenomenologica alla costruzione di una risposta giuridica efficace poggia sul cardine della definizione normativa delle condotte disinformanti. In assenza di una tassonomia giuridica precisa, il fenomeno del disordine informativo rischia di essere sottovalutato o affrontato con strumen-

21. Come nel caso dell’emergenza pandemica o degli eventi bellici d’Ucraina.

22. Può essere esemplificativo il seguente caso di scuola, costruito *ad hoc*, che racchiude tutte le condotte di disordine informativo.

Si verifica un evento inaspettato e il circolo mediatico reagisce, sulla spinta della “corsa all’esclusiva”, fornendo subito notizie parziali, affrettandosi a contendersi le esclusive, le stesse informazioni vengono però smentite dai fatti successivi e dagli stessi mass media; si crea confusione nell’opinione pubblica, che genera sfiducia nelle Istituzioni e nei canali di informazione tradizionali. A tal punto, uno Stato ostile, per aggravare la situazione, immette altre notizie false o vere segretate (quindi sottratte e desegretate) aumentando così il disordine informativo. Alcuni comparti della libera informazione (sia nazionali che esteri) riprendono quest’ultime notizie, presentandole come verità e accreditandole con la loro attendibilità di testata/autore, diffondendole ulteriormente. che racchiude tutte le condotte di disordine informativo.

ti giuridici generici e inadeguati, oppure al contrario, di generare interventi repressivi privi del necessario presidio di determinatezza.

Il principio di tassatività, espressione del più generale valore dello Stato di diritto, esige che la legge descriva in modo chiaro e univoco i comportamenti sanzionati, al fine di delimitare l'ambito di esercizio del potere punitivo e garantire la libertà dei consociati.

Una nozione vaga e onnicomprensiva, come quella giornalistica di "*fake news*", che non coglie le varie peculiarità delle differenti condotte disinformanti, non solo si presterebbe a potenziali applicazioni arbitrarie, con indubbia compromissione della libertà di manifestazione del pensiero (*ex art. 21 Cost.*) ma anche, paradossalmente, contribuirebbe a creare altro disordine informativo, sanzionando in modo indifferenziato ogni forma di espressione erronea o imprecisa.

La sfida costituzionale consiste, quindi, nel selezionare, all'interno del macro-fenomeno del disordine informativo, quelle condotte che, per la loro specifica modalità di realizzazione, l'intenzionalità fraudolenta o la concreta capacità di cagionare un pericolo per beni giuridici primari, superano la soglia di tolleranza dell'ordinamento e meritano una definizione giuridica propria, da cui far discendere delle opportune misure di contrasto. Tale operazione non deve essere intesa come un mero esercizio tecnico giuridico e legislativo, ma un come atto di bilanciamento necessario tra l'esigenza di tutelare la pubblica sicurezza e l'integrità del dibattito democratico e la libertà di espressione, che include, in una democrazia matura, anche la possibilità di esprimere (finanche di ritenere personalmente vere) opinioni minoritarie, dissenzienti o perfino erronee.

Solo una definizione rigorosa, che isoli le condotte caratterizzate da un nucleo di pericolosità oggettiva per lo Stato o di illiceità penale²³ può consentire di evitare la deriva verso una "Giustizia della Verità", applicata al dissenso e di preservare lo spazio costituzionalmente garantito al libero confronto delle idee.

5. Il caso della misinformazione "indotta"

Considerato l'apporto che le tecnologie digitali di internet hanno anche nel perpetrarsi del fenomeno del disordine informativo, la misinformazione è tra le condotte più elusive e al contempo pericolose; non tanto nella sua forma "semplice", legata alla sua dimensione inconsapevolezza o di erroneità colposa del diffusore all'origine, ma nella sua potenziale dimensione "indotta" da un

23. Come l'alterazione consapevole e finalizzata della realtà in un contesto che ne facilita la credibilità e la diffusione virale con potenziale lesivo per beni individuali o collettivi.

terzo attore malintenzionato. In tal caso, la fattispecie si sostanzierebbe in una condotta di creazione e/o diffusione inconsapevole di informazioni false, la cui origine, tuttavia, non è riconducibile a una mera negligenza del soggetto agente, bensì a un'attività di manipolazione esterna finalizzata a sfruttare l'apparente affidabilità e la buona fede della fonte.

L'elemento distintivo risiederebbe nel meccanismo eterodiretto della condotta: un soggetto terzo, tipicamente in possesso di finalità disinformati, progetta e immette nel circuito mediatico un contenuto falso (senza specificarne la natura e l'intento), costruito con tale perizia da superare i consueti controlli di veridicità, affinché venga ripreso e divulgato da soggetti – giornalisti, influencer, persino istituzioni – la cui credibilità funge da moltiplicatore di legittimazione e diffusione. Per questo, il *deepfake* può essere uno degli strumenti principali di tale induzione²⁴, ma essa può realizzarsi anche attraverso la contraffazione di documenti, la creazione di siti-web camuffati da fonti giornalistiche accreditate o l'alterazione maliziosa del contesto di notizie vere.

La rilevanza giuridica di questa condotta ibrida risiede nella difficoltà di inquadrare la responsabilità del diffusore in buona fede, il cui operato, sebbene colpevole per imperizia o per aver omesso i dovuti accertamenti, manca dell'elemento soggettivo proprio del dolo. D'altro canto, l'autore della manipolazione, spesso occulto e operante da giurisdizioni estere, risulta di difficile identificazione e sanzione. A tale proposito, è d'uopo far notare che la fonte falsa lasciata in rete Internet, intenzionalmente, con la volontà di indurre una misinformazione, non può essere un prodotto d'autore (anche *deepfake*) creato a fini artistici, ludici o satirico-parodistici; poiché altrimenti verrebbe meno l'*animus nocendi* dell'autore iniziale e ci sarebbe la colpa (per imperizia, scarsa professionalità, avventatezza o superficialità) del diffusore ingannato.

Tale complessità rende molto difficile il rilievo e, quindi, il contrasto della misinformazione indotta, rendendola – a tutti gli effetti – come la forma di condotta disinformativa più pericolosa nell'attualità e nel prossimo futuro.

Si impone, pertanto, una riflessione sull'adeguatezza degli strumenti giuridici tradizionali, spingendo verso un approccio che, fatti salvi i profili di colpevolezza, consideri l'introduzione di fattispecie di pericolo che colpiscano la condotta di chi forgia e immette deliberatamente un contenuto falsato, senza specificarne (ad esempio con una didascalia o una "bollinatura") la natura e il fine, nonché l'implementazione di obblighi di diligenza rafforzata per gli operatori dell'informazione che, per il loro ruolo e la loro visibilità, sono i vettori inconsapevoli ideali di tali manipolazioni.

24. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, 2023, p. 172 ss.

6. La possibilità di prevenire giuridicamente gli effetti delle condotte disinformanti

Alla luce della stratificata tipologia di condotte e della loro intrinseca pericolosità, si impone un'analisi sulla possibilità di prevenire giuridicamente gli effetti del disordine informativo. La risposta dell'ordinamento non può limitarsi alla sola repressione *ex post*, spesso inefficace quando il danno all'ecosistema informativo e alla fiducia pubblica si è ormai consolidato. La prevenzione giuridica deve articolarsi su un duplice binario: da un lato, il potenziamento di strumenti di natura civilistica e amministrativa, e dall'altro, un uso parsimonioso ma fermo dello *ius puniendi*. Sul primo versante, assumono cruciale importanza gli obblighi di trasparenza e responsabilità a carico delle piattaforme digitali, anche alla luce del regolamento europeo sul *Digital Services Act*, finalizzati a contrastare la viralità dei contenuti falsi attraverso meccanismi di tracciabilità e di correzione delle notizie inesatte. Parallelamente, l'investimento nell'educazione digitale e nell'alfabetizzazione mediatica rappresenta una forma di prevenzione sociale indiretta, volta a potenziare la capacità critica del destinatario finale dell'informazione. Sul versante penale, la prevenzione passa attraverso la definizione di fattispecie incriminatrici che colpiscano non l'errore in sé, ma le condotte a rischio elevato e caratterizzate da un dolo specifico, quali la manipolazione maliziosa durante emergenze nazionali o la creazione di un pericolo concreto per l'ordine pubblico e la sicurezza nazionale. In tale quadro, la categoria dei reati di pericolo, se utilizzata con cautela per non scadere in un diritto penale simbolico, può costituire uno strumento idoneo a intervenire prima della lesione del bene giuridico.

La prevenzione giuridica efficace, in definitiva, non risiede in una singola norma, ma in un approccio sistemico che, partendo da una definizione giuridica più accurata (nel solco dei principi costituzionali) della forma di disordine informativo, integri il contrasto delle condotte più gravi con politiche di promozione della trasparenza, dell'educazione e della responsabilità di tutti gli attori della filiera informativa.

LA CRISI DELLE DEMOCRAZIE COSTITUZIONALI AL TEMPO DELLA DISINFORMAZIONE: QUALE FUTURO PER LA LIBERAL-DEMOCRAZIA?

*Luca Maria Tonelli**

SOMMARIO: 1. Premessa. Cos'è il "costituzionalismo digitale" – 2. Alcune precisazioni terminologiche in ordine al fenomeno della disinformazione – 3. Sfera pubblica, sfera privata e libertà di manifestazione del pensiero nell'art. 21 Cost. – 4. Il *Digital Services Act* come quadro regolatorio per un'embrionale "democrazia protetta" europea – 5. Conclusioni. Un "nuovo" costituzionalismo fondato sul binomio verità-libertà.

1. Premessa. Cos'è il "costituzionalismo digitale"

Negli ultimi anni, l'espressione "costituzionalismo digitale"¹ ha acquisito crescente rilievo nel dibattito giuridico e filosofico-politico internazionale. Con tale locuzione si suole esprimere il tentativo di trasferire e adattare i principi, le garanzie e i meccanismi del costituzionalismo moderno all'ambiente digitale, in cui si producono, si scambiano e si governano oggi, gran parte delle interazioni sociali, economiche e politiche. In altri termini, il costituzionalismo digitale si prefigge l'obiettivo di ricercare nuovi strumenti di limitazione e di legittimazione del potere – non più solo statale, ma ora anche privato e tecnologico – nell'era delle piattaforme e dei dati.

Il costituzionalismo classico, nato tra il XVII e il XVIII secolo, mirava a vincolare il potere politico attraverso il diritto, ponendo al centro i diritti fon-

* Assegnista di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

1. Sul "costituzionalismo digitale" cfr., *ex multis*, E. Celeste, *Digital Constitutionalism. The Role of Internet Bills of Rights*, Routledge, 2022; V. Cavani, *Nuovi poteri, vecchi problemi. Il costituzionalismo alla prova del digitale*, in *Diritto pubblico comparato ed europeo*, n. 1/2023, p. 236 ss.; O. Caramaschi, *Il costituzionalismo al cospetto dell'intelligenza artificiale: nuove sfide, quali soluzioni?*, in *Rivista italiana di informatica e diritto*, n. 1/2025, p. 30 ss.; O. Pollicino, *Costituzionalismo digitale. Pensare la democrazia al tempo dell'IA*, il Mulino, 2025.

damentali dell'individuo e la separazione dei poteri. Nello spazio digitale, tuttavia, le fonti del potere si sono moltiplicate e frammentate: accanto agli Stati agiscono oggi grandi attori privati globali – piattaforme digitali, motori di ricerca, *provider* di servizi – che esercitano un controllo di fatto sull'accesso all'informazione, sulla *privacy* e persino sulla libertà di espressione².

In questo scenario, la tradizionale dialettica tra pubblico e privato, tra sovranità statale e libertà individuale, si mostra inadeguata. Le piattaforme digitali dispongono di un potere normativo e regolativo che si manifesta attraverso algoritmi, condizioni d'uso e *policy* di moderazione dei contenuti: un potere non elettivo, non territoriale e spesso privo di meccanismi di responsabilità democratica³. Il costituzionalismo digitale si propone dunque di ricondurre questo nuovo potere entro forme di diritto e di garanzia, mutuando i principi di legalità, proporzionalità e trasparenza nel linguaggio della tecnologia.

L'idea di fondo è che lo spazio digitale non sia un vuoto giuridico, ma un nuovo spazio costituzionale: una "sfera pubblica algoritmica" nella quale si esercitano diritti e si producono decisioni che incidono direttamente sulla vita delle persone⁴. Da qui l'esigenza di individuare "costituzioni digitali", intese come insieme di regole, valori e limiti che disciplinano l'uso delle tecnologie e garantiscono l'equilibrio tra innovazione e tutela dei diritti.

Il costituzionalismo digitale si articola oggi lungo due principali direttrici: da un lato, la dimensione statale e sovranazionale, che mira a integrare nei testi costituzionali esistenti la tutela dei diritti digitali, basti pensare alla Carta dei diritti fondamentali dell'Unione europea o alla recente Dichiarazione europea sui diritti e i principi digitali del 2022; dall'altro, la dimensione globale e multi-attore, che si interroga su come garantire principi di responsabilità, trasparenza e controllo democratico nei confronti di poteri tecnologici transnazionali⁵.

In questa prospettiva, il costituzionalismo digitale non è soltanto un'estensione del costituzionalismo tradizionale, ma una sua trasformazione paradigmatica. Esso invita a ripensare categorie come sovranità, cittadinanza e diritti

2. Piattaforme digitali che si atteggiavano come un vero e proprio "antisovrano", nel senso impiegato mirabilmente da M. Luciani, *L'antisovrano e la crisi delle costituzioni*, in *Rivista di diritto costituzionale*, n. 1, 1996, p. 124 ss. Che le grandi piattaforme digitali e, soprattutto, i *social network* avrebbero influito in modo determinante e dirimente sull'ossatura degli ordinamenti liberal-democratici, lo aveva ben chiaro, già nel 2004, S. Rodotà, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Laterza, 2004, spec. p. XXXIII ss. e p. 79 ss.

3. Di questo avviso è anche E. Cremona, *I poteri privati nell'era digitale. Libertà costituzionali, regolazione del mercato, tutela dei diritti*, Esi, 2023.

4. J.E. Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, Oxford University Press, 2019.

5. A. Lamberti, *Costituzionalismo digitale, poteri delle piattaforme, intelligenza artificiale e democrazia*, in *ConsultaOnline*, n. 2/2025, p. 1096 ss.

fondamentali in un contesto in cui la *governance* è distribuita e il potere si esercita attraverso infrastrutture tecniche e flussi di dati.

Come scriveva Lawrence Lessig, «il codice è la nuova legge»⁶: ma se il codice regola la condotta, allora il compito del costituzionalismo digitale è quella di assicurare che anche il codice sia sottoposto al diritto.

Nelle piattaforme digitali, soprattutto a seguito dello sviluppo dei *Large Language Models* (Llm), si trovano oggi ampie applicazioni nella generazione automatizzata di testi e nella diffusione sistemica della disinformazione, la quale rappresenta una delle minacce più insidiose alla tenuta epistemica delle democrazie costituzionali. In un contesto in cui la produzione e la circolazione delle informazioni sono mediate da piattaforme private globali, la capacità di discernere il vero dal falso non dipende più solo dall'educazione civica o dal pluralismo dei media tradizionali, ma anche – e sempre più – dalle architetture tecniche e normative che governano lo spazio digitale. Ed è, altresì, evidente che la diffusione e l'utilizzo su larga scala di tali tecnologie comportano rischi significativi per l'integrità dei processi democratici e per la qualità del dibattito pubblico.

In questo ambito il costituzionalismo digitale europeo può giocare sicuramente un ruolo cruciale.

2. Alcune precisazioni terminologiche in ordine al fenomeno della disinformazione

Per comprendere appieno la portata dell'impatto che l'intelligenza artificiale esercita – anche – sul fenomeno della disinformazione, è necessario affrontare preliminarmente la questione definitoria.

Sebbene nel linguaggio comune si sia progressivamente diffusa la locuzione inglese *fake news*, tale espressione è stata oggetto, negli ultimi anni, di critiche sempre più ampie, non solo in ambito accademico, ma anche istituzionale⁷.

In particolare, l'*High Level Expert Group on Fake News and Online Disinformation* (Hleg), istituito dalla Commissione europea, nel proprio rapporto finale del 2018, ha espressamente raccomandato di abbandonare l'uso dell'espressione *fake news* in favore di quella di “disinformazione”, per due ordini di ragioni.

In primo luogo, si è evidenziato come il concetto di “notizie false” sia ecces-

6. L. Lessig, *Code: And Other Laws of Cyberspace*, Basic Books, 1999.

7. Di questo avviso O. Pollicino, P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *federalismi.it*, n. 12, 2024, p. V s.

sivamente riduttivo, in quanto esclude molteplici manifestazioni del fenomeno che non si esauriscono nella mera falsità del contenuto.

In secondo luogo, l'uso mediatico e politico del termine *fake news* ha assunto, nel tempo, una connotazione distorta e strumentale, spesso impiegata da figure pubbliche per delegittimare informazioni o opinioni sgradite, più che per designare effettivi fenomeni di falsificazione.

Alla luce di tali considerazioni, appare preferibile il ricorso alla categoria concettuale di disinformazione⁸, intesa – secondo la definizione proposta dallo stesso Hleg – come la creazione, presentazione e diffusione deliberata di informazioni false, inaccurate o fuorvianti, prodotte con l'intento di conseguire un vantaggio economico, politico o ideologico, ovvero di arrecare un danno pubblico, ivi inclusa l'ingerenza nei processi elettorali e democratici.

Occorre tuttavia precisare che la nozione così delineata non coincide con le ipotesi penalmente o civilmente illecite previste dall'ordinamento (ad es. la diffamazione o la calunnia), ma abbraccia tutte quelle condotte che, pur non integrando un illecito giuridico, possono comunque arrecare un pregiudizio ai valori e ai principi democratici. Sono proprio tali ipotesi “di confine” – in cui la falsità non è punita in sé, ma produce conseguenze sistemiche sul discorso pubblico – a rappresentare la sfida più complessa per il legislatore nazionale ed europeo, chiamato a bilanciare la tutela del corretto funzionamento dello spazio democratico con la necessità di non comprimere in modo sproporzionato la libertà di espressione.

L'elemento intenzionale costituisce, in questa prospettiva, il criterio distintivo tra disinformazione e misinformazione.

La prima si caratterizza per la volontà consapevole di produrre e diffondere contenuti falsi con finalità manipolative o di vantaggio personale; la seconda, invece, si riferisce alla diffusione involontaria di informazioni inesatte, da parte di soggetti che agiscono nella convinzione della loro veridicità.

In altre parole, la *misinformazione* è un “disordine informativo” (*information disorder*) privo dell'elemento del dolo: la falsità deriva da errore, non da intenzione⁹.

Nell'ecosistema digitale, tuttavia, i confini tra le due fattispecie tendono a sovrapporsi. Spesso, contenuti disinformativi originati con dolo vengono amplificati da utenti o soggetti terzi che li ricondividono inconsapevolmente, contri-

8. L'analisi del fenomeno giuridico è svolta da L. Califano, *La libertà di manifestazione del pensiero... in rete; nuove frontiere di esercizio di un diritto antico. Fake news, hate speech e profili di responsabilità dei social network*, in *federalismi.it*, n. 26, 2021, p. 3 ss.; A. Ruffo, *Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, n. speciale 1/2024, p. 413.

9. Cfr., ancora, A. Ruffo, *op. loc. ult. cit.*

buendo alla loro viralità: lo stesso contenuto può dunque configurare disinformazione rispetto agli autori originari e misinformazione per i successivi diffusori.

Tale meccanismo, intrinseco alla logica delle piattaforme digitali, spiega la rapidità e pervasività con cui la falsità informativa si propaga in rete.

Fenomeno distinto, seppur contiguo, è quello della malinformazione, che consiste nella diffusione di informazioni formalmente veritiere – o basate su elementi reali – ma comunicate con l'intento di arrecare un danno, ad esempio attraverso la decontestualizzazione o la violazione della *privacy*¹⁰.

Sebbene con intensità differenti, disinformazione, misinformazione e malinformazione rappresentano tutti fattori di vulnerabilità democratica, in quanto incidono sui processi di formazione dell'opinione pubblica e, di conseguenza, sui meccanismi decisionali collettivi. La disinformazione, in particolare, mira deliberatamente a manipolare la percezione della realtà e ad alterare il giudizio politico dei cittadini, orientando o distorcendo il dibattito pubblico a vantaggio dell'agente disinformatore. Proprio per la combinazione di elemento oggettivo (falsità) ed elemento soggettivo (intenzionalità manipolativa), essa ha catalizzato l'attenzione del legislatore europeo e nazionale come fenomeno di rilevanza costituzionale e sistemica.

Occorre, infine, sottolineare che la disinformazione non è un fenomeno unitario, ma multiforme e graduato, che richiede – quindi – risposte differenziate e proporzionate.

Come osserva lo stesso Hleg, la disinformazione deve essere affrontata mediante un approccio multidimensionale, che integri misure normative con interventi di natura educativa e sistemica: promozione della trasparenza delle piattaforme, alfabetizzazione mediatica e informazionale dei cittadini, rafforzamento del giornalismo di qualità, nonché sostegno al pluralismo e alla sostenibilità del sistema mediatico. Solo un simile approccio complessivo consentirà di coniugare efficacemente la tutela della verità pubblica con la salvaguardia delle libertà fondamentali di informazione ed espressione.

3. Sfera pubblica, sfera privata e libertà di manifestazione del pensiero nell'art. 21 Cost.

Il nesso tra sfera pubblica e libertà di espressione rappresenta, sin dalle origini del costituzionalismo repubblicano, un terreno d'indagine complesso e tutt'altro che immediato, come testimonia l'articolato dibattito sviluppatosi intorno al diritto di manifestazione del pensiero. È noto, infatti, come l'art. 21

10. Così A. Ruffo, *op. ult. cit.*, p. 414.

della Costituzione si collochi al cuore dell'architettura dei principi e dei valori fondativi della Repubblica, segnando un netto superamento di quella concezione "funzionalizzata" al servizio dello Stato che aveva caratterizzato la stagione precedente, espressa nei numerosi reati d'opinione poi progressivamente espunti grazie all'intervento della giurisprudenza costituzionale¹¹.

Rispetto alla dimensione monolitica e gerarchica della società statutaria – in cui la libertà di stampa costituiva l'unico strumento di espressione pubblica, funzionale più al mantenimento dell'ordine che al pluralismo¹² –, la ricostruzione costituzionale del diritto di espressione riflette il mutato rapporto tra autorità e libertà individuale, ponendosi come strumento di attuazione del principio personalista¹³ e della pari dignità di ogni cittadino. Se sul piano prescrittivo tale evoluzione è già evidente nella formulazione del primo comma dell'art. 21, sul piano assiologico essa emerge anche nella ridefinizione del concetto di limite, che si emancipa dal tradizionale riferimento all'"ordine pubblico" per orientarsi verso la tutela della comunità democratica come luogo di espressione e realizzazione della persona, sia nella dimensione individuale sia in quella collettiva¹⁴.

La nuova sfera pubblica così delineata si configura, quindi, come spazio inclusivo, capace di accogliere le pluralità minoritarie e di valorizzare la diversità sociale e culturale, in netto contrasto con la rigidità dello Stato liberale, volto piuttosto alla conservazione del potere pubblico e alla neutralizzazione del dissenso¹⁵.

Questa lettura "evolutiva" della libertà costituzionale, derivante dalla sua collocazione nel sistema dei valori repubblicani, mette tuttavia in luce una tensione irrisolta: la vocazione relazionale del diritto si scontra con la difficoltà di individuare nel testo costituzionale un saldo ancoraggio alla dimensione collettiva della libertà di espressione. L'art. 21, infatti, non prevede esplicitamente limiti interni (a eccezione del riferimento al buon costume¹⁶) e non contiene

11. In questo senso si v. A. Di Giovine, *I confini della libertà di manifestazione del pensiero: linee di riflessione teorica e profili di diritto comparato come premesse a uno studio sui reati d'opinione*, Giuffrè, 1988, p. 100 ss.

12. Così U. De Siervo, *Stampa (dir. pubbl.)*, in *Enc. dir.*, vol. XLIII, Giuffrè, 1990, p. 580.

13. C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Giuffrè, 1958, pp. 8-9.

14. Coglie questo aspetto G. Vasino, *Lotta alla disinformazione, garanzie costituzionali e tutela dei diritti fondamentali nel panorama normativo dell'Unione europea*, in *federalismi.it*, n. 15, 2025, p. 108.

15. C. Caruso, *La libertà di espressione in azione. Contributo a una teoria costituzionale del discorso pubblico*, il Mulino, 2013, p. 84.

16. Sul limite in parola cfr. R. Perrone, *"Buon costume" e valori costituzionali condivisi. Una prospettiva della dignità umana*, Editoriale Scientifica, 2015.

elementi che connettano espressamente la pretesa individuale del soggetto alla costruzione di una sfera pubblica partecipativa e di uno spazio democratico condiviso.

Un'interpretazione congiunta, testuale e teleologica della disposizione ha messo in evidenza la sua impronta marcatamente individualistica¹⁷.

Da un lato, tale carattere esalta la primazia della persona nei confronti dello Stato; dall'altro, evidenzia la carenza di un esplicito riconoscimento del ruolo del cittadino come parte integrante della comunità democratica. Ne derivano notevoli difficoltà teoriche nel delineare una dottrina dei limiti che coniughi la tutela dell'autonomia individuale con le esigenze di costruzione di uno spazio pubblico pluralista¹⁸.

In sintesi, l'intento di codificare una frattura netta con l'esperienza autoritaria ha condotto i Costituenti a formulare l'art. 21 in chiave spiccatamente soggettiva.

Questa impostazione, tuttavia, ha ostacolato lo sviluppo di una teoria coerente e condivisa dei limiti, capace di resistere alla discrezionalità interpretativa e, soprattutto, di adattarsi ai profondi mutamenti del contesto sociale e tecnologico in cui la libertà di espressione oggi si manifesta.

Non sorprende, dunque, che le prime teorie sui limiti elaborate nei decenni successivi all'entrata in vigore della Costituzione abbiano mostrato segni di fragilità.

Il primo orientamento, di impronta formal-procedurale, tentava di colmare l'assenza di limiti "interni" attraverso l'individuazione di beni costituzionalmente rilevanti opponibili alla libertà di espressione, deducibili da altre norme della Costituzione o da fonti sub-costituzionali di attuazione¹⁹. Tale impostazione, accolta sia nella versione più ortodossa sia in quella evolutiva, ha avuto il merito di oltrepassare la mera lettera del testo, aprendo la strada a una prospettiva logico-sistematica in grado di considerare come limiti non solo i beni espressamente previsti, ma anche quelli implicitamente tutelati dall'ordinamento in funzione di protezione di diritti contrapposti²⁰.

In questo quadro, la garanzia della libertà non si fondava tanto sull'ancoraggio testuale, quanto sulla presenza di criteri interpretativi predeter-

17. Si v., per tutti, P. Barile, *Libertà di manifestazione del pensiero*, in *Enc. dir.*, vol. XXIV, Giuffrè, 1974, p. 429.

18. C. Esposito, *op. ult. cit.*, p. 12.

19. Ivi, p. 35 ss.

20. A. Pace, M. Manetti, *Art. 21*, in *Commentario della Costituzione*, diretto da G. Branca e continuato da A. Pizzorusso, il Mulino, 2006, p. 89. Nello stesso senso M. Mazzioti di Celso, *Appunti sulla libertà di manifestazione del pensiero nell'ordinamento italiano*, in Aa.Vv., *Scritti in onore di Vezio Crisafulli*, vol. II, Cedam, 1984, p. 518 ss.

minati, idonei a qualificare la condotta e a definire il significato materiale dell'espressione.

Tuttavia, anche questa elaborazione, pur caratterizzata da un solido rigore metodologico, non riusciva a occultare la presenza di una inevitabile dimensione valoriale, una sorta di «pre-comprensione materiale» del sistema dei principi costituzionali, che ne indeboliva la pretesa di assoluta oggettività²¹.

La riflessione successiva ha, dunque, intrapreso un percorso di ulteriore approfondimento, orientato verso una concretizzazione sostanziale della libertà di espressione, in cui il significato del diritto trova fondamento nelle garanzie diffuse dell'ordinamento a tutela del pluralismo e delle minoranze. In questa prospettiva, la libertà di manifestazione del pensiero viene riconnessa alla sua funzione di strumento di partecipazione democratica e di contrappeso contro-maggioritario, capace di alimentare il dibattito pubblico e di preservare la vitalità del pluralismo quale valore fondativo del sistema costituzionale.

Si afferma, in una fase successiva, un'interpretazione che non abbandona del tutto la nozione di limite formale, ma che la riqualifica alla luce dell'ispirazione garantista e del significato complessivo che promana dalla Costituzione intesa come sistema unitario. In tale prospettiva, prende forma una categoria di materie «privilegiate», la cui elevazione a settori dotati di particolare protezione non deriva tanto da una riserva di garanzia esplicitamente sancita dal testo costituzionale, quanto dal fatto di rappresentare l'espressione concreta della valorizzazione delle istanze minoritarie e delle esigenze pluralistiche che l'ordinamento, in via induttiva, riconosce e tutela.

Ne consegue che l'art. 21 non può essere interpretato unicamente come affermazione di un diritto soggettivo individuale, ma piuttosto come canone normativo generale volto a determinare quali forme espressive possano essere oggetto di legittima limitazione²². Tale impostazione segna un passaggio decisivo: essa non solo rompe con la concezione chiusa e formalistica dei vincoli che aveva confinato la libertà di espressione in un ambito di astratta neutralità, ma realizza anche una prima connessione strutturale tra la dimensione individuale del diritto e la trama dei principi fondamentali dell'ordinamento²³.

Il valore di questa interpretazione risiede, dunque, non soltanto nell'aver collegato la libertà del singolo alla comunità politica, ma soprattutto nell'aver

21. C. Caruso, *La libertà di espressione in azione*, cit., p. 94.

22. S. Fois, *Principi costituzionali e libera manifestazione del pensiero*, Giuffrè, 1957, p. 40.

23. Interpretazione avallata anche dalla Corte costituzionale: cfr. Corte cost., sentt. n. 234/1990, n. 21/1991, n. 155/2002, n. 312/2003.

proiettato la riflessione sui limiti della libertà in una dimensione sostanziale, collocando il diritto all'interno del sistema dei valori costituzionali e instaurando un dialogo diretto con le sue fondamenta democratiche. È proprio in questo cambiamento concettuale che si delinea, per la prima volta, una relazione strategica tra la libertà di espressione e la vocazione democratica del sistema costituzionale.

Tale lettura contribuisce a superare la presunta neutralità della libertà individuale, riconoscendo il suo ruolo centrale nella formazione dell'opinione pubblica e nel garantire la qualità del processo democratico, soprattutto nella prospettiva di assicurare la rappresentanza e la tutela delle minoranze²⁴.

Tuttavia, anche questa ricostruzione ermeneutica, per quanto coerente e feconda, non è rimasta immune da critiche e non può oggi fungere da chiave interpretativa definitiva rispetto al tema, attuale e controverso, della disinformazione.

Dietro tale percorso evolutivo, infatti, si intravede un rischio di rifunzionalizzazione ideologica della libertà di espressione: una tensione verso la subordinazione del diritto a finalità sistemiche che, pur ispirate alla tutela della democrazia, rischiano di condurre in un terreno teorico incerto e difficilmente governabile.

In particolare, una volta riconosciuto il nesso privilegiato tra libertà di espressione e integrità del processo democratico, risulta arduo fissare un punto di equilibrio stabile. La stessa impostazione, infatti, può generare esiti regolatori divergenti: da un lato, il riconoscimento della funzione garantista della libertà di espressione, volta ad assicurare che ogni voce – anche la più marginale – trovi spazio nel dibattito pubblico; dall'altro, la possibile giustificazione di limitazioni al contenuto espressivo in nome della tutela del corretto funzionamento della democrazia.

Questo paradosso conduce a una domanda di fondo: fino a che punto la democrazia può “proteggersi” da se stessa senza contraddirsi?²⁵

La soluzione potrebbe essere quella di ipotizzare l'esistenza, sul piano sovranazionale ed euro-unitario e – a cascata – anche a livello interno, di un modello di “democrazia protetta”, in cui la libertà di espressione venga bilanciata alla salvaguardia della stabilità del sistema democratico.

24. P. Barile, *op. loc. ult. cit.*; M. Luciani, *La libertà di informazione nella giurisprudenza costituzionale italiana*, in *Politica del diritto*, 1989, p. 613.

25. È la domanda che si pone, giustamente, nella sua ricostruzione anche G. Vasino, *Lotta alla disinformazione, garanzie costituzionali e tutela dei diritti fondamentali nel panorama normativo dell'Unione europea*, cit., p. 112.

4. Il *Digital Services Act* come quadro regolatorio per un’embrionale “democrazia protetta” europea

Proprio l’adozione del *Digital Services Act* (Reg. UE 2022/2065) segna un momento di svolta nel processo di maturazione del diritto europeo della rete e, più in generale, nel tentativo dell’Unione europea di definire un paradigma di costituzionalismo digitale capace di coniugare libertà fondamentali, responsabilità e salvaguardia della democrazia. In tale prospettiva, il Dsa può essere letto – pur nella sua natura prevalentemente regolatoria – come un primo embrione di un sistema a “democrazia protetta”²⁶ a livello euro-unitario e sovranazionale e, di riflesso, anche a livello interno.

Il riferimento ai valori fondativi dell’Unione, sanciti dagli artt. 2 e 7 TUE, costituisce il punto di avvio di questa riflessione. Tali disposizioni, com’è noto, consacrano la centralità del rispetto della dignità umana, della libertà, della democrazia, dell’uguaglianza e dello Stato di diritto quali pilastri comuni agli Stati membri, imponendo all’Unione e ai suoi organi il dovere di vigilare sulla loro effettiva attuazione. Se, in passato, tali valori trovavano concretizzazione principalmente sul piano politico-istituzionale, oggi essi si proiettano anche sul terreno digitale, dove si gioca una parte sempre più significativa della partecipazione civica e del processo di formazione dell’opinione pubblica.

Il Dsa, in questo contesto, si configura come strumento di tutela della democrazia digitale, capace di operare una sintesi fra libertà di espressione e integrità dello spazio informativo. L’obiettivo non è soltanto contrastare la disinformazione, ma preservare le condizioni strutturali che rendono possibile il confronto democratico, così come concepito nel paradigma costituzionale europeo. Tale prospettiva si avvicina all’idea, tipica della *militant democracy* – sviluppata originariamente da Karl Loewenstein negli anni Trenta del Novecento – secondo cui gli ordinamenti democratici possono, e talvolta devono, adottare misure di autodifesa per garantire la sopravvivenza delle proprie istituzioni e dei propri valori fondamentali²⁷.

Nel contesto europeo, tuttavia, questa logica di protezione non si traduce in forme di compressione diretta della libertà politica o associativa – come accaduto nei modelli nazionali storici – bensì in strategie di regolazione preventiva e di trasparenza sistemica volte a ridurre le distorsioni del discorso pubblico.

26. Sulla “democrazia protetta” cfr., *ex multis*, S. Ceccanti, *Le democrazie protette e semi-protette da eccezione a regola. Prima e dopo le Twin Towers*, Giappichelli, 2004; S. Ceccanti, F. Ferroni, *Democrazia protetta*, in *Dig. disc. pubbl.*, VI Agg., Utet, 2015, p. 55 ss.

27. K. Loewenstein, *Militant Democracy and Fundamental Rights*, in *Political Science Review*, 1937, p. 417 ss.

Il Dsa, attraverso l'imposizione di obblighi di diligenza e *accountability* ai fornitori di servizi digitali, in particolare alle *very large online platforms*, interviene sulle infrastrutture informative che condizionano la visibilità, la selezione e la circolazione dei contenuti, riconoscendo che il libero esercizio della libertà di espressione dipende oggi anche dall'architettura tecnologica che ne regola l'accesso e la diffusione²⁸.

L'innovazione concettuale di questo approccio emerge con chiarezza se letta alla luce dell'art. 3 del Primo Protocollo addizionale alla Cedu, il quale sancisce il diritto a libere elezioni come espressione essenziale del principio democratico²⁹. La qualità e trasparenza dello spazio informativo, di cui il Dsa si fa garante, diventano in tal modo condizioni sostanziali per l'effettivo esercizio dei diritti politici, riconnettendo la tutela della libertà di espressione alla protezione del processo democratico in sé. In altri termini, il diritto a partecipare alla vita politica e a formare liberamente un'opinione risulta inseparabile dalla garanzia che tale opinione si formi in un ambiente comunicativo non manipolato, dove l'informazione sia verificabile, pluralista e non alterata da meccanismi opachi di amplificazione o censura algoritmica.

Da questa prospettiva, il Dsa può essere considerato come un tentativo embrionale di costruire una "democrazia protetta" sovranazionale, fondata non su logiche di interdizione, ma su principi di *governance* e co-responsabilità. La "protezione", infatti, non assume la forma di una restrizione della libertà di parola, bensì quella di una regolazione delle condizioni tecnologiche e informative che ne determinano l'effettivo esercizio. Ciò segna un passaggio dalla concezione negativa della libertà – intesa come assenza di vincoli – a una visione relazionale e sostanziale, in cui l'intervento del potere (privato) diventa funzionale alla rimozione degli ostacoli che compromettono la qualità del discorso democratico.

In questa prospettiva, il costituzionalismo digitale fornisce la chiave interpretativa per comprendere la portata sistemica del Dsa. Esso si fonda sull'idea che la costituzionalità del digitale non derivi soltanto dall'applicazione dei diritti fondamentali esistenti al *cyberspazio*, ma dalla creazione di nuovi strumenti istituzionali e procedurali in grado di bilanciare il potere privato delle piattaforme e la sovranità democratica. Il Dsa, con i suoi meccanismi di *risk*

28. Sul Dsa cfr., per tutti, G. De Gregorio, *The Digital Services Act: A Paradigmatic Example of European Digital Constitutionalism*, in *Diritticomparati.it*, 17 maggio 2021.

29. Che così testualmente recita: «Le Alte Parti contraenti si impegnano a organizzare, a intervalli ragionevoli, libere elezioni a scrutinio segreto, in condizioni tali da assicurare la libera espressione dell'opinione del popolo sulla scelta del corpo legislativo». Si v., anche, le interessanti argomentazioni contenute nella recente pronuncia della Corte Edu del 22 luglio 2025, *Bradshaw et alii c. Regno Unito*, Appl. 15653/2022.

assessment, algorithmic transparency e *due diligence*, introduce per la prima volta un modello di responsabilità condivisa fra operatori economici e autorità pubbliche, volto a garantire la sostenibilità democratica dell'ecosistema digitale.

Il Dsa, pur confermando sostanzialmente il quadro delle esenzioni di responsabilità per gli intermediari già delineato dalla Direttiva 2000/31/CE sul commercio elettronico, innova profondamente sotto il profilo procedurale, introducendo un insieme articolato di obblighi di trasparenza, diligenza e tutela degli utenti. L'attenzione del legislatore europeo si concentra, in particolare, sulla moderazione e sulla selezione dei contenuti *online*, ambiti nei quali si manifesta con maggiore evidenza la dimensione giuridica del cosiddetto *potere privato delle piattaforme*³⁰. Si tratta, infatti, di un potere idoneo a incidere direttamente sull'esercizio di un diritto fondamentale della persona, quale la libertà di manifestazione del pensiero, ponendo così questioni cruciali di legittimazione e di garanzia.

In primo luogo, il Dsa impone ai prestatori di servizi intermediari l'obbligo di rendere pienamente trasparenti le regole che disciplinano l'utilizzo delle proprie piattaforme. L'art. 12, par. 1, stabilisce che tali operatori debbano inserire nei termini e condizioni del servizio informazioni chiare e accessibili sulle eventuali restrizioni all'uso, comprese le politiche, le procedure e gli strumenti impiegati per la moderazione dei contenuti, sia attraverso processi algoritmici sia mediante verifica umana. Tale disposizione assume un rilievo sistemico: essa introduce, di fatto, un principio di "spiegabilità *ex ante*" dei meccanismi algoritmici nei confronti degli utenti, ponendo le basi di una trasparenza preventiva che anticipa e accompagna l'esercizio del potere privato di moderazione. Inoltre, la disposizione richiede che le attività di moderazione siano condotte secondo criteri di diligenza, obiettività e proporzionalità, criteri che rinviano a una logica tipicamente pubblicistica di esercizio del potere in funzione di garanzia.

Il legislatore europeo rafforza poi il quadro di responsabilità imponendo agli operatori – con esclusione delle microimprese – di pubblicare una relazione annuale (art. 13, par. 1, lett. *c*) contenente informazioni dettagliate sulle attività di moderazione dei contenuti, comprese quelle avviate di propria iniziativa. Si tratta di un meccanismo di rendicontazione periodica che consolida la natura "pubblica" delle funzioni esercitate dalle piattaforme e che introduce una forma di *accountability* strutturale nel settore privato digitale.

Un ulteriore tassello è rappresentato dal meccanismo di notifica e azione (*notice and action*) previsto dall'art. 14, che obbliga i prestatori di servizi di

30. Così A. Lamberti, *Costituzionalismo digitale, poteri delle piattaforme, intelligenza artificiale e democrazia*, cit., p. 1093.

hosting a predisporre procedure chiare per la segnalazione dei contenuti illegali e per la comunicazione agli utenti delle decisioni adottate a seguito di tali segnalazioni. L'art. 15, par. 1 e 2, rafforza la tutela procedurale dell'utente imponendo alla piattaforma, in caso di rimozione di contenuti o disabilitazione dell'accesso, l'obbligo di informare tempestivamente il destinatario e di motivare in modo chiaro e specifico la decisione adottata. Questo obbligo di motivazione, di per sé inedito nell'ambito dei rapporti contrattuali privati, costituisce un autentico strumento di garanzia contro l'arbitrarietà dell'azione privata e può essere letto come segno di una progressiva "amministrativizzazione" del potere privato, vale a dire di una trasposizione nel settore digitale di garanzie e principi originariamente concepiti per il controllo dell'attività pubblica³¹.

Il Dsa prevede, inoltre, la creazione di meccanismi di reclamo interni (art. 17), cui gli utenti possono ricorrere contro le decisioni delle piattaforme, con l'esplicita esclusione del ricorso a strumenti interamente automatizzati per la loro risoluzione. In caso di insoddisfazione, l'utente può rivolgersi a un organismo indipendente di risoluzione extragiudiziale delle controversie (art. 18, par. 1), ferma restando la possibilità di adire l'autorità giudiziaria competente. Questo doppio livello di tutela – interna ed esterna – riflette l'intento del legislatore europeo di costruire un sistema di giustizia digitale ibrido, nel quale il potere privato sia sottoposto a forme di controllo ispirate ai principi di giusto procedimento e imparzialità, tipiche della sfera pubblica. Ne risulta una configurazione del rapporto tra utente e piattaforma che si avvicina a quella del cittadino di fronte all'Amministrazione, in cui il diritto di difesa e la trasparenza procedurale diventano elementi costitutivi dell'equilibrio dei poteri.

Particolarmente significativa, poi, è la disciplina delle piattaforme *online* di dimensioni molto grandi. L'art. 26 impone loro di effettuare valutazioni periodiche del rischio sistemico, prendendo in considerazione la diffusione di contenuti illegali, gli effetti negativi sull'esercizio dei diritti fondamentali (inclusa la libertà di espressione e di informazione), le manipolazioni intenzionali dei servizi, l'uso automatizzato o non autentico delle piattaforme, nonché i possibili impatti sulla salute pubblica, la tutela dei minori, il dibattito civico, i processi elettorali e la sicurezza pubblica. Si tratta di un passaggio cruciale che riconosce formalmente il legame tra integrità dell'ambiente informativo e tutela della democrazia, trasformando la governance algoritmica in un tema di diritto costituzionale sostanziale.

31. P. Piluso, *Poteri privati, diritti fondamentali e interessi legittimi nell'era delle piattaforme. Costituzionalismo digitale e "amministrativizzazione" del potere privato*, in *ConsultaOnline*, n. 2/2025, p. 1047 s.

A garanzia dell'effettività di tali obblighi, l'art. 28 prevede che le grandi piattaforme siano sottoposte ad *audit* indipendenti annuali, da effettuare a loro spese, da parte di soggetti terzi qualificati. Tale sistema di controllo esterno introduce nel diritto europeo del digitale una logica di verifica pubblica della conformità privata, rafforzando la cultura della responsabilità algoritmica.

Infine, il Dsa affida l'attuazione e il controllo del sistema a autorità nazionali competenti, eventualmente di nuova istituzione, dotate anche di poteri sanzionatori (art. 42). Le sanzioni pecuniarie, parametrare su una percentuale significativa del fatturato annuo, testimoniano la volontà dell'Unione di attribuire al Dsa un effettivo potere deterrente e di garantire la piena efficacia delle regole nel mercato digitale europeo.

Nel complesso, la disciplina delineata dal Dsa non si limita a introdurre nuovi obblighi di trasparenza e responsabilità per gli operatori privati, ma disegna una nuova architettura della *governance* digitale, fondata sulla proceduralizzazione del potere e sull'ibridazione tra garanzie pubblicistiche e rapporti contrattuali privati. È in questa prospettiva che il Dsa si configura non soltanto come uno strumento regolatorio, ma come un laboratorio di costituzionalizzazione del *cyberspazio* europeo, dove i diritti fondamentali non sono solo tutelati, ma resi effettivi attraverso meccanismi di trasparenza, motivazione e controllo.

Tuttavia, non sfugge che il legislatore europeo, pur concentrandosi prevalentemente su misure di carattere procedurale, conserva di fatto un ampio margine di discrezionalità per le piattaforme. Ciò emerge chiaramente dalla scelta di equiparare i contenuti illeciti a quelli potenzialmente "dannosi", ma formalmente leciti, attribuendo in entrambi i casi alle piattaforme il potere di rimuovere o limitare la circolazione dei contenuti senza l'intervento preventivo di un'autorità giurisdizionale. Si tratta di un paradosso evidente, poiché le autorità pubbliche, quando adottano provvedimenti con effetti analoghi, sono invece vincolate a un bilanciamento formale con i diritti fondamentali dei cittadini.

In sostanza, l'impianto di fondo rimane sostanzialmente invariato: il potere di moderare i contenuti resta nelle mani delle piattaforme, cui viene delegata la responsabilità di decidere quali contenuti ammettere o rimuovere, secondo le condizioni generali di contratto da esse unilateralmente stabilite o in conformità a codici di condotta che, comunque, hanno contribuito a elaborare.

Per di più – ed è questo l'aspetto che in questa sede più rileva – il Dsa stabilisce che le piattaforme *online* di dimensioni molto grandi debbano dotarsi di un sistema interno di identificazione, analisi e valutazione dei rischi connessi alla progettazione e al funzionamento dei loro servizi, inclusi i meccanismi algoritmici. Tali rischi riguardano, tra l'altro, potenziali effetti negativi, concreti

o prevedibili sul dibattito civico e sui processi elettorali. L'obiettivo è che, qualora si manifestino tali rischi sistemici, la piattaforma possa adottare misure adeguate, proporzionate ed efficaci per mitigarne gli impatti.

Inoltre, il Dsa prevede un meccanismo straordinario per gestire le cosiddette "crisi di sistema", ossia situazioni eccezionali che comportano una grave minaccia per la sicurezza pubblica nell'Unione europea. In tali casi, la Commissione europea, su raccomandazione del Comitato europeo per i servizi digitali, può imporre alle piattaforme di grandi dimensioni di attivare interventi urgenti, ad esempio modificando temporaneamente il funzionamento dei sistemi algoritmici o pubblicitari pertinenti, al fine di contenere i rischi emergenti (cfr. considerando 91 e art. 36 Dsa)³².

5. Conclusioni. Un "nuovo" costituzionalismo fondato sul binomio verità-libertà

Il costituzionalismo digitale agisce anche come garante del pluralismo, principio cardine del costituzionalismo liberal-democratico. La disinformazione prospera laddove l'ecosistema informativo è fragile o polarizzato. Garantire un pluralismo effettivo – non solo di fonti, ma anche di architetture comunicative – significa impedire che una piattaforma o un attore monopolizzi il discorso pubblico.

Il pluralismo digitale, quindi, non è soltanto una questione di mercato, ma una condizione di possibilità della verità pubblica. Il costituzionalismo digitale, attraverso regole sull'interoperabilità, sull'accesso ai dati e sulla neutralità delle piattaforme, può prevenire la formazione di spazi comunicativi chiusi o dominati da logiche di manipolazione delle coscienze.

Pur con i suoi limiti applicativi e con la necessità di un consolidamento interpretativo, il Dsa sembra dunque inaugurare una nuova fase del costituzionalismo europeo, in cui la protezione della democrazia non è più soltanto un compito degli Stati membri, ma una responsabilità diretta dell'Unione, esercitata attraverso strumenti normativi che anticipano la costruzione di una vera e propria sfera pubblica europea. In tale ottica, il Dsa può essere letto come una

32. È quanto accaduto a seguito dell'annullamento del primo turno delle elezioni presidenziali romene del 2024, a causa delle interferenze esterne sul procedimento elettorale attraverso il social network TikTok. Sul caso si rinvia a S. Sassi, A. Sterpa, *La Corte costituzionale della Romania difende la democrazia liberale dalla disinformazione. Prime note sulla sentenza n. 32 del 6 dicembre 2024*, in *federalismi.it*, n. 14, 2025, p. 162 ss., F. Rosa, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025, p. 53 ss.

risposta costituzionale alla crisi della democrazia informazionale, un tentativo di tradurre in diritto positivo l'idea che la libertà di espressione, per essere autenticamente democratica, debba essere non solo garantita, ma resa effettivamente praticabile all'interno di un ambiente informativo trasparente, pluralista e resistente alla manipolazione.

Rivelare le falsità della propaganda e garantire all'individuo gli strumenti necessari per formarsi un giudizio autonomo rappresenta il terreno naturale su cui il costituzionalismo può contrastare i tentativi di sovvertimento interno senza rinunciare ai propri tratti essenziali, preservando così la sua superiorità rispetto ai modelli c.d. illiberali. Per riuscire davvero a farlo, ci si potrebbe domandare se, data l'assoluta novità del fenomeno e la sua difficile riconducibilità alle tradizionali categorie del costituzionalismo liberal-democratico del Novecento, non sia forse necessario ipotizzare un cambio di paradigma: se finora ci si è concentrati sul rapporto tra autorità e libertà, non dovrebbe oggi l'attenzione spostarsi anche sul rapporto tra verità (*rectius*: le molteplici verità) e libertà³³? Il costituzionalismo digitale delineato a livello europeo, in questo senso, non limiterebbe davvero la libertà d'opinione in nome della lotta alla disinformazione, ma mirerebbe piuttosto a costruire un ambiente comunicativo affidabile, nel quale la libertà possa forse esercitarsi in modo più informato.

L'obiettivo – in altri termini – non è sostituire la verità con la “verità di Stato” o di piattaforma, ma creare le condizioni di possibilità della verità come pratica democratica. In questo senso, il costituzionalismo digitale è il presupposto materiale della libertà e della verità, entrambe coesistenti per la tenuta di un ordinamento autenticamente democratico.

33. Un rapporto che – accanto a quello che intercorre tra autorità e libertà, inquadrato e studiato nella sua dimensione ontologicamente dialettica (come lo inquadrava C. Mortati, *Istituzioni di diritto pubblico*, tomo I, Cedam, 1975, p. 133 ss., a proposito delle forme di Stato) – si fonda sulla complementarità e sinallagmaticità dei due termini: la verità è funzionale alla libertà tanto quanto la libertà è il presupposto naturale della verità per il mantenimento dell'ordine democratico.

Sezione II

INFORMAZIONE DIGITALE
E TUTELA DELL'INDIVIDUO

POTERI PUBBLICI, POTERI PRIVATI E GOVERNO COSTITUZIONALE DELL'INFORMAZIONE DIGITALE

*Giulio Enea Vigevani**

SOMMARIO: 1. Assetti di potere ed equilibri nella sfera pubblica “pentagonale” – 2. Il mercato dell’informazione digitale – 3. Informazione digitale e libertà fondamentali – 4. Modelli regolatori europei: trasparenza o censura?

1. Assetti di potere ed equilibri nella sfera pubblica “pentagonale”

La trasformazione digitale degli ultimi decenni ha profondamente mutato le forme dell’informazione e le strutture dei poteri che la governano. Per lungo tempo, la comprensione della sfera pubblica poteva essere ricondotta a un modello triangolare, nel quale il potere politico, il sistema dei media e l’opinione pubblica interagivano secondo equilibri variabili ma riconoscibili.

Equilibri certamente complessi, determinati non solo dalle regole costituzionali o legislative, ma anche dai reciproci condizionamenti, dalle ingerenze indirette della politica nelle decisioni editoriali e dall’uso da parte di grandi gruppi imprenditoriali dei media di cui sono proprietari come strumenti di pressione sul potere politico.

Si trattava di un’impostazione che, pur con i suoi limiti, rendeva possibile individuare dinamiche di influenza e contropoteri relativamente stabili, almeno nelle democrazie mature, ove i media sono stati capaci nel complesso di esercitare una funzione di controllo dei poteri pubblici.

Oggi questo schema risulta inadeguato. La trasformazione tecnologica e l’emergere di nuovi attori privati, dotati di inedite capacità di orientamento informativo e simbolico, hanno modificato radicalmente la morfologia del potere.

I nuovi soggetti dell’ecosistema digitale – le piattaforme online e, più recentemente, i sistemi di intelligenza artificiale generativa – hanno introdotto

* Professore ordinario di Diritto costituzionale e pubblico presso l’Università di Milano-Bicocca.

ulteriori centri di influenza che rendono più complesso il funzionamento della democrazia costituzionale.

Tale mutamento può essere rappresentato come un progressivo passaggio da una relazione a tre (un triangolo con ai vertici potere politico, sistema dell'informazione e opinione pubblica) a un quadrangolo del potere informativo, nel quale accanto ai media tradizionali si collocano gli intermediari digitali, i cosiddetti "signori della rete".

Più recentemente, l'irrompere dell'intelligenza artificiale generativa ha reso pertinente l'immagine di un vero e proprio pentagono dell'informazione, in cui i sistemi algoritmici non si limitano a veicolare contenuti, ma li generano, modificano e adattano in tempo reale, su scala industriale e senza costi marginali.

Un pentagono tutt'altro che in equilibrio, in cui la bilancia del potere tende a pendere verso quello politico, specie quando questo si fa "minaccioso" e ha strumenti di persuasione e ricatto nei confronti dei media vecchi e nuovi.

In questo nuovo scenario, la stampa appare il soggetto che più si è indebolito. La crisi strutturale dell'editoria tradizionale ha inciso sulla sua capacità di svolgere quel ruolo di controllo diffuso che ha rappresentato per decenni uno dei cardini della società aperta. La funzione di vigilanza sul potere, di denuncia degli abusi e di alimentazione della dialettica democratica è diventata più difficile da esercitare, non per un mutamento normativo, ma per un insieme di fattori strutturali.

La crisi economica del settore editoriale, la progressiva e talvolta inevitabile dipendenza da finanziamenti pubblici o da grandi gruppi economici, nonché la fragilità del modello dell'editore puro, mestiere in perdita in tutto il mondo e dunque sempre più raro anche a livello internazionale, hanno reso molte imprese più vulnerabili alle pressioni esterne.

D'altro lato, gli stessi giganti del web, per la loro struttura monopolistica e strettamente legata al potere pubblico, non paiono in alcun modo né potere, né volere svolgere quel ruolo di "guardiano delle libertà" tipico della stampa.

Il rapido allineamento delle piattaforme ai diktat di Trump e i repentini cambiamenti di rotta mi pare mostrino la pericolosità dell'alleanza tra potere politico e poteri privati e dunque i dubbi nell'attribuire a questi soggetti compiti di controllo sul discorso pubblico che tradizionalmente spetterebbero alle autorità pubbliche.

2. Il mercato dell'informazione digitale

Ciò premesso, come si può delineare un mercato dell'informazione digitale che sia al contempo libero, plurale e meno esposto a quelle pratiche di disinformazione particolarmente insidiose nel processo democratico?

Muovo da un interrogativo preliminare: quali sono state le innovazioni più rilevanti, sul piano costituzionale, introdotte dall'avvento delle piattaforme nel "mercato dell'informazione"? A mio avviso, se ne possono individuare due, strettamente connesse tra loro.

Da un lato, si è assistito alla progressiva accettazione dell'anonimato e della mancanza di trasparenza sull'origine delle informazioni come tratti ordinari del discorso pubblico digitale, con la conseguenza di esporre lo spazio informativo non tanto alle "fake news", quanto piuttosto a forme di disinformazione pianificata.

Nella tradizione costituzionale liberale, il nesso tra libertà e responsabilità era fondato sull'idea che chi esprime un'opinione dovesse assumersene la paternità. L'ampia diffusione di comunicazioni anonime e di identità digitali difficilmente riconducibili a soggetti reali ha incrinato tale principio, favorendo la circolazione di contenuti manipolati, campagne coordinate e, appunto, forme di disinformazione difficili da contrastare con gli strumenti giuridici tradizionali. La conseguenza è una progressiva perdita di *accountability* che indebolisce il senso stesso di responsabilità personale nel dibattito pubblico.

Dall'altro lato, si è consolidata l'accettazione della "moderazione dei contenuti", non solo quale facoltà discrezionale riconosciuta alle piattaforme per assicurare il rispetto delle proprie *policies*, ma anche quale strumento al quale il potere pubblico può fare ricorso per "ripulire" il web da contenuti irritanti, di cattivo gusto o anche falsi, benché non necessariamente illeciti, arrivando persino a incidere sulla preselezione delle notizie e delle idee effettivamente diffondibili all'interno della società attraverso sistemi algoritmici di categorizzazione dei contenuti.

In altri termini, la moderazione dei contenuti, originariamente pensata come strumento di gestione interna delle comunità digitali, si è trasformata in un potere regolativo di fatto, esercitato in modo opaco, talvolta in risposta a pressioni governative e senza garanzie procedurali. In questo contesto, il rischio non deriva soltanto dall'autonomia decisionale delle piattaforme, ma soprattutto dalla possibile convergenza tra poteri pubblici e poteri privati nella selezione e nella rimozione dei contenuti. Una tale convergenza può comportare forme di interferenza nel dibattito pubblico non compatibili con i principi fondamentali del costituzionalismo liberale.

Non è un caso che, in diversi ordinamenti, sia emersa l'esigenza di indagare più approfonditamente l'interazione fra autorità pubblica e piattaforme nella gestione del discorso politico online. Episodi recenti testimoniano una crescente preoccupazione per possibili interferenze e per il rischio di "censure soft", meno visibili ma non per questo meno incisive.

Tutto ciò finisce per sovvertire alcuni dei postulati fondamentali delle Costituzioni liberal-democratiche. In fondo, l'equilibrio tra autorità e libertà che il costituzionalismo classico ha delineato in materia di manifestazione del pensiero può essere riassunto, con brutale sintesi, in questi termini: “sei libero, ma devi dichiarare chi sei e assumerti la responsabilità di ciò che affermi”.

In termini più raffinati, sin dalle prime carte settecentesche, il costituzionalismo pretende l'estromissione dei poteri pubblici dalla selezione delle notizie e dalla valutazione della loro qualità – salvi i casi di illecito previsti dalla legge – attraverso il divieto di censura e la previsione di limiti stringenti al sequestro degli stampati. Contestualmente, tuttavia, esso ha imposto il “divieto di stampa clandestina”, l'obbligo di indicare un responsabile e la possibilità di ricostruire gli assetti proprietari (e dunque gli interessi) sottesi a ciascun mezzo di informazione.

E credo che il compito del costituzionalista – specie italiano, avendo la nostra Carta accolto in pieno il modello “liberale” della società aperta e del valore del libero confronto delle idee – sia quello di preservare tali principi e tale equilibrio, con i necessari adattamenti, nel mondo dell'informazione digitale.

Modello – lo dico per inciso – che non è nihilista per quanto attiene alla verità e alla preservazione di un habitat della sfera pubblica “civile”. Esso rifiuta, però, l'idea di verità assolute calate dall'alto e si fonda, per dirlo con Häberle, «su verità provvisorie, rivedibili, che assume in linea di principio al plurale e non al singolare e “per decreto”». E, soprattutto, l'accertamento della verità avviene attraverso procedure rigorose e predeterminate dalla legge.

In quest'ottica, è necessario valorizzare anche nel contesto digitale gli istituti di garanzia tradizionalmente previsti dalla Costituzione per la stampa: la riserva di legge e di giurisdizione, il divieto di censure preventive, la disciplina garantista del sequestro, e quindi anche dell'oscuramento online, almeno con riferimento alle informazioni riconducibili a un soggetto identificabile. Parallelamente, occorre frenare quelle tendenze che, seppur animate dalle migliori intenzioni, mirano a costruire un ordinamento separato per il mondo digitale, con regole che derogherebbero ai principi consolidati della nostra tradizione.

Si discute molto dell'utilità della nozione di “costituzionalismo digitale”, inteso come l'idea secondo cui le coordinate del contesto digitale non potrebbero essere regolate dagli stessi principi che disciplinano la libertà di espressione nel “mondo degli atomi”. Il *digital constitutionalism* mette in luce un'esigenza autentica: quella di sottoporre a limiti costituzionali poteri privati che oggi incidono in modo determinante sulla struttura del discorso pubblico. Le

piattaforme e, già ora, chi detiene le chiavi dei sistemi di intelligenza artificiale esercitano poteri di qualificazione, selezione, promozione e finanche produzione delle informazioni che possono influenzare la formazione dell'opinione pubblica. Ignorare tale fenomeno significherebbe rinunciare a una parte essenziale dell'effettività delle libertà fondamentali nell'era digitale.

Il rischio, tuttavia, è quello di affermare un paradigma separato che finisca per erodere la forza garantista del costituzionalismo classico. Le teorie che propongono una "costituzione della rete" o un ordine normativo autonomo del digitale finiscono con il trascurare l'esigenza di preservare principi consolidati quali il divieto di censura preventiva statale, la centralità della legge, la garanzia giurisdizionale e la responsabilità degli attori dell'informazione.

In tale prospettiva, nel tentativo di contenere il nuovo potere che le piattaforme digitali eserciterebbero di fatto in sostituzione dello Stato, si tende talvolta a dimenticare che le insidie più gravi provengono proprio dal potere pubblico, o più precisamente dall'alleanza tra poteri pubblici e poteri privati. Per usare le parole del Presidente Mattarella, il disegno dell'art. 21 Cost. mira anzitutto a porre un freno alle «mai abbandonate tentazioni dei poteri pubblici di fissare limiti agli spazi di libertà di informazione».

In questa prospettiva, la risposta più coerente con la tradizione costituzionale non consiste nella creazione di un ordinamento autonomo del digitale, ma nella estensione degli strumenti classici della garanzia ai nuovi ambiti di esercizio del potere privato. Ciò implica l'introduzione di obblighi di trasparenza algoritmica, la definizione di procedure di contestazione e ricorso, la tracciabilità degli interventi di moderazione e la possibilità di ricondurre le decisioni delle piattaforme entro un quadro giurisdizionale effettivo.

La vera sfida non risiede dunque nella costruzione di un nuovo costituzionalismo, ma nella capacità di aggiornare i fondamenti del costituzionalismo liberale in un contesto in cui i confini dell'autorità e della libertà si ridefiniscono continuamente.

Ciò non esclude la necessità di individuare regole e meccanismi che impongano alle piattaforme di accertare e indicare chi sia l'autore di un contenuto – specificando se si tratti di una persona fisica, di una persona giuridica, di un'autorità governativa, di un programma automatico come un bot o di un utente anonimo – e quale sia la sua provenienza, di garantire la riconoscibilità dei contenuti prodotti esclusivamente tramite sistemi algoritmici nonché di assicurare la trasparenza degli algoritmi che classificano, moderano ed eventualmente rimuovono le informazioni online (in questa direzione che si collocano interventi come il *Digital Services Act* e l'*AI Act*, che almeno per questi aspetti mirano a riconciliare l'innovazione tecnologica con la continuità dei principi fondamentali dello Stato costituzionale).

3. Informazione digitale e libertà fondamentali

Il nodo centrale riguarda l'equilibrio tra libertà di informazione e gli altri diritti o interessi che possono essere compromessi dal "cattivo uso" di tale libertà nell'ambiente digitale. La domanda di fondo è semplice: ciò che è lecito offline deve essere lecito anche online? Oppure è necessario un diverso "statuto dell'informazione in rete", perché il discorso pubblico digitale sarebbe "più pericoloso", dominato da manipolazioni e falsità, anche per il fatto che chiunque – e non solo i professionisti – può generare contenuti?

Per rispondere occorre fare un passo indietro. Nell'internet delle origini, l'elemento di maggiore novità consisteva nel fatto che i siti non erano "mezzi finiti": a differenza dei giornali o delle televisioni, non avevano l'obbligo di selezionare i contenuti, disponendo di uno spazio potenzialmente illimitato. Ciò ha indubbiamente abbattuto i limiti tradizionali del sistema informativo, dando vita a uno spazio aperto nel quale ogni cittadino poteva esprimere le proprie idee e confrontarsi con quelle altrui.

Questa evoluzione ha condotto – dopo alcuni entusiasmi anarcoidi e successive pulsioni censorie dettate dal consueto timore verso il nuovo – a un punto di equilibrio che può essere riassunto nel già citato parallelismo tra ciò che è lecito offline e ciò che è lecito online. Sia negli Stati Uniti sia negli ordinamenti europei si è infatti respinta l'idea di una regolamentazione dei contenuti specifica per il cyberspazio, così come l'ipotesi di una responsabilità editoriale generalizzata dei provider, nella logica di evitare che il potere pubblico eserciti un controllo più incisivo sui contenuti digitali rispetto a quelli del mondo fisico, avvalendosi della *longa manus* dei fornitori di servizi.

Cosa cambia con la "piattaformizzazione" della sfera pubblica? La moderazione dei contenuti viene oggi esercitata da un numero ristretto di soggetti, prevalentemente attraverso sistemi algoritmici. Per ragioni eminentemente pratiche, sono le piattaforme a decidere cosa possa rimanere online, incidendo direttamente su situazioni che coinvolgono diritti individuali (come diritto all'oblio e *privacy*) o interessi collettivi (come i casi di *hate speech*).

Si è inoltre sviluppata una tendenza sempre più marcata verso un rapporto dialogico e cooperativo tra piattaforme e poteri pubblici, secondo una logica di co-regolamentazione. Tale modello mira a garantire la pronta rimozione di contenuti considerati pericolosi per gli individui o per la democrazia nel suo complesso. Questa evoluzione riflette la presa d'atto che gli Stati non sono in grado di contrastare autonomamente, nel mondo digitale, la cosiddetta "zona grigia" ossia quei fenomeni "pericolosi" ma non necessariamente penalmente illeciti. Da qui l'idea di un sistema in cui gli ordinamenti politici fissano i fini e

le piattaforme mettono a disposizione gli strumenti necessari a realizzarli (algoritmi, risorse economiche, rapidità decisionale).

È precisamente in questo contesto che emergono i rischi di elusione delle garanzie costituzionali. La riserva di legge e di giurisdizione risultano aggirate quando l'identificazione dei contenuti come illeciti non deriva più esclusivamente da una fonte normativa statale ma da atti di *soft law* spesso opachi, frutto della cooperazione tra autorità pubbliche e soggetti privati nella co-regolamentazione. Parallelamente, manca un organo terzo e imparziale che valuti la legittimità delle limitazioni adottate. Allo stesso tempo, il divieto di censura preventiva viene sostanzialmente eluso attribuendo agli algoritmi – opportunamente addestrati – il compito di incidere in concreto sulla libertà di espressione e sul diritto dei cittadini a essere informati.

La conseguenza, paradossale rispetto alle concezioni originarie della rete, è che gli spazi della libertà di manifestazione del pensiero tendono oggi a contrarsi rispetto ai media tradizionali: viene meno, in altre parole, quel parallelismo tra ciò che è lecito offline e ciò che è lecito online che aveva caratterizzato la fase iniziale dello sviluppo di internet.

A queste trasformazioni si aggiunge oggi un ulteriore elemento: l'emergere dell'intelligenza artificiale generativa come nuovo attore della sfera pubblica. Per la prima volta nella storia, una parte del discorso pubblico è prodotta non da esseri umani ma da sistemi algoritmici in grado di generare testi, immagini e contenuti complessi. Questo fenomeno incide sul rapporto stesso tra autore e responsabilità, pilastro su cui si fonda una parte significativa della disciplina costituzionale dell'informazione. Ciò altera profondamente il rapporto tra libertà e responsabilità, perché il sistema costituzionale era costruito sull'idea che ogni opinione avesse un autore identificabile. L'automazione della produzione informativa rende più difficile attribuire responsabilità, favorisce la diffusione di contenuti manipolati e introduce nuovi rischi di disinformazione organizzata.

L'ingresso dell'intelligenza artificiale generativa non deve però essere interpretato come un fattore esclusivamente negativo: essa può rappresentare uno strumento di democratizzazione dell'accesso all'informazione e di ampliamento delle capacità espressive dei cittadini. Tuttavia, perché ciò avvenga, è necessario che l'ordinamento assicuri condizioni di trasparenza e tracciabilità, evitando che le dinamiche di produzione algoritmica finiscano per saturare lo spazio pubblico di contenuti indistinguibili tra loro.

Il risultato complessivo di questo scenario è che il parallelismo tra ciò che è lecito offline e ciò che è lecito online tende a sfumare, ma in senso contrario rispetto alle aspettative iniziali. La rete, nata come spazio di libertà privo di limiti strutturali, rischia di divenire un ambiente nel quale gli interventi, siano essi

pubblici o privati, comprimono il diritto alla libera manifestazione del pensiero in misura maggiore rispetto ai mezzi tradizionali. Ciò rende più urgente una riflessione sul ruolo delle garanzie costituzionali e sulla loro capacità di adattarsi alle nuove forme della comunicazione digitale.

4. Modelli regolatori europei: trasparenza o censura?

Una domanda preliminare è se questa tendenza – la crescente cooperazione tra poteri pubblici e piattaforme e il correlato rischio di compressione degli spazi di libertà – si stia accentuando nella “nuova stagione costituzionale dell’Unione”, rappresentata dal *Digital Services Act* (*Dsa*) e dal *Digital Markets Act* (*Dma*) del 2022, dall’*European Media Freedom Act* (*Emfa*), dall’*Artificial Intelligence Act* (*AI Act*) e dal Regolamento sulla trasparenza e sul *targeting* della pubblicità politica del 2024.

È difficile formulare un giudizio complessivo, ma la sensazione è che, almeno in parte, si sia registrato un freno. In particolare, il *Digital Services Act* attribuisce uno spazio relativamente limitato alla co-regolamentazione, preferendo ricorrere a strumenti più tradizionali e, sotto questo profilo, meglio allineati con le garanzie costituzionali in materia di libertà di espressione: mi riferisco soprattutto ai numerosi obblighi di trasparenza imposti alle piattaforme e ai sistemi di intelligenza artificiale.

Vi sono tuttavia alcune eccezioni rilevanti che segnalano un rafforzamento della cooperazione tra governi e piattaforme, con un corrispondente aumento dell’influenza delle autorità pubbliche sui contenuti. Emblematico è il meccanismo di risposta alle crisi previsto dall’art. 48 del *Dsa* (i cosiddetti “Protocolli di crisi”).

Il *Digital Markets Act*, invece, mira principalmente a favorire un ecosistema digitale più plurale e competitivo, condizione necessaria – anche se non sufficiente – per un sistema informativo equilibrato.

Un ulteriore tema cruciale riguarda il sostegno, anche economico, all’informazione di qualità. In questa direzione si muove l’*European Media Freedom Act*, che introduce garanzie volte a prevenire interferenze politiche nelle decisioni editoriali, ad assicurare la trasparenza dei finanziamenti e a contrastare il fenomeno del cosiddetto *media capture*, tanto nei confronti dei fornitori di servizi di media pubblici quanto privati.

Il Regolamento sulla pubblicità politica, inoltre, enuncia una serie di principi di rilievo: il diritto dei cittadini a riconoscere la pubblicità politica e i relativi finanziamenti; l’obbligo di etichettatura dei messaggi politici; limiti stringenti al micro-targeting; il divieto di pubblicità politica proveniente da Paesi terzi

nei mesi precedenti le elezioni europee (una misura chiaramente rivolta a contrastare interferenze esterne, *in primis* russe). La logica di fondo appare condivisibile – al di là della reazione irritata di alcune piattaforme, come Meta, che ha annunciato di non voler ospitare pubblicità politiche – anche se non manca un certo tasso di burocratizzazione. È comunque significativo che non si intervenga sui contenuti, ma sulla trasparenza circa la provenienza dei messaggi.

Quanto all' *AI Act*, sembra scongiurato il rischio che esso possa favorire derive censorie da parte dei poteri pubblici. La preoccupazione principale del legislatore è elevare la trasparenza a requisito fondamentale per i sistemi generativi: un contenuto prodotto artificialmente non è di per sé inaffidabile o inaccurato, ma deve essere riconoscibile come tale. Analogamente, i *deepfake* non sono vietati in quanto tali (salvo integrino contenuti illeciti, ad esempio materiali pornografici generati artificialmente con volti reali) ma devono essere chiaramente etichettati come contenuti generati o manipolati artificialmente.

In sintesi, il percorso intrapreso dal legislatore europeo appare, in linea di principio, condivisibile. Invece di optare per una regolamentazione contenutistica motivata dai timori suscitati dalle nuove tecnologie, l'Unione sembra rinunciare a una potenziale inclinazione censoria per lasciare spazio a un mercato delle idee quanto più possibile libero e trasparente. Che, in definitiva, è ciò che richiede l'art. 21 della nostra Costituzione.

AI E NUOVI MODELLI DI GARANZIA DEI DIRITTI FONDAMENTALI: CRITICITÀ E PROSPETTIVE APPLICATIVE

Giulia Vasino*

SOMMARIO: 1. Introduzione. Dalla disinformazione digitale ai nuovi meccanismi di tutela – 2. La Fria: potenzialità e caratteristiche costitutive di un nuovo strumento di tutela – 2.1. Le difficoltà teorico-applicative e i modelli esistenti di riferimento – 2.2. Possibili parametri operativi coerenti con un approccio giuridico e qualitativo alla tutela dei diritti fondamentali – 2.3. Alcune prospettive di concreta implementazione della Fria – 3. Conclusioni.

1. Introduzione. Dalla disinformazione digitale ai nuovi meccanismi di tutela

L'obiettivo di definire e bloccare la circolazione di contenuti disinformativi – inquadrati, appunto, come nocivi per la libera dialettica democratica e idonei potenzialmente a determinare un rischio sistemico – ha assunto una solidità normativa senza precedenti¹. Quest'ultima si estrinsecerebbe in un insieme di regole dotate di una forza performativa – un solido combinato disposto di atti – e una ampiezza – la pluralità di oggetti e destinatari – di inedita portata².

In merito al problema della disinformazione sono in particolare il Dsa e l'Aia a costruire una sinergia regolatoria all'apparenza formalmente impeccabile. Ambedue gli atti, com'è noto, si innervano su una logica di *accountability*

* Ricercatrice di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

1. Sull'evoluzione della sfera pubblica cfr. M. Manetti, *Internet e i nuovi pericoli per la libertà d'informazione*, in *Quaderni costituzionali*, n. 2/2023, p. 534.

2. Per una ricostruzione dell'articolato panorama normativo e le fonti più rilevanti in materia cfr. A. Iannuzzi, *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino, (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, spec. p. 13 e 23 ss.

graduata sulla base dell'entità del soggetto (Dsa)³ o dei rischi potenziali posti da una determinata tipologia di AI (Aia)⁴ con caratteristiche peculiari.

In forza di tale quadro, i destinatari delle prescrizioni devono assicurare di aver posto in essere i rigidi adempimenti individuati dall'atto normativo affinché possano essere considerati esenti da responsabilità per i contenuti ospitati ovvero per i danni potenziali prodotti da una determinata tecnologia. Tali obblighi si accompagnano poi a peculiari vincoli di trasparenza che in ambedue le discipline fungono da perno dell'impianto regolatorio nella sua totalità⁵.

È tuttavia nel pilastro rappresentato dal *risk-based approach* che in entrambi gli atti emergono i profili più rilevanti con riferimento al peculiare problema della disinformazione. Quest'ultimo acquisisce per la prima volta all'interno del Dsa un suo autonomo ed esplicito inquadramento, qualificandosi come prodotto collaterale connesso all'utilizzo dei servizi delle *tech companies*⁶. Gli artt. 34-35 del Dsa, più nello specifico, individuando una soluzione per contrastare gli effetti negativi connaturati al mezzo, invitano a prestare particolare attenzione alle ipotesi in cui le modalità d'uso contribuiscano a «diffondere o amplificare contenuti fuorvianti o ingannevoli, compresa la disinformazione»⁷.

L'AI Act tenta di perfezionare l'ambizioso scopo di mantenere un perfetto equilibrio fra la tutela dei diritti dei singoli e la sicurezza del prodotto⁸, vista la sua primaria base giuridica, l'art. 114 Tfu⁹. Il regolamento, nonostante l'oggetto che lo caratterizza, mantiene infatti la continuità con le fonti preesistenti e quelle in fase di elaborazione, prediligendo lo stesso *humus* valoriale e approccio metodologico proiettato alla trasparenza, alla responsabilizzazione dell'operatore, e al principio cardine del controllo umano¹⁰.

Benché le aspirazioni sottese all'Aia debbano forse essere ancora sondate e

3. Si fa riferimento, in particolare, al noto sistema a responsabilità condizionata (cfr. Capo III, Sez. V, Dsa).

4. Cfr. art. 6 e l'Allegato III dell'atto normativo.

5. Cfr., *ex multis*, l'art. 14 par. 1 del *Digital Services Act*. Emerge, inoltre, una diretta relazione fra trasparenza e dieta informazionale come evidenziano gli ulteriori obblighi di trasparenza posti per i sistemi di raccomandazione (art. 27) e i requisiti posti per le pubblicità online (art. 39).

6. Art. 34, par. 1. Tale correlazione si evince, inoltre, dal Considerando n. 2 del Dsa il quale si riferisce al contrasto a «contenuti illegali, disinformazione online e altri rischi per la società».

7. *Considerando* n. 84.

8. Per un ampliamento di questi profili cfr. M. Almada, N. Petit, *The EU AI Act: a medley of product safety and fundamental rights?*, in *SSRN Scholarly Paper*, Rochester, 2023.

9. Si veda il *Considerando* n. 1 dell'atto e l'art. 1 dell'Aia. In merito al punto specifico si rimanda a F. Fabrizzi, L. Durst, *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era dell'algoritmo*, Editoriale Scientifica, 2024, p. 7 ss.

10. Si vedano, in particolare, gli artt. 9 e 13 dell'atto normativo nonché gli artt. 43, 50 e 55.

analizzate in tutte le loro potenzialità¹¹, con riferimento alla disinformazione, elementi regolatori rilevanti possono trarsi dall'innalzamento degli obblighi di trasparenza previsti per i *deepfake*¹² e dalla specifica disciplina dedicata ai sistemi di intelligenza artificiale con finalità generali, con particolare riferimento al *genus* rappresentato dai sistemi di Gen-AI e quelle tecnologie che pongono rischi sistemici¹³.

In merito all'elevazione dei requisiti di trasparenza posti con riferimento ai sistemi di AI generativa che producano un contenuto audio, immagine, video o testuale manipolato artificialmente, la prescrizione emergente dall'atto consiste nell'obbligo di marcatura del contenuto, con previsione analoga per quegli specifici prodotti elaborati «allo scopo di informare il pubblico su questioni di interesse pubblico»¹⁴. Tale previsione colpisce, oltre che per la sua ambivalenza e vaghezza definitoria, poiché trasmetterebbe una posizione apparentemente neutra nei confronti del prodotto manipolato per il quale non è prevista alcuna «etichetta di falsità» ma soltanto una certificazione di trasparenza che renda il fruitore della notizia consapevole della natura del contenuto. Si fugherebbe così l'ipotesi che il legislatore continentale prediliga di per sé orientamenti «paternalistici» in relazione a contenuti non autentici¹⁵.

Tuttavia, al di là delle prescrizioni puntualmente dedicate alla disinformazione, è l'approccio regolatorio, nella sua generalità e innovatività, a invitare a una riflessione più radicale sul rapporto fra *fenomeno nocivo-rimedio-effettività della garanzia predisposta*. Di fronte a tale panorama normativo in fase di progressiva espansione e perfezionamento, si ha l'impressione, infatti, che il raffinemento della disciplina in chiave *human centric* – dichiaratamente modellata sul rafforzamento della posizione dell'individuo di fronte alla piattaforma o al fornitore/operatore di AI che si concreta anche in un articolato e convincente «*digital process of law*»¹⁶ e in un'emersione di istanze assimilabili a un diritto

11. Sul punto si veda A. Simoncini, *La co-regolazione delle piattaforme digitali*, in *Rivista trimestrale di diritto pubblico*, n. 4, 2022, p. 1031 ss.; e, in senso critico, O. Pollicino, *Regolazione e innovazione tecnologica nell'ordinamento della rete*, *Relazione al XXXIX Convegno annuale dell'Associazione Italiana dei Costituzionalisti*, 15-16 ottobre 2024, pp. 52-53; F. Donati, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, in *Rivista AIC*, n. 1, 2025, p. 1 ss.

12. Cfr. in part. art. 50, par. 4. Per una disamina approfondita cfr. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, n. 1, 2023, p. 170 ss.

13. Si v. art. 51 e i Considerando nn. 97 e 99.

14. Art. 50, par. 4.

15. Si v. G.E. Vigevani, *Potere politico e mezzi di comunicazione*, *Relazione al XXXIX Convegno annuale dell'Associazione Italiana dei Costituzionalisti*, 15-16 ottobre 2024, p. 33.

16. C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, n. 3, 2023, p. 559.

alla tutela giurisdizionale di fronte al privato – moltiplichi, piuttosto che attenuare, gli interrogativi teorici posti alla base.

Quest'ultimi, nella complessità del dibattito scientifico a essi dedicato, si sono da un lato ampiamente orientati verso la dimensione soggettiva. Parte della qualificata *querelle* giuridica sul tema ha infatti riservato uno spazio rilevante al “*chi*”, scandagliando la natura problematica degli attori coinvolti¹⁷. Le voci critiche si sono concentrate maggiormente sui profili inerenti ai titolari, focalizzandosi, da un lato, sul famoso dilemma della cessione, nelle mani del soggetto privato, di un obbligo di implementazione della disciplina euro-unitaria che vada ben oltre la mera applicazione di una previsione¹⁸; e, dall'altro, ponendo l'accento sulla pericolosa ricaduta che tale rinnovato rapporto privato-soggetto pubblico crei nei confronti di quest'ultimo¹⁹.

Più chiaramente, se, da un canto, non è stato taciuto il rischio di uno scivolamento orizzontale che metta il privato nella pericolosa posizione di maneggiare i diritti fondamentali²⁰, si è insistito, dall'altro, sull'esigenza di non rimanere meno vigili sul soggetto statale, ricordando le fondamenta delle garanzie di libertà e il paradigma classico autorità-individuo intorno al quale sono cuciti i diritti della persona.

In tal modo, in una prima ottica, è apparso ragionevolmente preoccupante lasciare alle piattaforme valutazioni sostanziali sui contenuti da eliminare, attuando spesso più che una delega di *enforcement* ma una vera e propria delega sostanziale²¹; contestualmente è stato altresì rilevato in modo altrettanto pertinente l'espansione della discrezionalità dell'operatore privato in presenza di tecnologie di AI.

Nel caso dell'AI Act, le perplessità del fronte soggettivo si fondono in modo

17. Fra i molti scritti dedicati dalla dottrina si rimanda a K. Klonick, *The new governors: the people, rules, and processes governing online speech*, in *Harvard Law Review*, vol. 131, n. 6, 2018, p. 1662 ss.; T. Gillespie, *Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media*, Yale University Press, 2018, p. 5; J. Grimmelmann, *The Virtues of Moderation*, in *Yale Journal of Law and Technology*, n. 17, 2015, p. 61 ss.

18. M. Bassini *Fundamental rights and private enforcement in the digital age*, in *European Law Journal*, vol. 25, n. 2, 2019, p. 187.

19. Cfr. J. Balkin, *Old-school/new-school speech regulation*, in *Harvard Law Review*, 2014S. F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, in *University of Pennsylvania Law Review*, 2006, p. 13 ss.

20. Tale premessa critica oramai classica di tali studi è ricostruita da M. Betzu, *I baroni del digitale*, Editoriale Scientifica, 2022.

21. Sul punto cfr. M. Monti, *Privatizzazione della censura e Internet platforms: la libertà d'espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, n. 1, 2019, p. 35 ss.; O. Grandinetti, *Le piattaforme digitali come “poteri privati” e la censura online*, in *Rivista italiana di informatica e diritto*, n. 1, 2022, p. 181.

ancora più rappresentativo con la dimensione oggettiva della questione. Difatti, la natura inedita dei fenomeni e delle tecnologie con le quali il legislatore della sfera digitale è costretto a confrontarsi spinge sempre più a domandarsi se la normativa euro-unitaria riesca oggi a superare quell'implicito "stress test" anche rispetto alle modalità e alle condizioni individuate per effettuare concretamente, e in modo effettivo, la tutela dei diritti coinvolti. In particolare, il problema della disinformazione – quale fenomeno sistemico "tipico" dell'era digitale poiché esacerbato dai sistemi di AI – induce a riflettere, in una dimensione più ampia, sugli anticorpi garantistici disposti dall'atto normativo nella loro intrinseca struttura e finalità.

Fra gli stessi si distinguono quelle valutazioni di impatto le cui potenzialità appaiono ancora da sondare ed esplorare. In particolare, risalta, proprio per la sua aspirazione universalistica, il primo comma dell'art. 27 dell'Aia. La previsione stabilisce che «i *deployer* che sono organismi di diritto pubblico o sono enti privati che forniscono servizi pubblici e i *deployer* di sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere *b* e *c*, effettuano una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema può produrre».

Partendo da tale prospettiva più *macro*, si cercherà di porre in luce brevemente la relazione problematica fra la "spasmodica" centralità ricoperta dall'obiettivo di tutelare i diritti fondamentali di fronte all'avanzamento delle nuove tecnologie, testimoniata dalla moltiplicazione delle norme e degli strumenti messi sul campo, e la intrinseca problematicità di tali modelli, costretti a "far i conti" con l'elemento tecnologico posto alla base.

A tale fine si utilizzerà la Fria come strumento esemplificativo di tale difficile obiettivo di rimodellamento delle tutele. I tentativi del legislatore unionale di coniugare l'avanzamento tecnologico con soluzioni normative che riecheggiano istituti garantistici classici "scontano", infatti, le difficoltà scaturenti dall'ibridazione fra meccanismi di natura privatistica e la vocazione universalistica e di tutela propria del costituzionalismo liberal-democratico. Tali aporie e problematicità connaturate allo strumento potrebbero essere parzialmente attenuate solo attraverso un'oculata traduzione applicativa del test valutativo: una sfida che oggi appare ancora in fase di costruzione.

2. La Fria: potenzialità e caratteristiche costitutive di un nuovo strumento di tutela

L'art. 27, imponendo un'analisi preventiva per i sistemi di AI ad alto rischio a garanzia dei diritti fondamentali – che si affianca, senza sostituirsi, a quella prevista dall'articolo 35 del Gdpr – appare esemplificativo delle aspirazioni

dell'intero atto normativo. Le stesse sembrano infatti tradotte anche a livello più strettamente operativo: imponendo al *deployer* una valutazione *ex ante*, il *Fundamental rights impact assessment*, ci si focalizza sui danni potenziali che una determinata tecnologia potrebbe porre per i diritti del singolo all'interno di un regolamento che fa del rischio il proprio criterio ordinatore e la propria ossatura-guida.

La Fria, inoltre, si affianca a ulteriori test valutativi, contribuendo alla edificazione di un sistema di filtri particolarmente serrato anche di fronte a fenomeni suscettibili di esercitare un impatto di carattere generale²². Tale griglia valutativa comprende:

- il generale RMS imposto dall'articolo 9, un *Risk Management System* che interessa tutte le tecnologie ad alto rischio e che si concentra sulla previa individuazione ed eventuale mitigazione relativa a beni quali la sicurezza e la salute²³;
- il CA posto dall'art. 43, un *Conformity Assessment*, che esige una valutazione di *compliance* per le tecnologie ad alto rischio indicate nell'All. III «in order to ensure a high level of trustworthiness of high-risk AI systems».

Mentre il CA avrebbe una declinazione maggiormente “tecnica” e ricade sul provider, il Fria non a caso ricade sul *deployer* e si concentra sull'impatto operativo di un'AI che si “interfaccia con il mondo reale”, con una particolare attenzione ai gruppi marginalizzati²⁴.

Infine, centrale, soprattutto in relazione a fenomeni quali la disinformazione, risulta lo SRA, imposto dall'art. 55, un *Systemic Risk Assessment* per i sistemi di AI con finalità generali che pongono rischi sistemici quali le Gen-AI²⁵. Mentre il Fria si concentra sul *deployment-specific risk*, lo Sra si concentra sui rischi sistemici

22. Sull'intreccio fra i vari test di valutazione anche esterni all'atto, quali il Dpia cfr. anche O. Pollicino, F. Paolucci, *Regulating AI Autonomy: A Constitutional Framework for the Digital Era*, SSRN Paper, 31 dicembre 2024, disponibile su: papers.ssrn.com/sol3/papers.cfm?abstract_id=5098433.

23. Si tratta, infatti di operatori pubblici o operatori privati che offrono un servizio pubblico (ad es. salute, educazione), operatori che per conto di autorità pubbliche sono chiamati a valutare l'ammissibilità delle persone fisiche nell'accesso a servizi essenziali (ad. es. sanità e credito).

24. In senso critico sulla distinzione di responsabilità fra *deployer* e *provider* cfr. F. Paolucci, *Shortcomings of the AI Act*, in *Verfassungsblog*, 14 marzo 2024.

25. Sulla natura dello Sra il quale è, in prima battuta, imposto dal Dsa e solo indirettamente incide anche sulla tecnologia utilizzata si rimanda all'analisi comparativa fra gli strumenti indicati di P. Chiara, F. Galli, *Normative Considerations on Impact Assessments in EU Digital Policy*, in *MediaLaws*, n. 1, 2024, p. 86; sull'ampia discrezionalità concessa al decisore politico circa

ci connessi a determinate tecnologie che hanno un impatto sociale significativo e, per questo, è sovente oggetto di critiche a livello teorico: si argomenta, infatti, come il *provider* vada a concentrarsi sulla tecnologia in sé in senso atecnico, ragionando su rischi di natura astratta che neppure potrebbero mai manifestarsi²⁶.

In realtà, da un punto di vista della strutturazione dello strumento e del suo livello di dettaglio regolatorio, il legislatore unionale pare esser senza dubbio pervenuto a un livello di sofisticatezza notevole soprattutto nel modo in cui si tende a coniugare l'equilibrio fra dato tecnico e tutela dei diritti. Il *Fundamental rights impact assessment* (Fria) è finalizzato infatti proprio a effettuare una valutazione preventiva del rischio in relazione ai diritti fondamentali attraverso, in estrema sintesi, tre momenti fondamentali:

1. individuare il rischio;
2. calcolarne l'entità;
3. prevenire e individuare misure di mitigazione (laddove possibili).

Gli elementi essenziali dell'*iter* valutativo appaiono tracciati con chiarezza ed efficacia a partire dal fronte definitorio, come si vede dalla qualificazione di "sistema ad alto rischio" fino alla enunciazione dei passaggi che costituirebbero il cuore operativo della Fria: i *deployer* effettuano una valutazione che comprende, in particolare, gli elementi seguenti:

1. una descrizione dei processi del *deployer* in cui il sistema di IA ad alto rischio sarà utilizzato in linea con la sua finalità prevista;
2. una descrizione del periodo di tempo entro il quale ciascun sistema di IA ad alto rischio è destinato a essere utilizzato e con che frequenza;
3. le categorie di persone fisiche e gruppi verosimilmente interessati dal suo uso nel contesto specifico;
4. i rischi specifici di danno che possono incidere sulle categorie di persone fisiche o sui gruppi di persone;
5. una descrizione dell'attuazione delle misure di sorveglianza umana, secondo le istruzioni per l'uso;
6. le misure da adottare qualora tali rischi si concretizzino, comprese le disposizioni relative alla governance interna e ai meccanismi di reclamo²⁷.

la definizione di rischio sistemico si esprime in senso critico F. Donati, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, cit., p. 16.

26. Sul punto cfr. A. Mantelero, *The Fundamental Rights Impact Assessment (Fria) in the AI Act: roots, legal obligations and key elements for a model template*, in *Computer Science Law Review*, vol. 54, 2024.

27. Art. 27, par. 1, Aia.

Le dettagliate indicazioni contenute all'art. 27 par. 2 vengono inoltre arricchite, in prospettiva ermeneutica, dalle ulteriori specifiche contenute al *Considerando* n. 96, il quale estende l'aspetto definitorio soprattutto nell'inquadrate lo strumento sotto il profilo finalistico, soggettivo e temporale. Si ribadisce, infatti, come l'obiettivo della valutazione d'impatto sui diritti fondamentali è consentire al *deployer* di determinare i rischi specifici per i diritti delle persone o dei gruppi di persone che potrebbero essere interessati e di individuare le misure da adottare al concretizzarsi di tali rischi.

Viene chiarito, inoltre, come l'*impact assessment*, per essere realmente una cartina di tornasole efficace, dovrebbe essere svolto prima del primo impiego del sistema di IA ad alto rischio e dovrebbe essere aggiornata quando il *deployer* ritenga che uno qualsiasi dei fattori costitutivi sia cambiato. Nell'effettuare tale esame, l'operatore dovrebbe tenere conto delle informazioni pertinenti per un'adeguata valutazione dell'impatto, comprese, tra l'altro, le informazioni trasmesse dal fornitore del sistema di IA ad alto rischio nelle istruzioni per l'uso. Nel valorizzare la sinergia con il fornitore, il sopraccitato *Considerando* pone l'accento sull'importanza del coinvolgimento dei soggetti istituzionali chiave – dall'Autorità di vigilanza all'ufficio europeo per l'AI – a cui deve essere notificato il risultato, rafforzando in tal modo le prescrizioni previste dai parr. 3 e 5 dell'art. 27 Aia²⁸.

La valutazione di impatto deve infine rendere chiari i processi in cui verrà adoperata una determinata tecnologia, con quale finalità e con quale frequenza. Sulla base delle informazioni fornite dal *provider*, devono essere individuati in modo accurato i soggetti che potrebbero essere negativamente condizionati dall'uso di quella tecnologia e dimostrare consapevolezza circa le misure che intendono adottare in caso in cui quel rischio di materializzi effettivamente. Inoltre, il *deployer* deve assicurare, ancora una volta, un sistema interno di reclami e gestione degli stessi, con la garanzia della supervisione umana, non solo per consentire al singolo di esercitare un ruolo nelle procedure che attengono alla sua posizione soggettiva ma anche perché tali garanzie possono essere strumentali ad assicurare una mitigazione del rischio in concreto.

Questo processo dovrebbe, sulla carta, favorire un concreto *empowerment* degli individui sia per il livello di dettaglio sia per la sua natura preventiva nonché per il costante controllo dell'autorità pubblica, designata, ai sensi del pa-

28. Risulta poi centrale, all'interno del *Considerando* n. 96, il riferimento alla società civile: si evidenzia come, qualora i sistemi di IA siano utilizzati nel settore pubblico, risulterebbe di estrema efficacia il coinvolgimento di portatori di interessi pertinenti, compresi i rappresentanti di gruppi di persone che potrebbero essere interessati dal sistema di IA, gli esperti indipendenti e le organizzazioni della società civile.

ragrafo 3 dell'art. 27, come soggetto destinatario della scheda valutativa: una scelta che testimonia, ancora una volta, un accoglimento di quel percorso di valorizzazione del principio di trasparenza e responsabilità alla luce del faro rappresentato dai diritti.

Nonostante ciò, la necessaria esigenza di implementazione di questo strumento di tutela ha inevitabilmente posto l'attenzione su come realmente rendere concreto ed effettivo un modello che muove delicatamente a cavallo fra la carica assiologica dei diritti e il fronte tecnico e sfaccettato dei sistemi di AI. La sfida attiene sia alla dimensione teorica e di principio, da un lato, sia alla dimensione più strettamente operativa, dall'altro, mancando attualmente un'univoca "griglia" e strategia applicativa che possa garantire uno standard uniforme a livello europeo.

2.1. Le difficoltà teorico-applicative e i modelli esistenti di riferimento

Con riferimento alla prima prospettiva si osserva come il Fria – ancora una volta e maggiormente come il Dpia, data la massimizzazione della sua vocazione garantistica – sugella il ponte teorico-operativo fra la dimensione del rischio e quella dei diritti, fra un modello che presuppone la quantificabilità e una sfera, quale quella dei diritti fondamentali, caratterizzata da una netta dimensione qualitativa, la quale ha sempre faticato a coniugarsi con ambiti governati dal *quantum*²⁹.

A maggior ragione questa ambizione si rivela ancora più problematica se affidata a una valutazione *ex ante* che esige, quindi, che si prospetti e si soppesi l'entità delle possibili conseguenze negative prima che il danno si sia effettivamente prodotto. Con riferimento a quest'ultimo, la definizione contenuta nel *Considerando* n. 75 del Gdpr di «rischio per i diritti fondamentali» rimane comunque determinante ancorché vaga, qualificandolo lo stesso come «*a composite result of the probability and severity of certain data processing activities, which could lead to physical, material or non-material damage*». In tal modo, sembra confermarsi l'idea che il danno (*harm*) costituisca il vero ponte, a livello concettuale prima che giuridico, fra un parametro probabilistico e quantitativo, da un lato, e la dimensione ampia e socialmente situata dei diritti, dall'altro.

29. La cui delicata dissertazione teorica si lega chiaramente all'ampio dibattito classico sulla proporzionalità e sulla bilanciabilità dei diritti fondamentali; sul punto cfr. R. Alexy, *Constitutional Rights, Balancing, and Rationality*, in *Ratio Juris*, vol. 16, n. 2, 2003, pp. 133 ss.; A. Barak, *Proportionality. Constitutional Rights and their Limitations*, Cambridge University Press, 2012, pp. 460; J. Silva Sampaio, *Proportionality in Its Narrow Sense and Measuring the Intensity of Restrictions on Fundamental Rights*, in D. Duarte, J. Silva Sampaio (a cura di), *Proportionality in Law: An Analytical Perspective*, Springer Verlag, 2018, p. 71 ss.

Da un canto, non vanno taciuti gli elementi positivi di una valutazione *risk-based*, la quale manifesta un'indubbia praticità di metodo nel momento in cui consente all'interprete, in astratto, di fare affidamento anche su una lista di rischi possibili già quantificati a livello legislativo e giurisprudenziale: in questo modo si procede ad anticipare l'evento fissando una linea di demarcazione di non tollerabilità oltre la quale il danno prodotto sarebbe ritenuto non più accettabile³⁰.

Dall'altra parte, tuttavia, anche un'osservazione preliminare sembra far sorgere più opacità teoriche e perplessità. Si è insistito, infatti, non solo sulla delicatezza di quantificare *a priori* un danno ma anche sulla aggravata difficoltà che la ponderazione numerica porrebbe ulteriormente soprattutto quando ci si sposta dalla sfera dei diritti individuali "classici" verso la sfera dei principi, come nel caso della dignità individuale. La menzionata difficoltà di quantificazione verrebbe inoltre aggravata dalla percezione e graduazione che i diritti assumono a seconda dei contesti valoriali e normativi di riferimento, i quali sono soggetti a una sensibile variazione nel panorama unionale.

Alla luce di queste premesse critiche, dunque, ancor prima di approfondire l'operatività in concreto di un inedito sistema di garanzia valutandone le prospettive applicative, i primi commentatori hanno paventato la limitata idoneità di un meccanismo di tal natura di poter ridurre la complessità della stessa categoria di diritto fondamentale, appiattendolo la stessa a una arida bidimensionalità³¹. Da quest'ultima discenderebbe, di conseguenza, soltanto la valutabilità di danni economicamente calcolabili e non il più multisfaccettato e poli-dimensionale impatto che un sistema di AI potrebbe avere sulla sfera della persona³².

Alla sfiducia di un modello *risk-based* – fondata proprio sulla asserita inadeguatezza radicale fra dimensione dei diritti e valutazione del rischio affidata, per di più, a un operatore economico – si è contrapposta pertanto la più rassicurante prospettiva mitigatoria di un modello *rights-based*³³. Quest'ultimo, a sua volta, tenta di eliminare le sfumature quantitative e adotta un approccio normativo che misura la violazione sulla base di un parametro violato: o vi è una violazione o non vi è. Tale linea direttiva, che pone la norma al centro dei

30. Si v. G. De Gregorio, P. Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, vol. 59, n. 2, 2022, pp. 473-500.

31. Una polemica teorica già ampiamente consolidata in dottrina già prima dell'entrata in vigore dell'atto, cfr. S. Engle Merry, *The Seductions of Quantification, Measuring Human Rights, Gender Violence, and Sex Trafficking*, The University of Chicago Press, 2016.

32. A. Mantelero, *Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI*, Springer, 2022, pp. 54-55.

33. Sul punto A. Mantelero, *The Fundamental Rights Impact Assessment (Fria) in the AI Act: roots, legal obligations and key elements for a model template*, in *Computer Science Law Review*, vol. 54, 2024.

diritti fondamentali, ha il pregio di evitare prospettive riduzioniste ma, dall'altra, avrebbe a sua volta una intrinseca forma di rigidità che si scontra con la formulazione testuale del Fria, la quale menziona esplicitamente la probabilità e la severità del rischio³⁴.

Con riferimento alla seconda dimensione, la principale difficoltà dipende dall'assenza di un *framework* applicativo seppur minimale. Da un lato, l'Aia, in merito alla Fria, fa esplicito riferimento ai principi della Carta di Nizza (Cfreu), la cui applicazione per via giurisprudenziale esige il rispetto del principio di proporzionalità nell'interferenza/limitazione di ciascun diritto fondamentale (art. 52).

Purtuttavia, a parte il riferimento a una *human centric AI* e alcuni specifici richiami al principio di trasparenza e sorveglianza umana (artt. 13, par. 3, e art. 14, par. 2) il test offerto dall'Aia rimane piuttosto generico, non offrendo un adeguato *benchmark* di riferimento, contribuendo così a determinare incoerenza e episodicità³⁵.

Infine, anche il problema dell'inquadramento tecnico risulta determinante. La reticenza normativa sotto questo profilo è esacerbata dalla specificità delle tecnologie contemplate dall'Aia che spaziano dall'apprendimento automatico alla robotica fino al ragionamento simbolico. A questo si aggiunge l'interdipendenza fra modelli e il riaddestramento continuo che rischia di trasformare ciclicamente gli eventuali standard elaborati con fatica in riferimenti obsoleti. Questi fattori contribuiscono a rendere anche le esigenze di conformità un "obiettivo mobile"³⁶.

A ogni modo, il Fria rappresenta l'evoluzione e uno strumento complementare rispetto ad altri sistemi/test di valutazione di impatto sui diritti umani a cui inevitabilmente i primi commentatori hanno guardato nella ricerca di standard adeguati, dallo *Human Rights Democracy and Rule of Law Impact Assessment* (Huderia) allo *Human Rights Impact Assessments* (Hria) elaborato in seno alle Nazioni Unite senza ovviamente tralasciare il ruolo chiave del già citato *Data Protection Impact Assessment* (Dpia) introdotto dal Gdpr³⁷.

34. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, in *Computer Law & Security Review*, n. 56, 2025.

35. Cfr. L. Gatt et al., *Fria implementation model according to the AI Act*, in *EJPLT*, n. 2, 2024.

36. Sul punto cfr. P. Ceravolo et al., *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, disponibile su: arxiv.org/abs/2503.18994.

37. Per una introduzione generale a tali strumenti si rimanda a N. Götzmann, *Introduction to the Handbook on Human Rights Impact Assessment: Principles, methods and approaches*, in N. Götzmann (ed.), *Handbook on Human Rights Impact Assessment*, Edward Elgar Publishing, 2019, p. 2 ss.

Con riferimento al supporto offerto dal diritto internazionale non va inoltre tralasciato il ruolo chiave svolto, almeno in una prima fase transitoria, dagli organismi di normazione. L'Organizzazione Internazionale per la Normazione (Iso) e la Commissione Elettrotecnica Internazionale (Iec), attraverso il comitato congiunto Iso/Iec Jtc 1/SC 42, hanno sviluppato un ampio insieme di standard dedicati all'intelligenza artificiale³⁸. Gli stessi tentano di individuare un primo "pacchetto" di linee guida di riferimento per la valutazione dell'impatto dell'IA sui diritti fondamentali che tenga conto della gestione del rischio, della qualità dei dati, della robustezza dei modelli e della governance organizzativa³⁹.

Tra gli standard più rilevanti si colloca Iso/Iec 23894 che fornisce linee guida per la gestione del rischio nei sistemi di IA. Questo standard definisce un processo strutturato per l'identificazione, l'analisi, la valutazione e il trattamento dei rischi associati all'IA, ponendo l'accento sull'interazione tra fattori tecnici, organizzativi e contestuali. L'approccio proposto è coerente con la logica *risk-based* adottata dall'AI Act e può essere efficacemente integrato nei sistemi di gestione del rischio richiesti dall'art. 9 del Regolamento⁴⁰.

Accanto agli standard Iso/Iec, anche l'*Institute of Electrical and Electronics Engineers* (Ieee) ha sviluppato una serie di standard e linee guida che pongono una maggiore enfasi sulle dimensioni etiche e sociali dell'intelligenza artificiale. In particolare, la famiglia di standard Ieee 7000 affronta temi quali la trasparenza, la responsabilizzazione, la mitigazione dei *bias* e l'allineamento dei sistemi di IA ai diritti umani⁴¹.

38. La cui rilevanza è richiamata e ricostruita in P. Ceravolo *et al.*, *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, 23 marzo 2025, p. 7 ss., disponibile su: arxiv.org/abs/2503.18994.

39. Per informazioni più dettagliate cfr. www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0.

40. Altri standard rilevanti includono l'Iso/Iec 23053, il quale fornisce un quadro di riferimento per l'IA basata su apprendimento automatico, e Iso/Iec 25059, che affronta la qualità dei sistemi di IA in termini di accuratezza, affidabilità e robustezza. Nel complesso, gli standard Iso/Iec offrono un insieme coerente di strumenti per affrontare le dimensioni tecniche della valutazione dell'IA. Tuttavia, essi tendono a concentrarsi prevalentemente sugli aspetti ingegneristici e organizzativi, lasciando in secondo piano una valutazione esplicita e sistematica dell'impatto sui diritti umani. Per un approfondimento si rimanda ugualmente al sito disponibile su: www.iso.org/committee/6794475.html.

41. Lo standard Ieee 7000 propone un modello per integrare considerazioni etiche nel processo di progettazione dei sistemi, promuovendo un approccio "*value-based*" allo sviluppo tecnologico. Altri standard della stessa famiglia, come Ieee 7001 (trasparenza), Ieee 7002 (data privacy) e Ieee 7003 (mitigazione dei bias), forniscono indicazioni operative per affrontare rischi specifici legati all'impatto dell'IA sui diritti fondamentali. Sul punto si rimanda al sito disponibile su: standards.ieee.org/industry-connections/activities/ieee-global-initiative/.

Tali linee guida risultano particolarmente rilevanti nel contesto degli *impact assessment* e offrono strumenti concreti per tradurre principi etici astratti in requisiti progettuali verificabili. Tuttavia, analogamente agli standard Iso/Iec, essi non sono concepiti specificamente per soddisfare gli obblighi giuridici dell'AI Act e non forniscono una metodologia integrata per la conduzione di una Fria completa⁴².

In primo luogo, tali parametri sono spesso frammentati, ciascuno focalizzato su un aspetto specifico del sistema di IA, senza offrire una visione complessiva dell'impatto sui diritti fondamentali. In secondo luogo, gli standard tecnici tendono a privilegiare una prospettiva *ex ante* e progettuale, mentre l'AI Act richiede una valutazione continua lungo l'intero ciclo di vita del sistema, inclusa la fase di utilizzo concreto e di monitoraggio *post-deployment*. Inoltre, la maggior parte degli standard omette di affrontare in modo esplicito il bilanciamento tra diritti fondamentali potenzialmente confliggenti, creando così un supporto applicativo discontinuo e parziale⁴³.

2.2. Possibili parametri operativi coerenti con un approccio giuridico e qualitativo alla tutela dei diritti fondamentali

Alla luce di tali criticità, è apparso subito necessario un approccio metodologico integrato che sia in grado di combinare i contributi degli standard tecnici con una precomprensione giuridica adeguata.

Da un punto di vista teorico risultano di rilievo quelle riflessioni che si sono orientate verso la valorizzazione, in fase di implementazione del modello, del concetto di "interferenza" con i diritti fondamentali piuttosto che di "violazione".

L'interferenza, in primo luogo, è un concetto che sfugge alla logica dicotomica della violazione soggiacendo a una logica di gradualità. Inoltre, l'apporto definitorio offerto dalla giurisprudenza della Corte di Giustizia UE e dalla Corte Edu risulta determinante trovando all'interno della stessa un ricorrente riferimento al concetto di «*seriousness of interferences*»⁴⁴. Entrambe le Corti

42. P. Ceravolo *et al.*, *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, cit., p. 9.

43. Ivi, p. 6.

44. Così la Corte Edu nel noto caso *Roman Zakharov v. Russia*, 2015, par. 232 in cui si ribadisce la necessità di individuare la «the seriousness of the interference with an applicant's right to respect for his or her private life» chiarendo come «the assessment depends on all the circumstances of the case, such as the nature, scope and duration of the possible measures, the grounds required for ordering them, the authorities competent to authorise, carry out and supervise them, and the kind of remedy provided by the national law. The Court has to determine

hanno infatti affermato che le interferenze costituiscono una scala graduabile, misurabile in termini di severità. Più chiaramente, è stato evidenziato come, ai fini di valutare la proporzionalità di un'interferenza, bisogna altresì misurarne la severità.

Se dunque ci si apre, a livello teorico, al necessario apporto di un modello *right-based* – il quale supererebbe i limiti dei meccanismi del rischio esclusivamente quantitativi – e, al contempo, la rigidità del concetto di violazione – tipico del modello *rights-based* – in favore di un concetto più mediato quale quello di interferenza, è opportuno individuare alcuni parametri.

Allo stato attuale, appaiono efficaci, in termini di approccio organico ed efficacia della valutazione, l'individuazione congiunta di alcuni parametri che, se applicati congiuntamente, risulterebbero oggettivi senza “inaridire” le implicazioni assiologiche sottese alla valutazione di un'interferenza con un diritto fondamentale. Fra le stesse si segnala un modello di azione integrato che tenga conto di:

- a. un parametro oggettivo, ossia una valutazione della gravità dell'interferenza valutata normativamente;
- b. un parametro soggettivo, finalizzato a soppesare il “valore sociale” di quella interferenza ovvero come la stessa venga percepita a livello individuale e collettivo in un determinato ambito;
- c. un parametro basato sull'osservazione delle conseguenze eventualmente prodotte, volto a identificare l'effetto reale causabile da quella interferenza, contemplato in termini economici e di benessere⁴⁵.

Con riferimento al parametro a) esso indubbiamente risulta, agli occhi del giurista, il criterio più affidabile in quanto si fonda, in primo luogo, sulla valutazione di una violazione congiunta di fonti primarie (previsioni costituzionali, della Carta dei diritti fondamentali, Convenzione Edu) e secondarie attuative di un determinato diritto fondamentale. Nell'ambito di tale criterio possono risultare determinanti anche ulteriori sub-criteri “interni” quali:

1. le indicazioni, contenute all'interno dello stesso atto normativo/previsioni coinvolte, utili a classificare e inquadrare l'intensità di una violazione, le

whether the procedures for supervising the ordering and implementation of therestrictive measures are such as to keep the “interference” to what is “necessary in a democratic society» (§232). Con riferimento alla Corte di Giustizia, si richiama, a titolo meramente esemplificativo, i passaggi dedicati delle note sentenze *Digital Rights Ireland* (§61) e *Google Spain* (§81).

45. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, in *Computer Law & Security Review*, n. 56, 2025, p. 5 ss.

quali offrono indicazioni circa la severità della interferenza, come ad esempio indicato all'art. 83 Gdpr⁴⁶;

2. l'interpretazione offerta dalla giurisprudenza delle Corti. Difatti, benché manchi una "gerarchia" fra diritti, non solo la Corte Edu si riferisce esplicitamente ad alcune disposizioni quali «*most fundamental provisions*», com'è noto, ma definisce la presenza di diritti non derogabili, nemmeno nelle situazioni emergenziali⁴⁷.

Rientrano all'interno di una valutazione oggettiva volta a stabilire la "severità", la valutazione della reversibilità del danno e la durata della violazione.

Con riferimento alla reversibilità, si tratta di un criterio più solido perché si intende soppesare l'entità della violazione al di là della percezione soggettiva del danno ma in termini di "obiettiva rimediabilità". In tale prospettiva, ad esempio, si considera irreversibile una mancata assunzione di un individuo per il suo orientamento sessuale in grado di violare il diritto antidiscriminatorio europeo (Directive 2000/78/EC) e l'art. 21 della Carta dei diritti fondamentali, mentre si potrebbe considerare di "lieve" severità, in termini di reversibilità, una mancata applicazione del principio di trasparenza nella fornitura di un servizio di intermediazione digitale (art. 15 Dsa) in grado di incidere negativamente sulla libertà di espressione⁴⁸. La valutazione della durata della violazione

46. L'art. 83, par. 2, Gdpr afferma infatti che in ogni singolo caso si debba tener conto di alcuni elementi, fra i quali, ad esempio, «a) la natura, la gravità e la durata della violazione tenendo in considerazione la natura, l'oggetto o a finalità del trattamento in questione nonché il numero di interessati lesi dal danno e il livello del danno da essi subito; b) il carattere doloso o colposo della violazione; c) le misure adottate dal titolare del trattamento o dal responsabile del trattamento per attenuare il danno subito dagli interessati; d) il grado di responsabilità del titolare del trattamento o del responsabile del trattamento tenendo conto delle misure tecniche e organizzative da essi messe in atto».

47. Gli artt. 2 e 3 della Convenzione vengono qualificati come disposizioni fondamentali in quanto idonee a rappresentare i valori fondativi del Consiglio d'Europa e della società democratica. Alcuni diritti, inoltre (artt. 15, 3, 4) sono esplicitamente inderogabili anche in circostanze emergenziali. Si tratta, tuttavia, di un sub criterio da implementare, eventualmente, con un certo grado di attenzione, considerando che alcuni diritti, pur non essendo espressamente qualificati come "fondamentali", potrebbero essere declassati su un gradino inferiore (ad. es art. 14) e, soprattutto, che altri diritti contengono, analogamente al menzionato art. 83 del Gdpr, delle linee interne volte a stabilire le condizioni e i limiti apponibili al diritto stesso, come dimostra la stessa libertà di espressione. Il compito di individuare quali siano le circostanze che giustificano la violazione, seppur tassativamente indicate dall'articolo, comporterebbero che in fase di valutazione preventiva si prefiguri un bilanciamento fra diritti fondamentali (la classica questione fra libertà di espressione e tutela della dignità individuale). Oltretutto, il collegamento fra disposizioni e legislazione secondaria attuativa non è sempre così diretto, come per alcune categorie di diritti fondamentali.

48. Così G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 10.

emerge, invece, come requisito che consente di attribuire al tempo, inteso sia in termini di svolgimento prolungato sia di frequenza dell'accadimento, un peso in termini di entità della interferenza⁴⁹.

Infine, risulta determinante nell'integrazione del primo criterio la valutazione della sussistenza di una violazione congiunta di diritti fondamentali, alla luce della logica generale secondo la quale una potenziale violazione contestuale di più previsioni concernenti i diritti fondamentali dovrebbe accrescere l'entità dell'interferenza.

La mera sussistenza di un criterio oggettivo, tuttavia, è apparsa sotto alcuni profili parziale. La violazione congiunta di più diritti fondamentali può rivelarsi, ad esempio, un sub-criterio di valutazione che non resiste alle accuse di eccessivo formalismo, considerando che la medesima obbligazione può essere contenuta in più articoli: questi ultimi potrebbero stabilire la stessa prescrizione sotto il profilo sostanziale senza che questo legittimi, come affermato dalla stessa Cgue, a effettuare un "*double-counting*" nel momento dell'esame dell'entità di un'interferenza⁵⁰.

Dall'altra parte, l'avvicinamento verso una "classificazione" di valore dei diritti fondamentali, letto attraverso l'ottica offerta dalla variegata giurisprudenza chiamata a effettuare un centrale filtro ermeneutico anche in fase di costruzione di una valutazione preventiva, risulta un'operazione delicata. Non si tratta soltanto di porre l'attenzione sul noto e delicatissimo problema teorico della costruzione di una gerarchia fra diritti, evidenziato anche dalla nostra giurisprudenza interna⁵¹ ed esasperato in una dimensione europea; quanto piuttosto attribuire ulteriormente al soggetto privato non solo il faticoso compito di valutare *a priori* il "peso" di un diritto ma anche quello di dover svolgere un bilanciamento a tutto tondo⁵².

Per tale ragione, è stato opportunamente suggerito di integrare il criterio oggettivo con un parametro soggettivo che lasci spazio anche alla percezione individuale e collettiva di una determinata violazione. Tale operazione consisterebbe in uno sforzo di "traslazione normativa" del significato sociale attri-

49. Si fa riferimento, ad esempio, a una privazione prolungata di una libertà o la ricorrente violazione di un parametro in termini di contesto, soggetti coinvolti, modalità della violazione.

50. Come ribadito dalla stessa Corte di Giustizia, cfr. CJEU, C-741/21, 11 aprile 2024, in cui esclude che la violazione congiunta di più previsioni del Gdpr determini una plurima violazione del diritto, riconoscendo l'identità del medesimo bene tutelato.

51. Fra le tante pronunce riconducibili al tema si richiama la paradigmatica sentenza della Corte costituzionale n. 85/2013, punto 9 del *Considerato in diritto*.

52. Si pensi all'articolo 10 che tutela la libertà di espressione, il compito di individuare quali siano le circostanze che giustificano la violazione, seppur tassativamente indicate dall'articolo, comporterebbero che in fase di valutazione preventiva si prefiguri un bilanciamento fra diritti fondamentali (la classica questione fra libertà di espressione e tutela della dignità individuale).

buito a una violazione, attribuendo così un ruolo chiave, in relazione a una potenziale lesione di un diritto fondamentale, alla «*perception prevailing in society at a given time*»⁵³.

Dietro tale sforzo integrativo, che tenta di introdurre una «*normative assessment of social significance*» emergerebbe il tentativo di flessibilizzare il parametro oggettivo, dando alla Fria quella ulteriore elasticità che le consentirebbe maggior completezza in termini di precomprensione del danno potenziale⁵⁴. Si tratta, d'altra parte, della costruzione di un criterio estremamente delicata la quale, come osservato, può facilmente sovrapporsi con la valutazione delle conseguenze negative sul benessere del singolo. La granularità di tale parametro emerge maggiormente quando il parametro soggettivo rileva anche in chiave individuale, al fine di cogliere la percezione prettamente personale del danno eventuale.

Il parametro basato sulla valutazione degli effetti (c) supera la dimensione normativo-oggettiva e quella soggettiva e propone di classificare la «*severity of the expected consequences*» ossia il potenziale pregiudizio che un determinato sistema potrebbe causare nell'esercizio di un diritto fondamentale, utilizzando come valore di riferimento il benessere individuale, valutato alla luce della salute fisica e mentale e delle possibilità economiche eventualmente pregiudicate⁵⁵.

53. Si richiama, in tal modo, l'articolato intervento dell'Avvocato Generale Sánchez-Bordona, v. *Opinion of Advocate General Campos Sánchez-Bordona delivered on 6 October 2022, UI v Österreichische Post AG*, in cui si afferma «I am in no doubt that there is a fine line between mere upset (which is not eligible for compensation) and genuine non-material damage (which is eligible for compensation) and I am also aware of how complicated it is to delimit, in the abstract, the two categories and apply them to a particular dispute. That difficult task falls to the courts of the Member States, which will probably be unable to avoid in their rulings the perception prevailing in society at a given time regarding the permissible degree of tolerance where the subjective effects of infringement of a provision in this area do not exceed a de minimis level».

54. Così G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 12. Al fine del corretto inquadramento, la dottrina a supporto individua la consultazione di esperti indipendenti del settore come una via chiave soprattutto perché offrirebbero anche il supporto empirico ovvero basato su una valutazione effettuata sul gruppo maggiormente colpito rispetto a un determinato diritto. A tal scopo si richiama il principio 18, par. 1, lett. b, dei *Guiding principles on Business and Human Rights* tracciati dalle Nazioni Unite, in cui si afferma che «in order to gauge human rights risks, business enterprises should identify and assess any actual or potential adverse human rights impacts» sia opportuno effettuare «meaningful consultation with potentially affected groups and other relevant stakeholders, as appropriate to the size of the business enterprise and the nature and context of the operation».

55. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 17.

2.3. Alcune prospettive di concreta implementazione della Fria

Proprio in forza di tali preliminari presupposti teorico applicativi tracciati e delle criticità segnalate, un'ulteriore riflessione dottrinaria ha condotto alcuni rilevanti tentativi di implementazione concreta di tale modello: è emersa progressivamente un'ottica semi-quantitativa, la quale sembra tener conto proprio della peculiarità dell'oggetto da valutare, a cavallo fra dato tecnico e sfumature di valore.

Fra gli stessi si segnala, da un lato, l'interessante proposta volta a costruire un modello di Fria in tre momenti operativi⁵⁶. Il primo step consisterebbe in un inquadramento della tecnologia adoperata: un'analisi finalizzata a comprendere se la stessa possa qualificarsi come sistema ad alto rischio, ai sensi delle previsioni del regolamento. In caso di esito positivo, si passerebbe a due ulteriori momenti di valutazione⁵⁷.

Il primo, più strettamente tecnico, servirebbe a produrre il cosiddetto Qri, ovvero una valutazione del rischio sotteso al sistema di AI testato, individuato sulla base di un *range* da 1 a 10⁵⁸. Il Questionario serve a comprendere le caratteristiche intrinseche della tecnologia utilizzata, la sua applicazione e le sue caratteristiche operative⁵⁹. Per far ciò, vengono dunque selezionati *key risk factors* come *fairness*, *transparency*, *human oversight* e assegnato un punteggio che va da uno a 10 sulla base di una classifica basata su una scala predefinita (Qri). Se in questa fase emerge un valore oltre una determinata soglia la valutazione si blocca perché per il sistema quel parametro è già troppo alto e quindi concepito come inaccettabile. Questa prima valutazione, in sintesi, pone delle basi numeriche concernenti il potenziale rischio intrinseco della tecnologia e della sua applicazione sulle quali poi si edificherà la matrice che guarda ai diritti fondamentali⁶⁰.

Quest'ultima consiste nel secondo momento operativo chiave. La valutazione si struttura come precipuo test per comprendere l'impatto della tecnolo-

56. Il modello sperimentale elaborato da A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, in *MediaLaws*, 4 marzo 2025, spec. par. 3 ss.

57. Un modello di analisi ripreso e sviluppato analogamente in S. Bertaina *et al.*, *Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act*, in *Computer Law & Security Review*, vol. 56, 2025, p. 5 ss.

58. Cfr. A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit., par. 3 ss.

59. Proprio perché può essere somministrato prima o dopo il *deployment*, lo stesso quindi si sdoppia in un questionario elaborato in fase di sviluppo della tecnologia e poi in fase operativa e applicativa.

60. Cfr. A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit., par. 3 ss.

gia sui diritti, così come definiti dalla Carta dei diritti fondamentali dell'Unione. Esige, dunque, un metodo semiquantitativo che si fonda sull'attribuzione di *Impact Scores* (IS) rispetto al singolo diritto fondamentale esaminato. Per essere efficace la matrice deve:

1. andare oltre i dati di *training* del sistema;
2. avere come panorama di individui presi a riferimento un numero adeguato e vario di persone potenzialmente capaci di subire l'impatto, come previsto dall'articolo 10 dell'Aia.

Nello svolgimento di tale analisi, vengono presi in considerazione tutti i diritti fondamentali previsti dalla Cfreu ma viene svolta una valutazione separata per ciascun diritto. Per ogni diritto fondamentale della Carta viene elaborato un IS score che va da 0 a 100 (con 1 indicante il minimo rischio e 100 indicante il massimo rischio).

Infine, il test tecnico e la valutazione sui diritti fondamentali devono operare congiuntamente per offrire un risultato sinottico: il questionario offre informazioni qualitative e una comprensione situazionale della macchina, la matrice offre un'analisi semiquantitativa relativa a tutti i diritti fondamentali⁶¹.

La Fria può dirsi così conclusa nel momento in cui si perviene a un valore finale che sia il complesso dei valori raggiunti che combino il Qri con il IS in un unico indice⁶².

Di analogo rilievo e di simile approccio appare la metodologia sperimentale HH4AI (*Human Rights & Human-centered AI Impact Assessment*), concepita come un quadro strutturato in linea con i requisiti dell'AI Act ma di diretta ispirazione ai principali *framework* internazionali in materia di diritti umani. Essa nasce dall'esigenza di superare la frammentazione degli approcci esistenti, integrando dimensioni giuridiche, etiche e tecniche in un unico processo coerente⁶³.

La metodologia si fonda su un approccio di valutazione progressiva "a strati" (*gate-based*), che consente di filtrare progressivamente i rischi rilevanti e di concentrare l'analisi approfondita solo sugli aspetti che presentano un poten-

61. Se il punteggio è inferiore a 50 il sistema ha passato il test; se il punteggio è uguale o superiore a 75 è necessario richiedere una seria analisi delle sezioni interessate; se il punteggio è medio (fra 50 e 75) bisogna mettere in atto necessari interventi e riprogrammare le fasi 2 e 3.

62. Nel calcolare il risultato finale si attribuisce un peso maggiore (del 70%) al valore IS perché è reputata maggiormente idonea nel cogliere la granularità e pervasività del danno potenziale e ha una base quantitativa più oggettiva. Cfr. A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit. par. 3.3.

63. P. Ceravolo *et al.*, *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, cit. p. 9 ss.

ziale impatto significativo sui diritti fondamentali. Tale struttura risponde a un'esigenza di proporzionalità e di efficienza, evitando valutazioni eccessivamente onerose per sistemi che non presentano criticità rilevanti⁶⁴.

Il processo HH4AI si articola in quattro fasi principali. La prima consiste nella raccolta di informazioni generali sul sistema di IA oggetto di valutazione. Questo step ha l'obiettivo di delineare il contesto di utilizzo, le sue finalità, le principali funzionalità e gli attori coinvolti. Questa analisi consente di determinare se la tecnologia rientri tra quelli classificabili come ad alto rischio ai sensi dell'AI Act e di individuare, in via preliminare, i diritti fondamentali potenzialmente coinvolti⁶⁵.

La seconda fase introduce una *checklist* sui diritti fondamentali, concepita come uno strumento di *screening* preliminare concernente i beni tutelati. Tale disamina si basa sui principali diritti riconosciuti dalle Carte internazionali ed europee, tra cui la Carta dei Diritti Fondamentali dell'Unione Europea, la Convenzione Europea dei Diritti dell'Uomo e la Dichiarazione Universale dei Diritti Umani.

L'obiettivo è identificare rapidamente se, e in che misura, il sistema di IA possa incidere su specifici sfere e principi oggetto di tutela. La analisi consente di individuare quali diritti richiedano un'analisi approfondita nella fase successiva e quali, invece, possano essere esclusi dalla valutazione dettagliata, riducendo così la complessità del processo.

La terza fase rappresenta il cuore della metodologia HH4AI ed è dedicata alla valutazione approfondita dell'impatto del sistema di IA sui diritti fondamentali individuati come rilevanti. In questa fase, l'analisi si concentra su specifiche aree principali di impatto, quali, ad esempio, *accountability* e trasparenza, per le quali vengono stimate la probabilità e la gravità del potenziale impatto negativo sui diritti fondamentali, tenendo conto del contesto di utilizzo e delle misure di mitigazione già previste⁶⁶.

L'approccio adottato appare coerente con la logica *risk-based* dell'AI Act e consente di distinguere tra rischi accettabili, rischi che richiedono misure correttive e rischi non accettabili che potrebbero rendere il sistema incompatibile con i requisiti normativi.

Infine, l'ultima fase della metodologia HH4AI produce un *output* strutturato che sintetizza i risultati della valutazione e individua le misure di mitigazione necessarie per ridurre i rischi individuati. Quest'ultime possono includere misure quali raccomandazioni tecniche fino a interventi di natura più pret-

64. Ivi, p. 10.

65. Ivi, p. 9.

66. *Ibidem*.

tamente giuridica che supportino il *deployer* nell'adempimento degli obblighi stabiliti dal regolamento.

3. Conclusioni

Il percorso di riflessione sopradescritto promuove una prospettiva di ricomposizione teorica e di implementazione concreta fra le peculiari sfide del contesto virtuale e la natura immutabile dei diritti fondamentali, utilizzando la Fria come modello di un più generale orientamento normativo.

Tale andamento racchiude con sé alla radice, a tutta evidenza, l'ormai quasi decennale dibattito sul costituzionalismo digitale. Quest'ultimo, al di là delle sfaccettate definizioni emerse in dottrina, persegue l'ambiziosa finalità di traslare gli obiettivi sottesi alla legalità costituzionale e ai principi della *rule of law* nella dimensione dell'innovazione tecnologica e digitale⁶⁷. Il modello di *impact assessment* sopra analizzato non sarebbe altro, infatti, che un tentativo ulteriore, dopo le prescrizioni del Dsa⁶⁸, di implementazione, sempre più articolata e sofisticata, di istituti, principi e paradigmi garantistici tipici del costituzionalismo "classico" in presenza di attori, contesti e fenomeni differenti⁶⁹.

Dietro tale direzione legislativa emergerebbe, tuttavia, quella premessa imprescindibile secondo la quale i fenomeni che il decisore politico contemporaneo si trova a dover affrontare non costituirebbero problematiche nuove quanto piuttosto criticità e interrogativi sempre esistenti che la globalizzazione e l'affermazione della dimensione virtuale e tecnologica avrebbero semplicemente esasperato, sfruttandone la forza iper-generativa⁷⁰.

67. L. Gill, D. Redeker, U. Gasser, *Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights*, in *Berkman Center Research Publication*, n. 1, 2015, p. 1 ss.; G. De Gregorio, *The rise of digital constitutionalism in the European Union*, in *International Journal of Constitutional Law*, vol. 19, n. 1, 2021, p. 41 ss. In termini definitivi ugualmente centrale è stata la riflessione di E. Celeste, *Digital constitutionalism: a new systematic theorisation*, in *International Review of Law, Computers & Technology*, vol. 33, n.1, 2019, p. 81; per un apporto a tale dibattito più risalente cfr. V. Karavas, *Governance of Virtual Worlds and the Quest for a Digital Constitution*, in C.B. Graber, M. Burri-Nenova (eds.), *Governance of Digital Game Environments and Cultural Diversity: Transdisciplinary Enquiries*, Elgar, 2010.

68. G. De Gregorio, *The Digital Services Act: A Paradigmatic Example of European Digital Constitutionalism*, in *Diritti Comparati*, 17 maggio 2021.

69. Sull'evoluzione di tale percorso, ancora, cfr. G. De Gregorio, *Digital constitutionalism across the Atlantic*, in *Global constitutionalism*, n. 11, 2022, p. 297 ss. e il già citato lavoro monografico dello stesso autore cfr. Id., *Digital constitutionalism in Europe*, cit., p. 41 ss.

70. Sul punto cfr. A. Jr. Golia, *The Critique of Digital Constitutionalism*, in *Max Planck Institute for Comparative Public Law & International Law – MPIL*, Research Paper, 2022, p. 12.

Questo presupposto logico ha portato con sé una conseguenza: se la società non è costretta a confrontarsi con un problema intrinsecamente e sostanzialmente differente la risposta regolatoria può essere strutturalmente identica, ossia quella ereditata dalle tradizioni giuridiche degli Stati membri e accolta anche nella Carta dei diritti fondamentali dell'Unione la quale attinge, appunto, a strumenti garantistici del costituzionalismo europeo⁷¹.

Purtuttavia, l'adattamento dello strumentario di tutele proprie della dimensione *offline* continua a esser accompagnato, a oggi, da irrisolte criticità. Non si tratta soltanto di evidenziare come l'azione intrapresa dal legislatore unionale di circoscrivere l'impatto delle nuove tecnologie e i nuovi poteri privati mediante un quadro strutturato e organico di garanzie poste a tutela della persona rimanga esso stesso limitato e talvolta schiacciato sull'orizzonte procedurale, senza un adeguato supporto sostanziale⁷²; quanto, più precisamente, fare i conti con il fatto che alcuni istituti tipici dello Stato di diritto costituzionale che presiedono la limitazione di un diritto fondamentale appaiono solo parzialmente rimodellabili e implementabili, come dimostra il *framework* preso in esame.

Sebbene l'AI Act compia, in una visione sinottica, passi ulteriori nell'espandere e dettagliare le modalità di massimizzazione della tutela dei diritti⁷³, come già accennato, ugualmente presenta in alcune sue previsioni-chiave una debolezza insuperabile che intacca elementi costitutivi della *rule of law*, nonostante i tentativi di ibridazione e gli sforzi di traduzione applicativa emersi alla luce dell'ampia riflessione dottrinarica ricostruita.

Tali criticità operative concernenti il corretto inquadramento delle tecnologie e l'adozione delle idonee misure preventive imposte dal Regolamento verrebbero soprattutto esasperate dalla difficoltà di non riuscire a disporre di modelli di implementazione soddisfacenti di tali ambiziosi *impact assessments* che

71. O. Pollicino, *Libertà di espressione, piattaforme digitali e cortocircuiti di natura costituzionale*, in *Privacy&*, n. 1, 2021, p. 7.

72. Per lo sviluppo di questa tesi si veda l'intervento di L. Califano, *La strategia normativa dell'Unione europea per un nuovo ordine digitale*, nel fascicolo speciale a cura di L. Califano, F. Fabrizzi, G. Sartor, *Information disorder e sistema democratico*, in *federalismi.it*, n. 15, 2025.

73. Sul punto, *ex multis*, cfr. ancora la ricostruzione multi prospettica contenuta nel volume a cura di F. Fabrizzi, L. Durst, *Controllo e predittività*, cit., spec. pp. 7-69. In chiave prospettica si veda altresì M. Orofino, *Obiettivi, ambito di applicazione e principi fondamentali dell'AI Act*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, p. 35 ss.; si veda anche A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. I, il Mulino, 2022, p. 156.

risultino uniformi ed efficaci rispetto alla materia trattata. La Fria e le sue prime vicende e ipotesi adattive testimoniano la essenziale e primigenia difficoltà di coniugare modelli semiquantitativi di giudizio con l'ambito ontologicamente qualitativo dei diritti fondamentali, caratterizzati per loro natura dal loro essere situati e pronti a sfumature di valore⁷⁴.

Di fronte a tale quadro che pare sempre più connotato da slanci regolatori e tentativi pionieristici di costruzione di soluzioni applicative univoche, è difficile, a ogni modo, lasciarsi dietro la convinzione che tale paradigma possa reputarsi una via imperfetta ma più facilmente percorribile, data l'edificazione della stessa su radici garantistiche familiari e consolidate, le quali rassicherebbero in merito alla costruzione di un adeguato ancorché imperfetto sistema di salvaguardia dei diritti fondamentali nel cyberspazio⁷⁵.

Dall'altra parte, tuttavia, le opacità teoriche e la difficile traduzione applicativa che numerosi di tali istituti mostrano rendono sempre meno distanti quelle elaborazioni alternative che interrogano le asserite "camaleontiche" potenzialità del costituzionalismo: si pone in dubbio, in sintesi, la sua imperitura capacità di modellarsi, lasciando intatta la sua forza di reazione a nuove forme di potere.

Le numerose incertezze che avvolgono tale re-impianto valoriale in uno scenario con attori e fenomeni nuovi contribuirebbero a rafforzare, ad esempio, le contro argomentazioni che suggeriscono di vagliare le potenzialità del *societal constitutionalism*, alla base del quale giacerebbe proprio la convinzione dell'inadeguatezza dei paradigmi tradizionali nel poter reagire alle svolte prodotte dalla globalizzazione nell'epoca contemporanea. In quest'ultima la frammentazione dei tradizionali rapporti di potere imporrebbe di andare oltre i classici modelli fondati sulla *rule of law*, sulla separazione dei poteri e sul ruolo garantistico dell'autorità giudiziaria⁷⁶ in quanto sistemi reputati inadatti a far fronte a una sfera pubblica "polverizzata" in cui si ergono nuove forme di dominio nei confronti della persona. Tali fenomeni si identificherebbero ben oltre i tradizionali poteri politici e sociali ma si configurerebbero, nell'ottica di tale teorica, quali nuovi "centri di potere": un mutamento di attori che

74. Sul profilo di metodo cfr. E. Longo, *La disciplina del "rischio digitale"*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, p. 75 ss.

75. M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista "Gruppo di Pisa"*, n. 2, 2021, pp. 180-181; G. Palombella, *È possibile una legalità globale?*, il Mulino, 2012, p. 161.

76. Sui "nuovi" principi guida dell'era digitale, si veda la interessante proposta di E. Longo, A. Pin, *Oltre il costituzionalismo? Nuovi principi e regole costituzionali per l'era digitale*, in *Diritto pubblico comparato ed europeo*, n. 1, 2023, p. 12 ss.

esigerebbe di tagliare quel legame indissolubile fra costituzionalismo e soggetto pubblico statale⁷⁷.

Più chiaramente, dunque, in un contesto in cui l'ordinamento giuridico è ormai ben distante dalla sua formulazione originaria, e in uno scenario in cui si intersecano inedite forme di coercizione, molto più sfumate, diffuse e policentriche, nonché sistemi normativi sovrapposti, non è facile non volgere lo sguardo verso teorie antitetiche che fanno appello a quelle intrinseche capacità di produzione normativa di un sistema sociale.

Si immaginerebbe, in sintesi, la costruzione alternativa di una costituzionalizzazione dei processi che non passa necessariamente attraverso lo Stato e la sua architettura tradizionale ma si nutre dell'interazione fra nuovi attori che, attivamente o passivamente, divengono protagonisti di nuove forme di dominio.

In tale scenario, quindi, appare sempre più lecito contemplare, anche in un'ottica di integrazione fra modelli di pensiero, la convinzione che istituti, regole e garanzie non possano configurarsi soltanto in modo "responsivo" di fronte a fenomeni inediti, affrettandosi a adattare all'occorrenza l'armamentario garantistico a disposizione del costituzionalismo liberale; ma, in una prospettiva nuova, che siano e debbano risultare normativamente e attivamente mobili e reattivi in presenza di un esistente mutato⁷⁸.

77. Per un ulteriore approfondimento teorico cfr. A. Jr Golia, G. Teubner, *Societal Constitutionalism in the Digital World: An Introduction*, Max Planck Institute for Comparative Public Law & International Law – MPIL Research Paper, n. 11, 2023, p. 1 ss. Per una collocazione della teorica nel più ampio dibattito sul global constitutionalism, cfr. A. Wiener et al., *Global constitutionalism: Human rights, democracy and the rule of law*, in *Global Constitutionalism*, vol. 1, n. 1, 2012, p. 3 ss.

78. A. Jr. Golia, *The Critique of Digital Constitutionalism*, cit., p. 8.

VIOLENZA DIGITALE DI GENERE IN AMBITO POLITICO: PROFILI COMPARATI TRA UNIONE EUROPEA E MESSICO

Rosa Iannaccone*

SOMMARIO: 1. La violenza politica di genere nell'ecosistema digitale – 2. L'approccio europeo alla disciplina del "digitale" e la questione della violenza politica di genere – 3. Configurazioni normative sulla violenza politica di genere in America Latina – 4. La risposta messicana alla violenza politica e digitale di genere – 5. La violenza politica digitale di genere come sfida democratica.

1. La violenza politica di genere nell'ecosistema digitale

La violenza esercitata contro le donne presenta caratteristiche peculiari rispetto a quella subita dagli uomini. Gli atti violenti rivolti alle prime, infatti, sono frequentemente connotati da stereotipi di genere, da linguaggi o rappresentazioni sessualizzate e da condotte di natura sessuale, che difficilmente si riscontrano nei confronti degli uomini, oltre alla violenza fisica in senso stretto. Tali dinamiche si manifestano anche nello spazio politico. Le donne che vi partecipano mettono in discussione uno *status quo* storicamente plasmato da logiche maschili e da una cultura politica permeata da pratiche di dominio e da forme di machismo.

Le persistenze patriarcali, le norme culturali discriminatorie e le disuguaglianze economiche si sono tradotte in prassi che hanno limitato l'effettivo godimento dei diritti umani delle donne e alimentato molteplici forme di violenza, spesso accompagnate da stereotipi e pregiudizi che distorcono la loro immagine pubblica e la loro capacità di azione politica. Queste condotte, oltre a rivelare una resistenza nel riconoscere pienamente la legittimità delle donne come soggetti politici, aggravano gli ostacoli che le militanti dotate di leadership devono affrontare all'interno dei partiti¹.

* Contrattista di ricerca in Diritto pubblico comparato, Università degli Studi di Sassari.

1. F. Freidenberg, K.M. Gilas, *¿Normas poco exigentes? Los niveles de exigencia normativa de las leyes contra la violencia política en razón de género en América Latina*, in *Política y Sociedad*, n. 59(1), 2022, p. 3.

La violenza politica di genere rappresenta un fenomeno ampiamente diffuso in tutti i sistemi democratici.

Secondo i dati dell'Unione Interparlamentare, nel 2016 almeno l'80% delle legislative nel mondo ha dichiarato di aver subito episodi di violenza psicologica, economica, fisica o sessuale².

Dalla nascita della stampa, passando per l'affermazione della radio e della televisione, fino all'avvento di Internet e delle sue evoluzioni – in particolare il web 2.0 – la mediatizzazione ha inciso profondamente sulle forme di partecipazione degli individui e dei gruppi alla vita pubblica. Oggi l'esperienza sociale si costruisce sempre più come una narrazione mediata.

In tale contesto, la violenza politica di genere si manifesta anche attraverso le piattaforme digitali, divenute luoghi privilegiati per l'espressione di nuove – e spesso più pervasive – modalità di aggressione. L'uso dei social media e di altri strumenti digitali ha infatti favorito la diffusione di discorsi d'odio, molestie e disinformazione, contribuendo ad amplificare la violenza contro le donne attive nella sfera politica.

Le campagne diffamatorie, il cyberbullismo e la circolazione di contenuti falsi rappresentano solo alcune delle strategie impiegate per delegittimare e silenziare le donne, dimostrando come le tecnologie digitali, pur ampliando le opportunità di partecipazione, possano contemporaneamente rappresentare spazi di accresciuta vulnerabilità³.

Il fenomeno assume particolare rilevanza proprio alla luce dell'espansione dell'ecosistema digitale e della sua crescente centralità nelle dinamiche sociali e politiche. Nella sua declinazione online, la violenza politica di genere si configura come una delle manifestazioni più diffuse e normalizzate del contemporaneo disordine informativo. Nel contesto digitale, inoltre, i tre livelli che compongono tale disordine – disinformazione, malinformazione e misinformazione⁴ – tendono a sovrapporsi e intrecciarsi, contribuendo alla costruzione di un ambiente ostile che facilita la delegittimazione, l'intimidazione e l'esclusione delle donne dalla vita politica.

La creazione e la diffusione deliberata di contenuti falsi, ossia la disinforma-

2. Inter-Parliamentary Union, *Sexism, Harassment and Violence against Women Parliamentarians*, Ginevra, 2016.

3. A.G. Antunez Nuñez, C. Flores Seefoó, E. Hernández Hernández, C.R. Rudas Murga, *Violencia política de género mediante el uso de las plataformas digitales en México*, in *Tribunal*, vol. 4, n. 9, 2024.

4. Si v. C. Wardle, H. Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe Report, 2017; D.M. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M. Schudson, *The science of fake news*, in *Science*, n. 359/2018, pp. 1094-1096.

zione⁵, colpiscono in modo particolare le donne che ricoprono cariche pubbliche. Le campagne disinformative includono notizie false, spesso caratterizzate da una marcata dimensione sessualizzata o diffamatoria, e arrivano fino alla produzione di *deepfakes*. Quest'ultima tecnica, basata su sistemi di Intelligenza Artificiale, affonda le sue origini nei meccanismi di discriminazione e violenza di genere, in particolare nel contesto della pornografia non consensuale. I *deepfakes*, infatti, vengono frequentemente utilizzati per atti di revenge porn o come strumento di sextortion ai danni di soggetti vulnerabili, prevalentemente donne; non a caso, oltre il 95% dei contenuti presenti online rientranti in questa categoria consiste in video pornografici non consensuali⁶.

Anche la diffusione di informazioni vere, ma condivise con finalità malevole, la c.d. malinformazione, può essere impiegata per danneggiare la reputazione delle donne in politica, ad esempio mediante la violazione della loro privacy o la manipolazione strumentale del contesto. Vi è poi la misinformazione, ossia la circolazione di contenuti falsi diffusi senza l'intenzione di nuocere, che tuttavia contribuisce alla perpetuazione di stereotipi di genere e pregiudizi socialmente radicati.

È significativo osservare come, nell'ambiente digitale, i contenuti disformativi tendano spesso a essere rilanciati da soggetti diversi da coloro che li hanno originati, amplificandone la portata. Nell'ecosistema di internet, infatti, tali contenuti possiedono una forte capacità di propagazione, potendo assumere rapidamente connotati virali⁷.

L'espansione della partecipazione attraverso i social media ha favorito la nascita di veri e propri spazi digitali dedicati alla produzione e alla diffusione di discorsi d'odio. Tra questi, un ruolo di particolare rilievo è svolto dalla cosiddetta "manosfera", una sottocultura online improntata a ideologie di supremazia maschile e apertamente ostile al femminismo. In tali ambienti si difende un modello di mascolinità tradizionale e si alimentano narrazioni misogine che, per la loro ampiezza e capacità di diffusione, si sono consolidate come un fenomeno sociale significativo, in grado di influenzare il discorso pubblico e le dinamiche politiche⁸.

5. Si v. M. De Cock Buning et al., *A multi-dimensional approach to disinformation. Report of the independent High-level Group on fake news and online disinformation*, Publications Office of the European Union, 2018.

6. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, n. 1/2023, p. 173.

7. O. Pollicino, P. Dunn, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Bocconi University Press, 2024, p. 125.

8. Si v. in particolare E. Ortega Fernández, *La Manosfera y su evolución en la dinámica socio-política: una revisión sistemática de publicaciones académicas (2018-2023)*, in *European Public*

Se, in termini generali, l'infodemia⁹ – ossia la diffusione virale e incontrollata delle informazioni – rappresenta una minaccia per il corretto funzionamento dei processi democratici¹⁰, la violenza digitale di genere incide in modo ancora più diretto sulla partecipazione politica delle donne. Essa configura una forma specifica di violenza politica di genere, poiché ostacola l'esercizio dei diritti politici, limita la libertà di espressione e contribuisce a escludere le donne dagli spazi deliberativi e decisionali.

Come si vedrà, a livello dell'Unione europea il fenomeno non trova ancora un trattamento compiuto nella normativa dedicata al digitale, mentre rilevanti spunti di riflessione provengono dall'area latinoamericana, e in particolare dall'esperienza messicana. Prima di procedere a una sintetica analisi di tali esperienze ordinamentali, risulta tuttavia necessario soffermarsi sulla questione definitoria, elemento imprescindibile per affrontare in modo sistematico la violenza digitale di genere in ambito politico.

L'espressione "violenza digitale di genere in ambito politico" è infatti complessa e stratificata, poiché incorpora almeno tre dimensioni fondamentali: la discriminazione di genere, con specifico riferimento alla sfera politica; la violenza contro le donne; e l'utilizzo delle tecnologie digitali. Si darà quindi conto, preliminarmente, delle principali definizioni elaborate rispetto a ciascuna di queste componenti, al fine di costruire una nozione che risulti idonea a rappresentare il fenomeno oggetto del presente contributo. Le definizioni disponibili sono numerose e provengono in larga parte dall'ambito internazionale.

È noto che la Convenzione delle Nazioni Unite per l'eliminazione di ogni forma di discriminazione nei confronti della donna del 1979 definisce come discriminazione contro le donne qualsiasi distinzione, esclusione o limitazione fondata sul sesso che abbia l'effetto o lo scopo di compromettere o annullare il riconoscimento, il godimento o l'esercizio, sulla base della parità tra uomini e donne e indipendentemente dallo stato civile, dei diritti umani e delle libertà fondamentali in ambito politico, economico, sociale, culturale, civile o in qualsiasi altro settore (art. 1). Ai sensi della Convenzione di Belém do Pará

✂ *Social Innovation Review*, n. 10/2025. Sul contesto italiano v. D. Azzolari, V.M. Borodi, D. Garusi, *Il Forum degli Incel: la costruzione del consenso politico nella manofera italiana*, in *Rivista di Digital Politics*, n. 2/2021.

9. Termine utilizzato dall'Organizzazione Mondiale della Sanità (Oms) per riferirsi alla enorme quantità di informazioni, non tutte veritiere, diffuse a proposito del Coronavirus. v. F. López Veneroni, *De la pandemia a la infodemia: el virus de la infoxicación*, in *Revista mexicana de ciencias políticas y sociales*, vol. 66, n. 242, 2021.

10. L. Califano, F. Fabrizio, G. Sartor, *Information disorder e sistema democratico*, in *federalismi.it*, n. 15, 2025, p. iv.

(*Convenção Interamericana para Prevenir, Sancionar e Erradicar a Violência contra a Mulher*), adottata nel 1994 nell'ambito del sistema interamericano per prevenire, punire ed eliminare la violenza contro le donne, si intende per violenza contro la donna qualsiasi atto o condotta basato sul genere che provochi morte, danno o sofferenza fisica, sessuale o psicologica, sia nell'ambito pubblico sia in quello privato (art. 1). Questo strumento internazionale è stato il primo specificamente dedicato alla protezione delle donne contro la violenza e ha trasformato tale fenomeno da questione privata a responsabilità degli Stati e tema centrale in materia di diritti umani. Nel 2011, anche il Consiglio d'Europa ha adottato una Convenzione sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica, la c.d. Convenzione di Istanbul, la quale precisa che l'espressione "violenza nei confronti delle donne" designa una violazione dei diritti umani e una forma di discriminazione, comprendente tutti gli atti di violenza fondati sul genere che provocano o siano suscettibili di provocare danni o sofferenze fisiche, sessuali, psicologiche o economiche, incluse le minacce di tali atti, la coercizione o la privazione arbitraria della libertà, sia nella vita pubblica sia nella vita privata (art. 3).

Con riferimento alla dimensione digitale, il report di UN Women *Technology-facilitated Violence against Women: Towards a common definition* del 2022 definisce la violenza contro le donne facilitata dalla tecnologia come qualsiasi atto commesso, agevolato, aggravato o amplificato mediante l'uso delle tecnologie dell'informazione e della comunicazione o altri strumenti digitali, che provochi o possa provocare danni fisici, sessuali, psicologici, sociali, politici o economici, ovvero altre violazioni di diritti e libertà fondamentali. Già nel 2018, la Relatrice speciale sulla violenza contro le donne, le sue cause e conseguenze, in un rapporto dedicato alla violenza online contro donne e ragazze dal punto di vista dei diritti umani, aveva qualificato la violenza contro le donne come forma di discriminazione e violazione dei diritti umani sanciti dalla Convenzione del 1979 e da ulteriori strumenti internazionali e regionali. In tale prospettiva, la definizione di violenza online contro le donne si applica a qualsiasi atto di violenza di genere commesso, integralmente o parzialmente, mediante l'uso delle tecnologie dell'informazione e della comunicazione, o da esse aggravato, come telefoni cellulari, smartphone, internet, piattaforme di social media o posta elettronica. Nella stessa relazione si evidenzia, inoltre, che donne impegnate nella difesa dei diritti umani, giornaliste e politiche sono specificamente prese di mira, minacciate, molestate o persino uccise per il loro lavoro, ricevendo online minacce di natura misogina, spesso sessualizzate ed esplicitamente correlate al genere.

Anche l'Osce, nel 2022, attraverso la pubblicazione di un *toolkit* contenen-

te strumenti operativi e raccomandazioni rivolte a legislatori, governi, parlamenti e partiti politici, nonché indicazioni per la società civile e per le donne politiche vittime di violenza, ha sottolineato che la violenza contro le donne nella politica costituisce una minaccia non solo ai diritti individuali, ma anche alla democrazia, poiché indebolisce la *governance* democratica, il pluralismo politico e l'inclusività delle istituzioni. Infine, una risoluzione dell'Assemblea Parlamentare del Consiglio d'Europa (Pace) rileva che le donne in politica sono frequentemente bersaglio di discorsi d'odio online e di minacce che mirano ad allontanarle dalla sfera pubblica, compromettendo così la loro piena partecipazione democratica.

A livello internazionale, dunque, l'Onu, l'Osce e il Consiglio d'Europa hanno riconosciuto l'esistenza della violenza politica di genere, sebbene non sempre denominandola esplicitamente in questi termini. Hanno infatti evidenziato come gli attacchi digitali rivolti a politiche, giornaliste, attiviste e rappresentanti istituzionali costituiscano veri e propri strumenti di esclusione dal dibattito pubblico: attraverso campagne coordinate di odio, disinformazione e molestie online si alimenta un clima di intimidazione che induce molte donne ad autocensurarsi o a ridurre la propria presenza online; a rinunciare alla candidatura o al confronto politico pubblico; a vedere compromessa la propria competenza o credibilità istituzionale.

Alla luce di quanto delineato precedentemente, infine, la violenza digitale di genere in ambito politico può essere definita come l'insieme di comportamenti, pratiche e contenuti veicolati mediante tecnologie digitali che hanno lo scopo o l'effetto di intimidire, umiliare, silenziare o delegittimare le donne in quanto tali, colpendo la loro identità di genere o la loro espressione pubblica.

2. L'approccio europeo alla disciplina del “digitale” e la questione della violenza politica di genere

A livello europeo, il divieto di discriminazione di genere è sancito direttamente dai Trattati. In particolare, gli articoli 2 e 3 del Trattato sull'Unione europea (TUE) riconoscono la parità tra uomini e donne come principio fondamentale dell'Unione, attribuendole il compito di promuoverla in tutte le sue attività e di contrastare ogni forma di discriminazione fondata sul sesso. L'articolo 8 del Trattato sul funzionamento dell'Unione europea (TFUE) affida altresì all'Unione l'obiettivo di eliminare le disuguaglianze e di favorire la parità tra uomini e donne in tutte le sue azioni. Tali principi sono confermati anche dalla Carta dei diritti fondamentali dell'Unione europea: l'articolo 21 vieta qualsiasi forma di discriminazione, compresa quella basata sul sesso, mentre l'articolo

lo 23, dedicato espressamente alla parità tra uomini e donne, sancisce il diritto alla parità di trattamento e di opportunità in ogni settore¹¹.

Approfondendo il rapporto tra, da un lato, parità di genere e divieto di discriminazione in ambito politico e, dall'altro, piattaforme digitali e intelligenza artificiale, nel 2019 la Commissione europea ha adottato le "Linee guida etiche per un'IA affidabile", fondate su sette requisiti essenziali. Tra questi figurano la non discriminazione e l'equità, volte a prevenire i *bias* algoritmici, nonché il principio di responsabilità, che rimane in capo agli esseri umani che progettano o utilizzano sistemi di IA. Tali linee guida, tuttavia, non introducono misure operative in caso di violazioni e non hanno natura vincolante.

Il successivo Libro Bianco sull'Intelligenza Artificiale riconosce che l'uso crescente degli algoritmi all'interno dell'Unione comporta rischi specifici per la tutela dei diritti fondamentali, con particolare riguardo ai principi di uguaglianza e di non discriminazione. Tali criticità sono state richiamate anche nella Strategia per la parità di genere 2020-2025 della Commissione, che evidenzia come l'intelligenza artificiale possa accentuare le disuguaglianze di genere. In tale prospettiva, l'Unione ha invocato la costruzione di un "ecosistema di fiducia", nel quale l'intelligenza artificiale sviluppata e impiegata in Europa sia conforme ai valori e ai diritti fondamentali dell'UE, tra i quali il diritto all'uguaglianza e alla non discriminazione occupa una posizione centrale.

Ciò nonostante, i due principali strumenti normativi dell'Unione in materia – il Digital Services Act del 2022 sulle piattaforme online e l'AI Act del 2024, che stabilisce regole armonizzate in materia di intelligenza artificiale – non affrontano in modo esplicito il fenomeno della violenza digitale di genere, in particolare nella sua manifestazione in ambito politico.

Il Digital Services Act, pur non configurandosi come una normativa espressamente orientata alla dimensione di genere, mette comunque a disposizione strumenti utili al contrasto della violenza digitale, inclusa quella di genere in ambito politico. Tra gli elementi maggiormente innovativi vi è l'obbligo, posto in capo alle piattaforme, di valutare e mitigare i "rischi sistemici" connessi al funzionamento dei propri servizi. La Commissione include tra tali rischi, in

11. Si v. *ex multis* E. Palici Di Suni, *La tutela contro le discriminazioni nell'Unione Europea*, in *Minori giustizia*, n. 2/2008; D. La Rocca, *Principio di non discriminazione e approccio di genere nel diritto europeo*, in R. Palidda (a cura di), *Donne, Politica e Istituzioni – percorsi di ricerca e pratiche didattiche*, edit Press, 2012; D. Fisichella, *La parità di genere nell'ordinamento giuridico dell'Unione Europea*, in R. Palidda (a cura di), *Donne, Politica e Istituzioni – percorsi di ricerca e pratiche didattiche*, cit.; A. Sciortino, *L'uguaglianza di genere nell'UE: categorie giuridiche e tutele*, testo riveduto ampliato e corretto della Relazione tenuta al XXXIV Convegno annuale dell'Associazione italiana dei costituzionalisti "Uguaglianza e discriminazione nell'età contemporanea", Università degli studi di Bergamo, 15-16/11/2019.

modo esplicito: la diffusione di disinformazione e contenuti manipolatori; la violazione dei diritti fondamentali – tra cui rientrano l’uguaglianza di genere e la non discriminazione –, nonché l’impatto negativo sul dibattito pubblico e sui processi democratici.

Vengono imposti, altresì, obblighi specifici alle piattaforme digitali in tema di moderazione dei contenuti ritenuti nocivi. In particolare, esse sono tenute ad adottare misure proattive per contrastare i contenuti illegali, tra cui l’*hate speech* e la disinformazione; implementare sistemi di segnalazione efficaci e accessibili alle vittime; e garantire adeguati livelli di trasparenza sugli algoritmi e sui criteri di raccomandazione dei contenuti. In linea teorica, ciò consente il monitoraggio, il tracciamento e la tempestiva rimozione di contenuti sessisti, denigratori o violenti nei confronti delle donne.

Anche l’AI Act non contiene riferimenti espliciti e puntuali alla discriminazione di genere, collocandola piuttosto nella più ampia categoria delle discriminazioni e delle violazioni dei diritti fondamentali. Pur in assenza di una disciplina specifica, il regolamento riconosce che i sistemi di intelligenza artificiale non sono neutri e possono riprodurre, amplificare o automatizzare discriminazioni e abusi. Nel delineare i rischi connessi ai *bias* algoritmici, richiama più volte il principio di “non discriminazione” in diversi considerando, mentre i riferimenti diretti all’uguaglianza di genere risultano significativamente più circoscritti: essi compaiono nei Considerando 27 e 48 e, sul piano operativo, nell’articolo 95, che attribuisce all’AI Office e agli Stati membri il compito di valutare e mitigare gli impatti negativi dei sistemi di IA sui gruppi vulnerabili, includendo la dimensione di genere. Il testo menziona inoltre le “donne” nei Considerando 56 e 57, riconoscendo forme di discriminazione basate sul sesso e sull’orientamento sessuale. Sul piano precettivo, l’articolo 5 vieta l’impiego di sistemi che sfruttino la vulnerabilità delle persone in relazione all’età, alla disabilità o a una specifica situazione sociale o economica. Per quanto il genere non sia incluso tra tali fattori, resta possibile un’interpretazione estensiva del riferimento a una «specifica situazione sociale o economica»¹².

Le normative europee sul digitale, dal Digital Services Act all’AI Act, segnano un passaggio cruciale nell’approccio dell’Unione Europea al rapporto tra tecnologia, diritti fondamentali e dinamiche democratiche; tuttavia, esse

12. Si v. *ex multis* M. Peruzzi, *Il diritto antidiscriminatorio al test di intelligenza artificiale*, in *Labour & Law Issues*, vol. 7, n. 1, 2021; A. Karagianni, *The EU Artificial Intelligence Act through a Gender Lens*, Friedrich-Ebert-Stiftung, 2025; Id., *Gender in a stereo-(gender)typical EU AI law: A feminist reading of the AI act*, in *Cambridge Forum on AI: Law and Governance*, n. 1/2025; A. Santana González, L. Rampino, *A design perspective on how to tackle gender biases when developing AI-driven systems*, in *AI and Ethics*, vol. 5, 2025; P. Gambatesa, *Alla ricerca degli “anticorpi” contro le discriminazioni di genere nell’AI Act*, in *BioLaw Journal*, n. 1/2024.

si mantengono formalmente neutrali rispetto alla dimensione di genere e non affrontano in modo esplicito il fenomeno della violenza digitale di genere, né le sue ricadute sulla partecipazione politica delle donne.

Inoltre, la disciplina dell'Unione europea in materia di piattaforme digitali e intelligenza artificiale configura un articolato "pacchetto" regolatorio che riflette, tuttavia, una strategia di *meta-regulation* e di responsabilizzazione del settore privato, dalle piattaforme digitali ai *deployer* di sistemi di IA. Il legislatore europeo ha infatti adottato un approccio cooperativo pubblico-privato, che attribuisce ai *private governors* il ruolo di garanti attivi del rispetto della *rule of law* all'interno dei propri "ordinamenti" di riferimento. Tale evoluzione normativa, osservabile anche in una prospettiva diacronica, evidenzia il tentativo di "elevare" il privato all'esercizio di funzioni di natura pubblicistica¹³. Questo spostamento di responsabilità verso soggetti privati presenta effetti problematici rispetto al contrasto della violenza politica di genere, lasciando a decisioni gestionali interne, spesso automatizzate attraverso algoritmi non trasparenti, e a logiche prevalentemente economiche la funzione di prevenire, individuare e mitigare contenuti che possano configurare tale fattispecie.

3. Configurazioni normative sulla violenza politica di genere in America Latina

Se, in ambito europeo, la disciplina vigente si configura prevalentemente come una cornice normativa di principio – suscettibile di costituire un punto di partenza per affrontare i rischi sistemici connessi alle discriminazioni di genere nell'ambiente digitale – in America Latina, e in particolare in Messico, è stato sviluppato un approccio più diretto, normativamente tipizzato e orientato alla tutela immediata delle vittime. Prima di procedere all'analisi del caso messicano, risulta pertanto utile delineare alcuni cenni sulle origini e sulla regolamentazione della "violenza politica di genere" nella regione latinoamericana.

Questa forma di violenza è riconducibile alla mancanza di condizioni paritarie nella competizione politico-elettorale sotto il profilo del genere, fenomeno strettamente connesso ad altre disuguaglianze di carattere strutturale che caratterizzano gran parte delle società dell'America Latina. La variabile di genere costituisce l'elemento propulsore di tale illecito, le cui vittime sono le donne che partecipano attivamente alla vita politica e alla competizione elettorale. Tra le pratiche riconducibili a questo fenomeno rientrano, tra le altre: la costrizione delle elette a dimettersi affinché un uomo subentri al loro posto; la pre-

13. L. Califano, F. Fabrizzi, G. Sartor, *Information disorder e sistema democratico*, cit., p. vi.

sentazione di sole candidature femminili in distretti considerati perdenti; la comunicazione di dati falsi sull'identità delle candidate agli organismi elettorali; minacce rivolte alle donne elette; disparità nella distribuzione dei tempi radiofonici e televisivi – o una copertura mediatica negativa – durante la campagna elettorale o l'esercizio della carica; ostacoli al regolare svolgimento delle loro funzioni; divieti di espressione; diffamazione; molestie attraverso i media e i social network; insulti; calunnie; violenza sessuale; aggressioni fisiche; palpeggiamenti o altri approcci sessuali indesiderati; forme di dominio economico in ambito domestico e politico; persecuzioni ai danni di parenti o sostenitori, fino ad arrivare, nei casi estremi, al femminicidio. Tali condotte possono essere attuate dallo Stato o dai suoi agenti, dai partiti politici o dai loro rappresentanti, dai mezzi di comunicazione, dagli utenti dei social network, dagli opinion leader, nonché da qualsiasi persona o gruppo di persone, direttamente o per interposta persona, indipendentemente dal sesso degli autori.

La maggiore visibilità e rilevanza del problema è strettamente legata all'aumento della presenza femminile nella sfera politico-elettorale, reso possibile in particolare attraverso misure di azione positiva. Tale circostanza consente di affermare che gli avanzamenti normativi finalizzati a promuovere l'inclusione politica delle donne nella regione non si sono necessariamente tradotti in trasformazioni culturali profonde delle pratiche relative alla partecipazione femminile¹⁴.

In America Latina, tali forme di violenza rappresentano oggi uno dei principali ostacoli all'accesso delle donne alla cittadinanza e all'esercizio dei loro diritti¹⁵. Le evidenze empiriche mostrano come le molestie e/o la violenza politica di genere siano penetrate nell'agenda legislativa e governativa di numerosi Paesi della regione, determinando l'adozione di diverse tipologie di interventi normativi. Tra queste si annoverano: l'introduzione di leggi specificamente dedicate a tali fenomeni, come avvenuto in Bolivia nel 2012; il riconoscimento della violenza di genere all'interno di alcune istituzioni politiche, pur senza una definizione esplicita di molestia e/o violenza politica di genere, come avvenuto in Argentina nel 2010 e in El Salvador nel 2011; l'inserimento di una definizione specifica nelle leggi generali sulla violenza di genere, come in Argentina nel 2019, in Bolivia nel 2013, in Ecuador nel 2018, a Panama nel 2013, in Paraguay nel 2016 e in Uruguay nel 2018; la qualificazione della fattispecie

14. L. Albaine, *Marcos normativos contra el acoso y/o violencia política en razón de género en América Latina*, in F. Freidenberg, G. Del Valle Pérez, *Cuando hacer política te cuesta la vida. Estrategias contra la violencia política hacia las mujeres en América Latina*, UNAM, 2017, pp. 118-119.

15. F. Freidenberg, K.M. Gilas, *¿Normas poco exigentes? Los niveles de exigencia normativa de las leyes contra la violencia política en razón de género en América Latina*, cit., p. 1.

come reato elettorale tramite la riforma del Codice Elettorale, come accaduto in Ecuador nel 2020; e, infine, la sua integrazione in varie norme che disciplinano la competizione politico-elettorale, la violenza di genere e altre regolamentazioni relative alle istituzioni dotate di competenze e poteri d'azione in materia di violenza politica di genere¹⁶. È questa, in particolare, la strategia adottata dal Messico nel 2020, come si illustrerà nel prosieguo.

La legislazione adottata in America Latina presenta un carattere eterogeneo, poiché emergono differenze sia nella definizione della violenza politica di genere, sia nei meccanismi previsti per il coordinamento dei sistemi statali competenti in materia; nella determinazione e nell'efficacia delle sanzioni; nelle misure di protezione; e negli strumenti predisposti per il risarcimento delle vittime. Gli interventi normativi spaziano dall'adozione di normative particolarmente rigorose (come nel caso del Messico) a quelle di severità intermedia (Ecuador, Panama), mentre la maggior parte dei Paesi presenta ancora discipline deboli (Argentina, Bolivia, Brasile, El Salvador, Perù, Paraguay, Uruguay)¹⁷.

4. La risposta messicana alla violenza politica e digitale di genere

Il Messico costituisce oggi uno dei laboratori più avanzati in tema di violenza digitale di genere e di violenza politica di genere, grazie a un quadro normativo che riconosce tali fenomeni come categorie autonome e specifiche di violazione dei diritti umani delle donne. Tale sviluppo è il risultato di un contesto caratterizzato da livelli allarmanti di violenza di genere, in cui oltre 9 milioni di donne subiscono annualmente forme di violenza digitale e circa il 22% dichiara di essere stata vittima di molestie online¹⁸. A ciò si aggiunge la persistente incidenza della violenza politica di genere, che colpisce in modo sistematico candidate, elette e attiviste. L'Istituto Nazionale Elettorale (Ine) rileva che, in un arco temporale di quasi quattro anni, a decorrere dal 1° dicembre 2020, sono state sanzionate 470 persone – di cui 90 donne e 380 uomini – per violenza politica contro le donne basata sul genere¹⁹.

16. Onu Mujeres, Cim/Oca, *Violencia contra las mujeres en política en América Latina: mapeo legislativo y proyectos parlamentarios*, 2020, pp. 15-16.

17. F. Freidenberg, K.M. Gilas, *¿Normas poco exigentes? Los niveles de exigencia normativa de las leyes contra la violencia política en razón de género en América Latina*, cit., p. 3.

18. Inegi, *Estadísticas a propósito del Día Internacional de la Eliminación de la Violencia contra la Mujer*, n. 20/11/2025.

19. Ine, *Registro Nacional de Personas Sancionadas en Materia de Violencia Política contra las Mujeres en Razón de Género*, dato aggiornato al 24/11/2025 (12:01:34 p.m.).

Il Messico è il Paese dell'America Latina che ha adottato il maggior numero di riforme a livello federale per affrontare la violenza politica contro le donne, intervenendo sulle leggi generali relative alle istituzioni e alle procedure elettorali, ai partiti politici, ai reati elettorali, al sistema di ricorso in materia elettorale e all'accesso delle donne a una vita libera dalla violenza. Nel 2019 sono state modificate e integrate numerose disposizioni di leggi fondamentali al fine di disciplinare la violenza politica di genere, con l'obiettivo di costruire un quadro giuridico volto a prevenire, sanzionare ed eliminare tali condotte. In particolare, sono state riformate la Legge generale sull'accesso delle donne a una vita libera dalla violenza, la Legge generale sulle istituzioni e le procedure elettorali e la Legge generale sui partiti politici. Con la riforma della *Ley General en Materia de Delitos Electorales* è stata introdotta, come fattispecie autonoma, la violenza politica di genere: la norma punisce atti di intimidazione, discriminazione, minaccia, diffamazione o sabotaggio ai danni di donne candidate, elette o impegnate in attività politiche, quando tali condotte mirano a ostacolare o limitare la loro partecipazione. Le sanzioni previste includono misure penali e amministrative, oltre a strumenti preventivi e riparatori. L'attuazione è affidata a un sistema istituzionale integrato che comprende il Tribunale Elettorale del Potere Giudiziario della Federazione, competente per la risoluzione dei casi; l'*Instituto Nacional Electoral* (Ine), incaricato di monitorare le violazioni, promuovere la formazione politica con prospettiva di genere e realizzare campagne di prevenzione; nonché le autorità locali, responsabili dell'applicazione territoriale delle misure. Oltre alle sanzioni amministrative e penali, è previsto che l'autore o l'autrice della violenza perda i diritti politici relativi all'eleggibilità ed è stato istituito un registro pubblico nazionale delle persone sanzionate per tale reato.

Gli elementi che rendono il Messico l'ordinamento dotato della normativa più rigorosa in materia sono molteplici: una definizione ampia e non restrittiva degli atti di violenza politica di genere; l'individuazione chiara ed esplicita delle condotte che integrano il reato, articolate in ventidue fattispecie; un sistema di collaborazione stretta e coordinata tra istituzioni statali e attori politici, volto a garantire un'efficace assistenza alle vittime; la previsione di ordini di protezione, misure di prevenzione e garanzie di non ripetizione; strumenti di riparazione a favore delle vittime; e, infine, un complesso articolato di sanzioni, sia amministrative che penali, che include la perdita dell'eleggibilità per coloro che siano stati condannati per atti di violenza politica di genere²⁰.

Alla struttura normativa appena descritta si affianca, completandone la por-

20. F. Freidenberg, K.M. Gilas, *¿Normas poco exigentes? Los niveles de exigencia normativa de las leyes contra la violencia política en razón de género en América Latina*, cit., pp. 9-10.

tata sistemica, la c.d. “Ley Olimpia”, anch’essa espressione di un approccio particolarmente avanzato alla tutela delle donne contro le nuove forme di violenza. La riforma – nata da una mobilitazione sociale guidata da Olimpia Coral Melo Cruz e avviata, a livello federale e statale, a partire dal 2020 – costituisce infatti il primo intervento, nel panorama comparato, a riconoscere la violenza digitale come una forma autonoma di violenza di genere.

Questa normativa sanziona la diffusione non consensuale di contenuti sessuali, prevedendo aggravanti qualora la vittima sia minorenne o l’autore sia un partner o ex partner, e impone ai soggetti privati operanti nello spazio digitale obblighi puntuali di cooperazione con le autorità per la rapida rimozione dei materiali illeciti. L’innovazione principale non risiede soltanto nella penalizzazione della condotta, ma nel disegno complessivo della riforma, che integra la violenza digitale nel più ampio sistema di protezione delle donne, assicura procedure giudiziarie rapide ed efficaci e istituisce solidi meccanismi di coordinamento tra potere giudiziario, autorità di telecomunicazione e piattaforme digitali. In tale prospettiva, la Ley Olimpia si colloca coerentemente nel solco dei principi affermati dalla Convenzione di Belém do Pará (1994), che impone agli Stati l’obbligo di prevenire, punire e sradicare ogni forma di violenza contro le donne, comprese quelle emergenti nell’ambiente digitale²¹.

Nel contesto messicano, dunque, pur essendo disciplinate da testi legislativi distinti, la violenza digitale e la violenza politica di genere non costituiscono fenomeni separati, ma rappresentano i due estremi di un medesimo continuum. L’interazione tra la Ley Olimpia e la legislazione elettorale, come riformata nel 2020, dà vita a un ecosistema normativo coerente e integrato, nel quale la tutela delle donne si sviluppa su più piani – penale, amministrativo e istituzionale – senza perdere di vista la necessità di un cambiamento culturale profondo, essenziale per garantire l’effettività delle misure adottate.

5. La violenza politica digitale di genere come sfida democratica

Il digitale non introduce una forma di violenza “nuova”, ma amplifica e rende più pervasive dinamiche già esistenti, estendendole all’ambito della comunicazione e dell’immagine. La risposta legislativa messicana riconosce tale in-

21. Si v. *ex multis* A.E. Nava Garcés, *La violencia digital en México (Ley Olimpia)*, in *Revista Criminalia*, vol. 87, 2020; D. López Téllez, M.P. Fernández-Cuevas, A. Callejas-Téllez, *La ley Olimpia: Un punto de inflexión en la regulación normativa penal de la violencia digital y mediática en México*, in *Divulgare*, n. 9(18), 2022; D.J. Flores Ramírez, R. Pérez Ramírez, *Del Ciberactivismo a la Construcción de una Gobernanza Feminista. La Incidencia de la Ley Olimpia y la Ley Ingrid en México*, in *Enfoques de la Comunicación*, n. 14, 2025.

terconnessione e configura la violenza digitale non come un'eccezione, bensì come una "estensione tecnologica" della discriminazione politica tradizionale.

La violenza digitale, infatti, opera come veicolo e amplificatore della violenza politica, trasformando lo spazio online in un nuovo terreno di conflitto per il controllo dell'espressione e della rappresentanza pubblica delle donne. Campagne di disinformazione, diffamazione sessista, manipolazioni visive o la diffusione non consensuale di contenuti privati sono utilizzate per delegittimarne la credibilità, minarne l'autorità e scoraggiarne la partecipazione. Ne deriva un effetto strutturale: la compressione della partecipazione democratica femminile e la riproduzione di una cittadinanza diseguale.

La violenza digitale contribuisce così ad alimentare una discriminazione sistemica, erodendo i principi di uguaglianza e rappresentanza. Storicamente, la conquista del suffragio femminile nella prima metà del Novecento non è stata sufficiente a superare quello che Pateman ha definito il "contratto sessuale"²², che per secoli ha sancito una gerarchia tra uomini e donne, riservando ai primi l'accesso alla sfera pubblica e confinando le seconde a quella privata e affettiva. Oggi, le nuove tecnologie e l'intelligenza artificiale rischiano di ampliare ulteriormente tale divario. È infatti nello spazio digitale che, sempre più frequentemente, si forma l'opinione pubblica e si costruisce la reputazione delle persone, rendendo questo ambito uno dei luoghi decisivi della vita democratica contemporanea. Per questo sarebbe essenziale che la dimensione di genere divenisse un criterio centrale nella valutazione della qualità democratica dell'ecosistema digitale e degli ordinamenti contemporanei.

La violenza politica di genere, soprattutto nella sua manifestazione digitale, non costituisce soltanto un problema di sicurezza individuale, ma rappresenta una minaccia strutturale per la democrazia, incidendo direttamente sulla libertà di espressione, sull'integrità personale e sulla partecipazione delle donne alla vita politica.

22. Si v. C. Pateman, *Il contratto sessuale. I fondamenti nascosti della società moderna*, Moretti&Vitali, 2015.

L'INFORMAZIONE NELL'ERA DELL'ALGORITMO: INTELLIGENZA ARTIFICIALE E LIBERTÀ DI ESPRESSIONE NELLA GOVERNANCE EUROPEA

Giulia Napoli*

SOMMARIO: 1. La diffusione di notizie false nello spazio digitale: dal mito della rete alla crisi dell'informazione – 2. Intelligenza artificiale e manipolazione del consenso: la democrazia nella società algoritmica – 3. Dalla *soft law* alla *Constitutional governance*: il modello europeo di contrasto alla disinformazione *online* tra Dsa e AI Act – 4. *Pars construens*: l'IA al servizio del pluralismo informativo.

1. La diffusione di notizie false nello spazio digitale: dal mito della rete alla crisi dell'informazione

Ogni ordinamento democratico si costruisce attorno a un presupposto tanto semplice quanto decisivo: che il processo di formazione della volontà collettiva poggi su un'informazione libera, pluralista e attendibile¹. Quando tale pre-

* Dottoranda di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Milano-Bicocca.

1. La formula riflette l'idea di democrazia deliberativa cara, tra gli altri, a J. Habermas (al quale si rimanda, in particolare in *Storia e critica dell'opinione pubblica*, alias *Strukturwandel der Öffentlichkeit*, 1962; Laterza, 2005), J. Rawls, R. Dahl e S. Rodotà, secondo i quali il consenso democratico non è autentico se non si fonda su un dibattito pubblico informato. Più di recente, questo assunto è stato sviluppato da M. Luciani, *Libertà di espressione e pluralismo politico nella giurisprudenza della Corte costituzionale*, Corte Costituzionale, 2025, ove si legge che «senza pluralismo non c'è democrazia e senza libertà di espressione non c'è pluralismo», perché «fra libertà di espressione, libertà di informazione, pluralismo e democrazia vi [è] una corrispondenza biunivoca». Si rimanda altresì a G. Sartori, *The Theory of Democracy Revisited*, Chatham House, 1987, ove egli scrive che «il potere elettorale è di per sé la garanzia meccanica della democrazia; ma la garanzia sostanziale è data dalle condizioni sotto le quali il cittadino acquisisce l'informazione». Sul rapporto tra pluralismo informativo e nuove tecnologie, si vedano anche, tra gli altri, G.E. Vigevari, *I media di servizio pubblico nell'età della rete. Verso un nuovo fondamento costituzionale, tra autonomia e pluralismo*, Giappichelli, 2018; A. Lamberti, *Libertà di informazione e democrazia ai tempi della società digitale: problemi e prospet-*

supposto si incrina, anche la sovranità popolare vacilla, poiché si altera la capacità dei cittadini di partecipare alla vita democratica consapevolmente. In questo delicato equilibrio tra libertà e verità, tra informazione e potere, si colloca il fenomeno della disinformazione: non una novità del nostro secolo o un portato della contemporaneità, ma una costante della storia e della comunicazione politica, che le tecnologie digitali e l'intelligenza artificiale hanno reso più pervasiva ed efficace, fino a trasformarla in una minaccia sistemica per l'integrità del dibattito pubblico e, di riflesso, per la salubrità dei processi democratici².

Già Marc Bloch, nel celebre saggio *Réflexions d'un historien sur les fausses nouvelles de la guerre* (1921) osservava come le "fausses nouvelles" non costituissero soltanto l'espressione di meri difetti di percezione o di inaccuratezze comunicative, bensì un fenomeno politico e sociale, che rivela le credenze, le paure e le passioni di un'epoca e che risulta capace di incrinare il rapporto fiduciario tra cittadini e istituzioni³. Quell'intuizione, lungi dal risultare ormai datata, conserva del tutto intatta la propria attualità. Tuttavia, se nel passato la diffusione del falso era limitata dalla lentezza dei mezzi analogici di comunicazione, oggi la velocità e la capillarità della rete – alle quali si aggiunge la "deprofessionalizzazione" dell'informazione, favorita dalla facilità di accesso a strumenti di pubblicazione più immediati e dalla progressiva erosione del ruolo degli intermediari tradizionali – amplificano esponenzialmente il rischio di manipolazione dell'informazione. Il risultato è un ecosistema informativo frammentato, certamente polifono ma forse eccessivamente rumoroso, in cui la distinzione tra fonti autorevoli e contenuti amatoriali si assottiglia e l'attenzione dell'utente diventa una risorsa scarsa, contesa da logiche di mercato⁴.

In questo scenario si inseriscono le grandi piattaforme digitali – alcune delle quali molto note, quali Facebook, X, YouTube, TikTok –, attori privati divenuti in poco tempo parte integrante dell'infrastruttura pubblica della comunicazione e *tertium* nel tradizionale rapporto tra Stato e cittadino⁵. Per la loro

tive, Consulta Online, II, 2022; R. Mastroianni, O. Pollicino, M. Bassini (a cura di), *Il T.U. dei servizi di media audiovisivi*, Giuffrè, 2024.

2. Sul rapporto tra pluralismo, libertà di espressione e nuove tecnologie si veda M. Bassini, *Intelligenza Artificiale generativa e media*, in R. Mastroianni, O. Pollicino, M. Bassini (a cura di), *Il T.U. dei servizi di media audiovisivi*, Giuffrè, 2024, pp. 269-293.

3. M. Bloch, *Réflexions d'un historien sur les fausses nouvelles de la guerre*, 1921. Nella traduzione del saggio del 1997 edita Einaudi, dal titolo *Riflessioni d'uno storico sulle false notizie della guerra*, si parla proprio di "percezioni distorte", magari anche giuste al principio «ma mal interpretate – unanimemente stravolte per accordarsi agli ardenti desideri di tutti».

4. F. Nanni, *I rischi dell'utilizzo dell'IA sul diritto di informazione – La disinformazione*, in *Rivista di diritto dei media*, I, 2025.

5. Il riferimento imprescindibile qui è J. Balkin, *Free speech is a triangle*, in *Columbia Law Review*, n. 118/2011.

capacità di connettere, selezionare e diffondere contenuti su scala globale, esse si configurano di fatto come la principale *agorà* del discorso pubblico, lo spazio in cui si formano le opinioni, si consolidano le credenze e si plasmano gli orientamenti politici.

Ma proprio in virtù di tale indubbia centralità, che le ha rese le vere e proprie *gatekeepers* dell'informazione⁶, tali piattaforme meritano oggi un'attenzione particolare. In effetti, le *digital platforms* non si limitano a svolgere – o, quantomeno, non più – una funzione di intermediazione neutrale dei contenuti immessi dagli utenti: il principio di neutralità della rete, che in origine le voleva meri vettori del flusso comunicativo, scerve da interventi di moderazione e immuni da qualsivoglia responsabilità, ha ceduto il passo a una realtà assai diversa, nella quale gli operatori digitali assumono un ruolo attivo di selezione, filtraggio e orientamento dei messaggi⁷. Tale potere – talvolta esercitato attraverso sistemi algoritmici automatizzati di *content moderation* sulla base dei *community standards* – comporta, nei fatti, la possibilità di sopprimere, oscurare o declassare contenuti e profili sulla base di criteri non sempre trasparenti. In questo senso, le piattaforme si trovano a esercitare un potere di incidenza sulla libertà di espressione che, pur muovendosi formalmente nel perimetro del diritto privato, assume effetti sostanzialmente pubblici: esse decidono, infatti, quali discorsi restano visibili, quali vengono rimossi e, in ultima analisi, quali idee possono circolare nello spazio digitale.

Quel che ne deriva è un rischio strutturale di “censura privata”, ossia di limitazione del discorso pubblico non imposta da un'autorità pubblica, ma da regole contrattuali e algoritmiche – spesso opache e comunque governate da logiche di mero profitto – che sfuggono a ogni forma di controllo democratico e che, in concreto, determinano in misura sempre crescente la forma e la qualità del dibattito pubblico⁸.

Eppure, per struttura e per logica di funzionamento, la rete avrebbe dovuto ampliare e rendere più immediata – e in parte effettivamente lo ha fatto – la fruizione da parte dei cittadini dei diritti e delle libertà fondamentali, in particolare la libertà di manifestazione del pensiero e il correlato diritto all'informazione: l'accesso diretto a una quantità potenzialmente illimitata di fonti,

6. Questa fortunata espressione è stata istituzionalizzata dal Digital Markets Act del 2022: sul punto, si veda C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, n. 3/2023, pp. 558-559.

7. Si veda G.E. Vigevani, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritti di libertà e nuove tecnologie*, a cura di A. Morelli, *Diritto Costituzionale*, 1, VI, 2023.

8. Sul tema della cosiddetta “privatizzazione della censura” si rimanda a M. Monti, *Privatizzazione della censura e Internet Platforms: la libertà d'espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, n. 1/2019.

la possibilità di produrre e diffondere contenuti in tempo reale e la riduzione delle barriere all'ingresso nel dibattito pubblico sembravano tradurre in pratica l'ideale che aveva caratterizzato Internet ai suoi albori. Quest'ultimo, infatti, incarnava l'immagine di un'arena orizzontale, capace di emancipare la parola dai vincoli della censura e dai condizionamenti del potere politico ed economico, divenendo simbolo di una libertà finalmente svincolata da mediazioni istituzionali. Sintetizza perfettamente questa immagine la *Declaration of the Independence of Cyberspace* del poeta statunitense John Perry Barlow, nella quale egli scriveva «Governments of the Industrial World, you weary giants of flesh and steel [...] You have no sovereignty where we gather», cristallizzando quella fiducia libertaria nel *cyberspazio* come nuova frontiera dell'autodeterminazione collettiva e individuale⁹.

Tuttavia, come si è detto poc'anzi, l'esperienza ha poi rivelato il rovescio di quell'originaria promessa, mostrando il definitivo tramonto del mito della rete come spazio neutrale e incontaminato. Alla visione di un'arena comunicativa aperta e orizzontale si è sostituita la realtà di un ecosistema informativo governato da pochi attori privati, titolari di un potere inedito di selezione e orientamento dei contenuti. Le stesse infrastrutture che avrebbero dovuto "liberare" la parola si sono trasformate in dispositivi di filtraggio e profilazione, nei quali la visibilità dei messaggi è determinata da logiche di mercato più che da criteri di rilevanza o veridicità delle informazioni. A ciò si aggiunge un ulteriore elemento di criticità: la rete, che in astratto avrebbe dovuto ampliare lo spazio della libertà comunicativa, si è rivelata anche un potente veicolo di contenuti che, pur non necessariamente illeciti, si dimostrano parimenti nocivi o manipolatori. La combinazione tra l'accesso illimitato, l'assenza di "filtri" professionali e la presenza di algoritmi di amplificazione ha reso più agevole la diffusione di messaggi falsi, distorti o deliberatamente ingannevoli, capaci di incidere non solo sull'informazione individuale, ma sull'integrità stessa del dibattito democratico.

Per qualche anno, le istituzioni hanno osservato questo processo con una prudenziale inerzia, ritenendo, tra l'altro, che la rapidità dell'innovazione tecnologica rendesse inadeguato un intervento normativo, che sarebbe repentinamente tramutato in un atto obsolecente. Di qui il ricorso a forme di autoregolamentazione, con le quali le piattaforme si sono impegnate – su base volontaria – a contrastare la disinformazione e i discorsi d'odio attraverso codici di condotta e *policy* interne.

Questa strategia, tuttavia, ha presto mostrato i propri limiti: l'assen-

9. J. Perry Barlow, *A Declaration of the Independence of Cyberspace*, in *Stanford Law Review*, 1996.

za di controlli esterni e di criteri uniformi ha reso infatti evidente che la *self-regulation* sia una pratica meramente discrezionale e, come tale, suscettibile di decisioni arbitrarie e spesso guidate da interessi economici piuttosto che da principi posti a tutela dei diritti fondamentali. Quel che ne deriva, come anticipato poc'anzi, è una forma di privatizzazione della censura, nella quale attori privati esercitano poteri pubblici senza il corrispettivo di adeguate garanzie di trasparenza o di ricorso.

È proprio per superare tale squilibrio che l'Unione europea ha progressivamente elaborato un modello alternativo, fondato su una logica di coregolamentazione e, più di recente, sull'adozione di strumenti di *hard law*¹⁰. Diversamente dall'approccio statunitense – ancorato a una concezione negativa della libertà di espressione – e da quello cinese, centrato sul controllo politico del discorso, l'Europa ha perseguito un paradigma *rights-based*, nel quale libertà e responsabilità coesistono come poli inscindibili della cittadinanza digitale. Il percorso che conduce dal Codice di condotta del 2018 fino al Digital Services Act¹¹ e all'Artificial Intelligence Act¹² riflette questa ambizione: riportare la *governance* della comunicazione *online* entro il perimetro del diritto, garantendo la responsabilità algoritmica, la trasparenza dei processi decisionali automatizzati e il bilanciamento tra libertà di espressione, pluralismo e protezione degli utenti.

Da questa tensione prende le mosse la domanda che orienta il presente contributo: le misure oggi vigenti – in particolare il Dsa e l'AI Act – sono davvero idonee a garantire un'informazione pluralista e affidabile nell'ecosistema digitale o rischiano invece di tradursi in un sacrificio eccessivo della libertà di manifestazione del pensiero?

2. Intelligenza artificiale e manipolazione del consenso: la democrazia nella società algoritmica

Come anticipato nel paragrafo che precede, una dialettica democratica salubre presuppone un circuito informativo aperto, pluralista e quanto più possibile attendibile. Quando tale circuito viene saturato dall'immissione di con-

10. A dire il vero, prima dell'adozione del Digital Services Act non vigeva soltanto un regime di *self-regulation*, essendo già stata emanata la Direttiva n. 31 del 2000 nota come Direttiva “*E-commerce*”. Pur rimanendo una forma di regolamentazione piuttosto debole in materia, infatti poi superata con estrema velocità dall'evoluzione tecnologica, essa rappresentava senz'altro un intervento di *hard law* degno di menzione.

11. Regolamento UE n. 2065 del 2022.

12. Regolamento UE n. 1689 del 2024.

tenuti falsi o fuorvianti, il diritto dei cittadini a essere informati subisce un *vulnus* che incide sulla libera formazione della volontà popolare, momento essenziale per l'esercizio della relativa sovranità. In quest'ottica, l'informazione, tradizionalmente intesa come "cane da guardia della democrazia" o "quarto potere" dello Stato¹³, svolge un ruolo imprescindibile in ogni ordinamento costituzionale, perché «soltanto un'opinione pubblica consapevole, che si nutre di un processo costante di diffusione delle idee, può dare effettività al diritto dei cittadini di concorrere alla determinazione della politica nazionale», fornendo loro gli strumenti – e dunque le informazioni necessarie – per esercitare consapevolmente i propri diritti politici¹⁴.

In questa prospettiva, la disinformazione non si esaurisce soltanto in un problema di accuratezza delle notizie riportate o di veridicità dei fatti esposti, ma rappresenta un fenomeno delicato, capace di incidere direttamente sulla legittimazione del potere democratico.

A rendere ulteriormente complesso il già precario ecosistema informativo è intervenuto lo sviluppo dei sistemi di intelligenza artificiale, che ha favorito la produzione e la successiva proliferazione bulimica di contenuti sintetici, amplificando i rischi di una cattiva informazione. Queste tecnologie – oggi impiegate tanto nella forma generativa, capace di creare autonomamente contenuti testuali, visivi o sonori di elevata verosimiglianza, quanto in quella moderativa, utilizzata dalle piattaforme per determinare la visibilità, l'ordine e la rilevanza nella modalità di diffusione dei contenuti condivisi dagli utenti, secondo criteri di mercato e di massimizzazione dell'*engagement* – esercitano un'influenza crescente sul modo in cui le informazioni vengono generate, veicolate e, soprattutto, recepite.

E proprio l'interazione tra IA generativa e IA moderativa incide profondamente sulla struttura del discorso pubblico, perché ciò che deriva dalla loro combinazione è un ambiente comunicativo polarizzato, nel quale i contenuti più divisivi e sensazionalistici – spesso falsi o fuorvianti – prevalgono su quelli più complessi o analitici, alimentando un processo di radicalizzazione delle opinioni e di impoverimento del confronto pubblico. La disinformazione non appare più come una deviazione patologica, ma come un effetto fisiologico di un ecosistema informativo governato da logiche di attenzione e profitto.

13. Come noto, si tratta di principi ormai radicati nella giurisprudenza costituzionale italiana e, soprattutto, della Corte di Strasburgo, ove la stampa e i giornalisti in particolare sono definiti i "*public watchdog*" della democrazia.

14. G.E. Vigevari, *Informazione e potere*, in *Enciclopedia del diritto*, a cura di M. Cartabia e M. Ruotolo, Giuffrè, V, 2023, p. 230.

I rischi di manipolazione si accentuano nei momenti di competizione elettorale, quando la sovrapposizione tra tecniche di *microtargeting* politico, *profiling* comportamentale e uso coordinato di *social bot* o *cyborg accounts* consente di orientare in modo mirato la comunicazione verso specifici segmenti specifici di elettorato. Le elezioni presidenziali statunitensi del 2016, il referendum sulla Brexit e, più di recente, le elezioni rumene del 2024 mostrano quanto fragile sia la linea di confine tra libertà di informazione e manipolazione del consenso: anche in assenza di prove definitive in merito a esiti elettorali concretamente alterati dall'impiego di sistemi di IA, il rischio di interferenze mirate appare ormai sempre più tangibile¹⁵.

In questo orizzonte, assume rilievo la scelta del legislatore europeo di riconoscere, all'interno del Regolamento europeo sull'intelligenza artificiale, i processi democratici come bene giuridico meritevole di protezione. Pur non rappresentando una normativa "sulla disinformazione", il regolamento individua nelle classificazioni per rischio come particolarmente sensibili i sistemi di IA idonei a influenzare l'esito delle elezioni o a orientare il comportamento di voto, di fatto riconoscendo che la tutela della democrazia non si esaurisce nel diritto formale al suffragio, ma include anche le condizioni cognitive e comunicative che ne garantiscono l'effettività.

Tale impostazione comporta due importanti implicazioni costituzionali: in primo luogo, questa anticipa l'intervento della tutela. La protezione dei diritti politici, infatti, non si limita più al momento dell'espressione del voto, ma si estende alla qualità del circuito informativo che lo precede e lo rende possibile. In secondo luogo, essa solleva il problema della misurabilità del pregiudizio: il modello di regolazione del rischio, centrato su nessi causali e soglie di danno individuale, fatica a cogliere gli effetti diffusi e cumulativi che erodono la qualità della deliberazione pubblica senza incidere immediatamente sul singolo individuo.

È in virtù di questa tensione – tra tutela preventiva e garanzia sostanziale, tra rischio individuale e rischio sistemico – che ci si interroga sulle modalità di preservazione del diritto dei cittadini a essere informati e, conseguentemente, l'integrità dei processi democratici in un ecosistema informativo "mediato" dall'intelligenza artificiale, senza con ciò sacrificare la libertà di espressione e il pluralismo informativo.

E il punto di raccordo tra queste istanze, allo stato, pare doversi ricercare nel quadro regolatorio ora vigente e, in particolare, nel rapporto tra Digital Services Act e AI Act.

15. G.E. Vigevari, *Informazione e potere*, cit., p. 230.

3. Dalla *soft law* alla *Constitutional governance*: il modello europeo di contrasto alla disinformazione *online* tra Dsa e AI Act

Una volta delineato il perimetro del problema, chiarito perché la disinformazione costituisca oggi una minaccia più grave rispetto al passato ed evidenziate le sue ricadute sui fondamenti costituzionali della democrazia rappresentativa, occorre interrogarsi sulle risposte che l'Unione europea ha progressivamente elaborato per fronteggiare il fenomeno e sulle peculiarità che distinguono l'approccio europeo rispetto ad altri contesti ordinamentali.

Svanito, infatti, l'iniziale entusiasmo che aveva accompagnato la nascita del *cyberspazio*¹⁶ – un fervore che, come accade nelle svolte epocali, si intrecciava a timori distopici e ad allarmismi fisiologici – si è gradualmente fatta strada la consapevolezza dell'esigenza di intervenire sul “lato oscuro della rete”¹⁷: la diffusione di contenuti nocivi, l'incitamento all'odio, la radicalizzazione dei discorsi pubblici e, più di recente, la manipolazione informativa integrata mediante i sistemi di intelligenza artificiale hanno progressivamente rivelato la necessità di un intervento regolatorio volto a limitare i rischi generati da simili fenomeni¹⁸.

Nei primi anni del dibattito europeo sul tema, la scelta – lo si è anticipato – è stata quella del *laissez-faire*, giustificata da una pluralità di ragioni: innanzitutto, le tendenze anarcoidi che avevano connotato l'internet degli albori; in secondo luogo, la fiducia nella capacità autocorrettiva del libero mercato delle idee; ancora, la percezione che la rigidità del diritto positivo fosse intrinsecamente incompatibile con la mutevolezza e la rapidità evolutiva delle tecnologie digitali, a fronte di operatori privati dotati di risorse tecniche e capacità di intervento più immediate; infine, la convinzione che una normativa vincolante potesse compromettere la libertà d'impresa di tali attori e, in ultima analisi, incidere sulla stessa libertà di espressione.

A ciò si aggiungeva la difficoltà di mettere a fuoco e di inquadrare giuridicamente quella «zona grigia» che connota in misura peculiare i contenuti disinformativi¹⁹: messaggi ambigui o “*borderline*”, formalmente leciti – perché non

16. Sul punto, si veda Christopher S. Yoo, *Free Speech and the Myth of the Internet as an Unintermediated Experience*, in *George Washington Law Review*, vol. 78, n. 4, 2010; M. Bassini, *Fundamental rights and private enforcement in the digital age*, in *European Law Journal*, n. 25, 2019, pp. 184-187.

17. Per un approfondimento sul tema, si veda, tra gli altri, A. Missiroli, *The Dark Side of the Web: Cyber as a Threat*, in *European Foreign Affairs Review*, n. 24, 2019, pp. 135-152.

18. D. Keller, *Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users*, in *The University of Chicago Law Review*, 2022.

19. G.E. Vigevari, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritti di libertà e nuove tecnologie*, cit., p. 44.

lesivi di altri diritti o libertà costituzionalmente rilevanti – ma potenzialmente idonei a compromettere la qualità del dibattito pubblico e ad alterare il regolare svolgimento dei processi democratici. Tutti questi elementi indussero le istituzioni europee a ritenere, in un primo momento, «che l'autoregolamentazione dei titolari delle piattaforme bastasse a contrastare la disinformazione»²⁰, evitando il rischio che un intervento normativo eccessivamente prescrittivo si traducesse in un ostacolo alla libertà di manifestazione del pensiero.

Su tali premesse si fondava il Codice di condotta sulla disinformazione del 2018²¹, espressione di quella regolamentazione volontaria che impegnava le principali piattaforme digitali a introdurre misure interne di contrasto alla diffusione di notizie false, a migliorare la trasparenza nella pubblicità politica e a collaborare con i verificatori di fatti indipendenti.

Questo primo tentativo, tuttavia, rivelò ben presto i propri limiti strutturali: l'assenza di obblighi giuridicamente vincolanti, la mancanza di meccanismi sanzionatori, l'opacità delle procedure di attuazione e la sostanziale discrezionalità nella definizione degli standard ridussero l'efficacia del codice a quella di un mero strumento reputazionale.

Nel frattempo, la centralità assunta da questi attori rese evidente l'ampiezza del potere da essi esercitato. È oggi innegabile che il concreto esercizio della libertà di espressione e il godimento effettivo del diritto all'informazione dipendano, in larga misura, dalla presenza dell'individuo all'interno delle piattaforme, rivelandosi così condizionato dalle regole di accesso e di visibilità imposte da tali “nuovi governatori” dello spazio digitale²².

Pertanto, la consapevolezza dell'insufficienza di tale modello ha condotto, in una seconda fase²³, all'elaborazione di un paradigma di *coregolamentazione*, fondato sulla cooperazione tra poteri pubblici e attori privati. L'obiettivo dichiarato era quello di responsabilizzare le piattaforme senza esautorare la loro capacità tecnica e innovativa, mantenendo un equilibrio tra la flessibilità operativa e la salvaguardia dei valori costituzionali dell'ordinamento europeo.

In tale prospettiva, le istituzioni europee hanno perseguito una logica di

20. C. Pinelli, U. Ruffolo, *I diritti nelle piattaforme*, Giappichelli, 2023, p. 13.

21. Istituito nel 2018, come noto, il Codice di buone pratiche è stato poi notevolmente rafforzato prima nel 2022, con l'obiettivo di essere riconosciuto come Codice di condotta ai sensi della legge sui servizi digitali, e poi il 13 febbraio 2025, quando la Commissione e il Comitato europeo per i servizi digitali hanno approvato l'integrazione del Codice di buone pratiche sulla disinformazione del 2022 come Codice di condotta sulla disinformazione nel quadro della legge sui servizi digitali.

22. K. Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, in *Harvard Law Review*, n. 131, 2018.

23. Sulle fasi di intervento europeo di contrasto alla disinformazione, si veda O. Pollicino, P. Dunn, *Intelligenza artificiale e democrazia*, cit., pp. 127-134.

shared responsibility: lo Stato definisce gli obiettivi di interesse pubblico (trasparenza, sicurezza, tutela dei diritti fondamentali), mentre le piattaforme conservano la libertà di individuare gli strumenti più idonei al loro raggiungimento.

È in questo quadro di progressiva maturazione che si inserisce, nel 2022, l'adozione del Digital Services Act, che ha costituito un vero e proprio punto di svolta nella regolazione dello spazio digitale europeo. Il Dsa rappresenta, infatti, il primo tentativo di istituzionalizzare la cooperazione tra poteri pubblici e privati, attraverso l'introduzione di obblighi giuridici vincolanti per le *very large online platforms* (Vlop) e i *very large online search engines* (Vlose), chiamati a valutare annualmente i rischi sistemici derivanti dai propri servizi, tra i quali rientrano espressamente la disinformazione, la manipolazione elettorale e le interferenze informative.

Pertanto, la logica sottesa a detto Regolamento si dimostra essenzialmente prevenzionistica e orizzontale, poiché non mira alla rimozione selettiva dei contenuti, bensì alla gestione del rischio attraverso strumenti di trasparenza, *risk assessment* e *audit* indipendenti sugli algoritmi. Si tratta, in altri termini, di una regolazione che, più che incidere sul singolo contenuto, interviene sulle condizioni strutturali della comunicazione digitale, riconducendo la responsabilità delle piattaforme a un piano sistemico.

Nel solco di tale approccio, e in perfetta sintonia con la logica della “società del rischio” che permea il diritto europeo contemporaneo²⁴, l'Unione ha ulteriormente arricchito il proprio quadro normativo con l'adozione nel 2024 del Regolamento europeo sull'Intelligenza Artificiale, adottato al fine di mettere a punto «un assetto giuridico adeguato alla complessità e poliedricità della tecnologia» in questione²⁵. Il risultato è una disciplina che pare confermare il modello di regolazione europeo orientato su una logica *rights-based*, ove la tutela dei diritti fondamentali non è considerata un limite alla libertà ma, al contrario, la condizione del suo esercizio pieno e responsabile.

Già nel 2018 la Commissione europea aveva adottato il Piano coordinato sull'intelligenza artificiale, ove aveva tentato di individuare «le coordinate attorno alle quali dovesse ruotare l'intelligenza artificiale»²⁶, evidenziando che soltanto la sussistenza di un «quadro di riferimento chiaro e completo» avrebbe consentito ai cittadini di avere fiducia nello sviluppo tecnologico²⁷.

24. O. Caramaschi, *Il costituzionalismo al cospetto dell'intelligenza artificiale: nuove sfide, quali soluzioni?*, in *Rivista italiana di informatica e diritto*, n. 1/2025, p. 30.

25. M. Bassini, *Intelligenza Artificiale generativa: alcune questioni problematiche*, in *Rivista di diritto dei media*, n. 2/2023, p. 392.

26. D. Martire, *Intelligenza artificiale e Stato costituzionale*, in *Diritto Pubblico*, n. 2/2022, p. 404.

27. *Ibidem*.

In tal senso, l'AI Act rappresenta senz'altro un tentativo degno di plauso, al quale si deve riconoscere il merito di aver messo a punto una normativa di carattere settoriale che considera l'IA come un insieme di tecnologie rischiose. Si tratta, nello specifico, di tecnologie a rischi graduati, classificati secondo il potenziale impatto sui diritti e sulla sicurezza degli individui e della collettività.

Pur non affrontando in modo diretto il tema della disinformazione, il Regolamento lo richiama indirettamente all'art. 50, imponendo ai fornitori di sistemi di IA capaci di generare contenuti sintetici – testi, immagini, audio, video – obblighi di *disclosure* e tracciabilità, volti a evitare che tali prodotti possano essere confusi con informazioni autentiche.

Tuttavia, come si anticipava nel paragrafo che precede, la vera novità concettuale pare risiedere nella qualificazione dei processi democratici come ambiti “ad alto rischio”. L'AI Act, infatti, include espressamente tra i sistemi sottoposti a obblighi rafforzati quelli destinati a influenzare l'esito delle elezioni o delle consultazioni referendarie, nonché a incidere sul comportamento di voto degli elettori.

In tal modo, il legislatore europeo riconosce in modo esplicito che la tutela della democrazia nell'era dell'algoritmo non si esaurisce nel momento formale del voto, ma comprende anche le condizioni cognitive e informative che ne determinano la genuinità.

Si tratta di un passaggio concettuale di grande rilievo, perché anche in assenza di un espresso riferimento alla disinformazione l'AI Act ne presuppone l'incidenza potenziale, collocando la protezione dei processi deliberativi e dell'opinione pubblica tra le finalità stesse della regolazione.

Resta, tuttavia, un nodo irrisolto: il rapporto tra Dsa e AI Act. Se il primo regola la circolazione dei contenuti *online*, il secondo, invece, disciplina i sistemi che li generano o li elaborano. Entrambi perseguono la tutela dei diritti fondamentali nello spazio digitale, ma si muovono lungo assi regolatori differenti: *systemic risk management* nel Dsa, *risk-based approach* tecnologico nell'AI Act. Ci si chiede, dunque, se – e in quale misura – le garanzie poste dal Dsa possano o debbano estendersi anche ai *deployer* e ai *provider* di sistemi di intelligenza artificiale, e se un simile ampliamento possa realmente rafforzare la tutela del diritto all'informazione, senza introdurre forme di controllo preventivo incompatibili con la libertà di espressione²⁸. La risposta non sembra risiedere in un'estensione formale dell'applicabilità, bensì in una integrazione sistematica e funzionale tra i due regimi: un dialogo normativo capace di preservare il nu-

28. Sul punto, si veda M. Bassini, *Speech without a Speaker: Constitutional Coverage for Generative AI Output?*, in *European Constitutional Law Review*, 2025.

cleo duro della libertà di manifestazione del pensiero e, al contempo, di garantire la responsabilità algoritmica degli attori privati che operano nell'ecosistema informativo digitale.

4. *Pars construens*: l'IA al servizio del pluralismo informativo

Le considerazioni sin qui svolte mostrano come la disinformazione, amplificata dall'intermediazione algoritmica e dalla pervasività delle piattaforme digitali, rappresenti oggi una delle principali sfide per la tenuta della democrazia.

Resta tuttavia una domanda di fondo, che travalica il piano strettamente normativo: è possibile che la stessa tecnologia che alimenta la crisi del discorso pubblico diventi anche parte della sua soluzione?

L'intelligenza artificiale, in effetti, si presenta come un fenomeno intrinsecamente ambivalente, rappresentando certamente una fonte di rischio per la diffusione di contenuti disinformativi, ma al contempo anche una potenziale risorsa per contrastarne gli effetti e promuovere la produzione di informazione pluralistica e di qualità²⁹. Se da un lato, infatti, i sistemi generativi e predittivi possono amplificare la circolazione di contenuti falsi, polarizzanti o sintetici, dall'altro le medesime architetture possono essere ripensate in funzione garantista, per promuovere la diversità delle fonti, la trasparenza delle scelte algoritmiche e l'inclusione di prospettive eterogenee nello spazio informativo.

È in questo orizzonte che si ritiene di poter inserire una riflessione su quello che la dottrina ha definito "*diversity by design*", concetto elaborato nell'ambito dell'ingegneria dei sistemi informativi per indicare la possibilità di intervenire, sin dalla fase di progettazione degli algoritmi, prevedendo dei meccanismi che favoriscano l'esposizione degli utenti a contenuti diversificati e che contrastino la formazione di bolle informative³⁰.

Trasposto nel dibattito giuridico e politologico, tale principio può divenire una chiave di lettura per le politiche europee sulla *media diversity*: non più solo come obiettivo di *policy* da perseguire *ex post* mediante i più canonici strumenti regolatori, ma come criterio intrinseco di progettazione degli ambienti digitali. In altre parole, il *diversity by design* rappresenta la traduzione in termi-

29. Sul tema, si rimanda al saggio di G. Contissa, F. Galli, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025, pp. 129-131. Per una puntuale ricostruzione delle numerose modalità di impiego "positivo" dei sistemi di intelligenza artificiale si rimanda a O. Pollicino, P. Dunn, *Intelligenza artificiale e democrazia*, cit., pp. 150-153.

30. N. Helberger, *Diversity by design*, in *Journal of Information Policy*, n. 1/2011, pp. 441-469.

ni “tecnologici” del principio del pluralismo informativo: volto a incorporare nei codici delle piattaforme quei valori di apertura e molteplicità che l’art. 21 Cost. e l’art. 11 Cdfue pongono a presidio della sfera pubblica democratica, sostanzialmente orientando l’algoritmo a promuovere l’esposizione degli utenti a informazioni diversificate, invece di consolidare meccanismi di chiusura cognitiva come le *filter bubbles* e le *echo chambers*. In questi termini, se orientata secondo i riconosciuti principi di trasparenza, *accountability* e “diversità”, l’intelligenza artificiale potrebbe così svestire gli abiti del vettore della crisi per divenire alleata nella garanzia della qualità del discorso pubblico.

Tuttavia, anche tale prospettiva non è esente da criticità. Se l’IA viene istruita privilegiando soltanto fonti considerate “autorevoli” si corre il rischio di un appiattimento informativo che soffoca l’eterogeneità dei punti di vista e riduce il pluralismo a mera omologazione. Se, al contrario, il sistema viene nutrito di dati viziati, incompleti o polarizzati, il rischio è opposto: la diversità non sarebbe più una misura posta a sostegno del pluralismo, bensì una forma sofisticata di relativismo informativo.

In entrambi i casi pare riaffiorare una domanda antica, che nessuna tecnologia riesce a eludere: chi decide che cosa è vero e secondo quali criteri? L’illusione di poter delegare a un algoritmo il discernimento tra il vero e il falso porterebbe infatti a riprodurre, in forma automatizzata, le stesse asimmetrie di potere che il diritto costituzionale si propone di evitare.

Da questa prospettiva, pertanto, la sfida non è quella di integrare al *diversity by design* un modello di “verità programmata”, ma nel costruire un equilibrio dinamico tra apertura e affidabilità: un sistema che preservi la pluralità delle opinioni senza rinunciare all’attendibilità delle informazioni; un sistema in cui la tecnologia non pretenda di determinare la verità, ma si limiti a rendere le condizioni della sua ricerca più accessibili e trasparenti.

In quest’ottica, la tecnica e il diritto non si porrebbero più in contrapposizione, ma come strumenti complementari di una stessa architettura regolatoria: la prima come strumento di attuazione del secondo, in un’ottica di co-evoluzione tra garanzie costituzionali e innovazione tecnologica. Forse, raggiunto un simile equilibrio, l’intelligenza artificiale potrà realmente divenire alleata nella tutela del discorso democratico, contribuendo a preservare un flusso informativo aperto, affidabile e plurale.

DISCLOSURE, EXPLAINABILITY E HUMAN OVERSIGHT: NUOVI DIRITTI FONDAMENTALI NELL'ERA DELL'IA E DEI *PRIVATE GOVERNORS*

*Andrei-Mihai Pop**

SOMMARIO: 1. Introduzione – 2. La genesi dei diritti fondamentali – 3. *Human first* e l'IA – 4. Dai poteri pubblici ai *private governors*: lotta alla disinformazione come banco di prova per i nuovi diritti – 5. Conclusioni.

1. Introduzione

Alla luce dello sviluppo della tecnologia appare opportuno ripensare il catalogo dei diritti fondamentali. Invero, essi dovranno tutelare il singolo dalle ingerenze dei poteri pubblici, ma anche dall'ascesa di un potere algoritmico che proviene dal mondo privato – in particolare dai c.d. *private governors*.

Ogni rivoluzione ha necessitato di una elaborazione e previsione di nuovi diritti e norme, quindi appare coerente operare nello stesso modo anche per l'IA. Un primo elemento da comprendere è quello legato al quesito: cosa è un diritto fondamentale? La ragione di una simile domanda è semplice, ossia per sostenere la necessità di introdurre nuovi diritti appare opportuno comprendere l'oggetto della discussione.

Si segnala che un diritto fondamentale è composto da due termini: *in primis*, da diritto, ossia una situazione giuridica di vantaggio, tutelata dall'ordinamento, rispetto all'attore privato e pubblico; *in secundis*, da un aggettivo, ossia fondamentale, il quale richiama l'idea che di una certa utilità-vantaggio non si può fare a meno, poiché fondativa di un certo ordinamento. Allo scopo di distinguere la generica categoria di diritti da quella di diritti fondamentali occorre effettuare un certo sforzo, poiché, di per sé, questa seconda nozione non ha un contenuto predeterminato¹. Infatti, la fundamentalità dei diritti non è

* Dottorando di ricerca in Filosofia del diritto, Università degli Studi di Milano.

1. La Costituzione utilizza il termine diritto fondamentale una sola volta all'articolo 32, ma ciò non equivale alla inammissibilità di altri diritti.

un attributo che, in sé e per sé, assume un significato all'interno dell'universo giuridico essendo necessario ricercare le risposte all'esterno di esso. Dunque, un diritto verrà considerato come fondamentale in base anche a criteri esterni, quindi a dei giudizi di valore. Invero, osservando la migliore dottrina, si possono individuare due grandi filoni di pensiero nell'interpretazione della nozione di diritto fondamentale.

Una prima tesi formalistica afferma che i diritti sono fondamentali a seconda della fonte che li afferma². In tale visione, essi sarebbero quelli enunciati nelle Costituzioni e nelle varie leggi. Una seconda impostazione, adottando una prospettiva sostanziale, osserva i contenuti e le posizioni giuridiche oggetto di tutela per affermare se un diritto sia dotato dell'attributo della fundamentalità³. A oggi i più, preferendo la seconda tesi, si concentrano maggiormente sul contenuto degli stessi. Infatti, i diritti fondamentali andrebbero a proteggere uno zoccolo duro di libertà nei confronti di ogni manifestazione del potere.

A questo punto risulta opportuno comprendere se risulti necessario un intervento del legislatore, ovvero se esista già, all'interno della nostra Carta dei diritti, uno strumento idoneo all'accoglimento delle nuove istanze della *societas*. La risposta a tale interrogativo potrebbe avere ricadute immense sulla tutela dei diritti dei singoli, poiché accogliendo la prima tesi poc'anzi esposta non si potrebbe che dubitare della capacità del Legislatore di intervenire tempestivamente con una disciplina organica all'altezza delle sfide lanciate dall'IA⁴. Ovvero, allo stato, la via preferibile da percorrere sarebbe la seconda per il tramite dell'articolo 2 della Costituzione.

L'interpretazione di tale norma ha generato immensi scontri tra gli esperti: è una clausola chiusa o una aperta? Secondo l'impostazione chiusa⁵, i diritti fondamentali sarebbero solamente quelli riconosciuti dalla Costituzione, ovvero quelli potenzialmente riconoscibili per il tramite di una interpretazione estensiva. Invece, la prospettiva aperta⁶ osserva nella detta norma una finestra tramite la quale nuovi diritti potrebbero essere affermati. Tale questione non può essere definita come una astratta. Ovverosia, le sfide dell'IA difficilmente

2. Così N. Bobbio, *Presente e avvenire dei diritti dell'uomo*, in *L'età dei diritti*, Einaudi, 2014.

3. Si veda G. Pino, *Il costituzionalismo dei diritti. Struttura e limiti del costituzionalismo contemporaneo*, il Mulino, 2017.

4. L'*AI Act*, e i vari interventi nazionali, indubbiamente introducono nuovi diritti fondamentali, ma non riescono a enuclearli tutti. Inoltre, l'*iter* legislativo richiede delle specifiche tempistiche e lo stesso *AI Act* entrerà pienamente in vigore solamente gradualmente.

5. Si rimanda ad A. Pace, *Problematica delle libertà costituzionali. Parte generale*, Cedam, 2003.

6. Così A. Barbera, *Art. 2*, in *Commentario della Costituzione. Principi fondamentali*, Zanichelli, 1975.

potrebbero consentire una risoluzione dei futuri problemi semplicemente con l'estensione delle vecchie libertà e garanzie.

Come accennato, a seguito di lunghi confronti, la tesi della fattispecie aperta è prevalsa. La scelta di riconoscere la possibilità che l'evoluzione dei costumi e dei bisogni della società possa innovare il catalogo dei diritti fondamentali non deve sorprendere. Invero, uno dei timori di una costituzione scritta è quella di divenire "lettera morta".

2. La genesi dei diritti fondamentali

Alla luce della ricostruzione svolta, è assodato che l'interpretazione dell'articolo 2 della Costituzione come clausola aperta fornisce lo strumento per accogliere nuove istanze di tutela, occorre analizzare la genesi di tali nuovi diritti⁷. L'IA è un potere, anzi sarà "Il Potere" nel futuro prossimo, quindi occorrono nuovi interventi.

Analizzando la genesi di tale categoria, si possono svolgere alcune osservazioni. In genere un diritto segue una certa sequenza: *in primis*, esso nasce da una istanza sociale; *in secundis*, riceve lentamente espliciti segnali di riconoscimento da parte della giurisprudenza; successivamente, essa verrà positivizzata; infine, e a una certa distanza temporale, trova un accoglimento all'interno delle carte costituzionali. Di seguito si concentrerà l'attenzione sul quarto momento di tale processo con lo scopo di individuare quei nuovi diritti fondamentali che potrebbero, alla luce dell'IA, trovare accoglimento all'interno della Costituzione. Purtroppo, tale analisi non verrà svolta con uno sguardo panoramico, bensì calando i diritti all'interno della vita e dell'azione dei *private governors*, in modo da avere un contesto di riferimento. Il rapporto tra i *private governors* e i cittadini dovrà essere modificato e adattato all'impatto che l'IA avrà sia nel settore privato sia in quello pubblico.

Occorreranno nuovi protocolli, procedure e diritti a favore dei soggetti coinvolti, poiché le piattaforme digitali sono la *longa manus* di un potere algoritmico, oltre che spesso l'attore più vicino al territorio e agli utenti⁸. Invero, i *private governors* impattano, con la loro attività, sulle basilari libertà democratiche del nostro ordinamento, in quanto la manipolazione delle informazioni è pericolosa specialmente se il soggetto privato rimane irresponsabile.

7. Si veda A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Rivista di Biodiritto*, n. 1/2019.

8. Innegabile la pervasività delle piattaforme digitali all'interno della quotidianità (lo *smart phone* è l'esempio per eccellenza).

A seguito di queste riflessioni iniziali, si può già intuire che la questione non è “se” l’IA verrà utilizzata dai *private governors*, bensì il “come” e con quali limiti: il cittadino avrà un controllo circa l’andamento del procedimento informativo, oppure diverrà passivo rispetto al calcolo della macchina? I *private governors* saranno un potere responsabile?

Per tale ragione, al fine di tutelare l’esistente, servirà una ricerca e una operazione di individuazione al fine di enucleare dal presente i diritti fondamentali futuri imprescindibili per una *societas* complessa come quella attuale.

Tra i più importanti candidati, pur non esaurendo essi la lista, vi sarebbero: il diritto a una *AI disclosure* (riconoscibilità-conoscibilità), ossia il conoscere la natura artificiale di un determinato sistema con il quale si interagisce; il diritto a una spiegazione dei risultati dell’IA, quindi una *AI explainability*; il diritto a un controllo umano sul funzionamento del sistema di IA, definito come *human oversight*, il quale si connette al principio di non-esclusività dell’IA nella decisione⁹. Tali potenziali nuovi principi giuridici si andrebbero a collocare nel seno del c.d. *European Constitutional Law of Technology*. Nel successivo paragrafo ci si concentrerà, in particolare, sull’*human oversight*, dunque l’intervento e controllo dell’Uomo sull’operato dell’IA.

3. *Human first* e l’IA

L’Uomo fin dalle origini si è prodigato di migliorare le proprie condizioni di vita con l’uso della tecnica. Inizialmente, quest’ultima rimaneva strutturalmente uno strumento asservito agli scopi del proprio creatore. Tuttavia, lo sviluppo della tecnica impone un sempre maggiore grado di autonomia delle invenzioni per incrementare il grado di soddisfazione della *societas*. Ciononostante, doveroso tentare un contenimento di tale “slittamento esistenziale” attraverso lo strumento del diritto, il quale rimane una delle opere più longeve, potenti e umane. È stato già evidenziato che la tecnica è presente nella vita dell’Uomo fin dai suoi albori. Purtroppo, la *téchne* ha una gemella, la quale nacque in un periodo contiguo e, pressoché, in contemporanea¹⁰. Ci si riferisce allo *ius*, il quale è già rinvenibile nelle prime regole di condotta di un dato gruppo di uomini rispetto ai propri membri (su un piano interno) e agli altri gruppi (su di un piano esterno). In altri termini, la tecnica come massima espansione delle ambizioni e potenze, alle volte incontrollate, dell’umanità; al contrario,

9. Così V.G. Sapuppo, *Etica e IA*, in M. Iaselli (a cura di), *AI Act. Principi, regole e applicazioni pratiche del Reg. UE 1689/2024*, Maggioli, 2024.

10. Si consulti N. Irti, E. Severino, *Dialogo su diritto e tecnica*, Laterza, 2001.

lo *ius* come limite e guida verso un determinato fine¹¹. Negli ultimi decenni tale rapporto non è mutato, tuttavia la tecnica ha subito una netta accelerazione, e il diritto non è riuscito sempre ad adattarsi. Invero, sono addirittura state avanzate tesi che vorrebbero la sostituzione di uno *ius* umano con un c.d. *ius* sintetico, in quanto oggettivo e imparziale. In tale ultimo scenario, la stessa esistenza del diritto rischierebbe di essere elisa. Dunque, si può cogliere la necessità di ripudiare la tesi che vorrebbe un diritto assolutamente neutrale¹² allo scopo di favorire la tecnica, poiché una delle sue essenze è proprio quella di controbilanciarla e di consentirne la manifestazione alla condizione di salvaguardare i diritti fondamentali degli individui. In particolare, non occorre un diritto neutrale, bensì un buon diritto¹³.

Detto ciò, quello che potrebbe essere ammesso è l'enucleazione di un insieme di norme coordinate su di un certo territorio per evitare disallineamenti – coerente a tale prospettiva l'approccio dell'UE con l'*AI Act*. Questa tensione tra il dominio umano sulle proprie opere e la ricerca continua di miglioramenti della propria esistenza impone un temperamento che solamente il diritto può svolgere.

In tale elaborato si accoglierà la tesi che impone la preferenza di un Uomo-*dominus* rifiutando una deriva totale che imponga di mutare la visione di una umanità al centro del mondo. Quest'ultima vorrebbe che si passasse da una prospettiva tendenzialmente umano-centrica a una tecno-centrica. In tal guisa, la visione umano-centrica è stata cristallizzata anche dall'*AI Act* che ha voluto imporre una netta preferenza. L'idea di Uomo al comando dell'IA è il fulcro dei diritti tratteggiati nel precedente paragrafo, poiché essi presuppongono che l'umanità sia sempre il principio e la finalità ultima dei sistemi artificiali.

Il cittadino, dunque, dovrà mantenere un certo grado di controllo sui sistemi di IA e sul loro operato se vorrà garantire il rispetto dei diritti costituzionali. Questo criterio di gestione dell'IA ha differenti gradi, ossia la pervasività del controllo rispetto alle operazioni compiute dall'IA aumenterà in virtù del concreto scenario.

Tali categorie sono conosciute con le espressioni: *human in command*, *hu-*

11. Questa considerazione impone, altresì, che il diritto intervenga a monte senza dover inseguire lo sviluppo costante e imprevedibile dell'IA. In altri termini, la sua portata e la forza deve inserirsi in un approccio *by design* all'IA che riesca a giocare d'anticipo evitando, laddove possibile, di giungere al tradizionale approdo violazione-sanzione.

12. Si veda R. Abbott, *The Reasonable Robot. Artificial Intelligence and the Law*, Cambridge University Press, 2020.

13. Così M. Angelone, *Alla ricerca del paradigma ottimale. La governance dell'IA tra legislazione e regolazione*, in *Tecnologie e Diritto*, n. 2/2024.

man in the loop, *human on the loop* e *human out of the loop*¹⁴. L'insieme di tali espressioni contribuiscono alla formazione del principio dell'*human oversight*, ossia il controllo dell'Uomo sull'IA.

Incominciando dalla categoria più debole, ossia l'*Houtl*, si può comprendere come tale categoria, nell'ipotesi in cui l'Uomo non svolga alcuna attività nella decisione dell'IA, non potrà trovare applicazione nel caso in cui siano in gioco diritti delle persone. In un simile scenario, vi è una assenza totale dell'intervento umano, ovvero il *dominus* assoluto è l'IA. Di conseguenza, un simile livello di pervasività dell'IA appare, in via astratta, incompatibile con i trattamenti che riguardino prerogative degli individui.

Procedendo nell'analisi emerge l'*Hotl*. In tale caso, l'IA ha l'iniziativa e la gestione del compito da svolgere, mentre l'Uomo assume il compito di sorvegliare i suoi procedimenti potendo intervenire in una fase *ex post* rispetto all'*output* del sistema. Questo approccio potrà essere perseguito laddove il sistema di IA non comporti pericoli rilevanti per i diritti fondamentali delle persone. Invece, un approccio all'*human oversight* che consenta un più penetrante controllo dell'essere umano è quello dell'*Hitl*. Quest'ultimo implica una gestione-controllo dell'essere umano dell'attività da svolgere in cui l'IA assiste, non decide. Questo modello pare compatibile con la gestione della maggior parte dei dati che riguardino diritti fondamentali e che comportino un rischio medio-alto. Inoltre, il quadro normativo attualmente vigente sostiene la necessità di applicare *Hitl* in tali situazioni in base al *risk-based approach*¹⁵. In ultimo luogo, si trova l'*Hic*. In tale ipotesi, l'IA avrà un ristretto ambito applicativo essendo l'intero procedimento controllato dall'Uomo. In tale caso è proprio l'Uomo a essere il *dominus*.

Tale rigore troverà applicazione laddove le attività da svolgere possono coinvolgere e influenzare diritti fondamentali con un rischio alto-inaccettabile. Dunque, il diritto allo *human oversight* verrà declinato, nella sua ampiezza e profondità, in base allo stato tecnologico e all'interesse oggetto di bilanciamento. Da un lato, più trasparenza e spiegabilità avrà un sistema di IA e mag-

14. Una parte della dottrina ha enucleato una ulteriore categoria, ossia lo *human post the loop* (*Hptl*). Quest'ultima si porrebbe tra lo *Houtl* e lo *Hotl* in termini di grado di controllo umano. Così M. Durante, *Potere computazionale. L'impatto delle Ict su diritto, società, sapere*, Meltemi, 2019.

15. Tale impostazione è particolarmente sostenuta dal legislatore europeo. Invero, già il Gdpr era fondato su ciò. Ovviamente la tutela dei diritti fondamentali e l'approccio basato sul rischio necessitano anch'essi di una continua opera di contemperamento, poiché quest'ultimo è tipicamente uno strumento di stampo privatistico-aziendale orientato più all'utile del mercato che alla tutela delle libertà. In altri termini, nessun diritto può essere considerato assoluto e tiranno (Corte Costituzionale, n. 85/2013), tuttavia ciò non equivale alla possibilità di una sua erosione oltre una minima soglia.

giori saranno le possibilità per seguire un approccio *HotI*; viceversa, laddove il problema della *black box* non venga sufficientemente contenuto, risulterà opportuno incrementare il controllo umano. Dall'altro lato, assumendo una differente prospettiva, all'aumento di criticità del diritto oggetto di attenzione da parte del sistema di IA, vi sarà la necessità di utilizzare sistemi trasparenti e spiegabili.

Questa tensione tra stato tecnologico e diritti dovrà essere oggetto di un contemperamento continuo da parte del legislatore, dei giudici e degli operatori del diritto. In ogni caso, e prescindere dalla etichetta, tali espressioni possono essere tradotte con il diritto-dovere a una sorveglianza umana sull'IA. Invero, si impone la regola secondo la quale in caso di applicazione di sistemi di IA, in settori giudicati come critici, occorrerà la supervisione dell'Uomo – c.d. *collaborative intelligence*¹⁶. Questo criterio ha avuto già nel *Gdpr* una sua iniziale, benché parziale, enunciazione all'articolo 22. A una prima lettura, tale previsione risulta essere perentoria pur potendosi applicare solamente nei casi in cui l'IA utilizzi dati personali. Purtuttavia, tale enunciazione subisce importanti deroghe ai successivi commi rendendo, *de facto*, il divieto di trattamenti automatizzati non la regola, bensì l'eccezione. Ovvero, ai sensi del comma 2, il divieto non si applicherà laddove vi sia un contratto, una legge o il consenso del soggetto destinatario di tale trattamento. In particolare, la deroga relativa all'accordo dell'interessato acuisce il rischio del *blind consent*, ossia la tendenza di prestare il proprio consenso ai fornitori dei servizi *internet* senza effettuare una sostanziale lettura delle condizioni del trattamento¹⁷. Quindi, la previsione del *Gdpr* è particolarmente lacunosa.

Alla luce di quanto osservato, appare fondata la visione secondo la quale non si potesse ritenere esistente nel *Gdpr* un diritto sufficientemente solido al controllo umano nei trattamenti automatizzati da parte dell'IA.

Il regolamento europeo sull'IA (n. 1689/2024) tenta di offrire una soluzione a tali aspetti problematici. In particolare, l'articolo 14 dell'*AI Act* – rubricato *Human oversight*. In tale sede, al comma 1, si prevede che al fine di garantire la supervisione dell'Uomo sull'IA rientrante nella categoria ad alto rischio, quest'ultima dovrà essere ideata, progettata e realizzata in un modo compatibile con le capacità di comprensione e di gestione dell'essere umano (*human friendly by design*).

16. Si veda S. Quintarelli *et al.*, *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale: principi diritti e raccomandazioni*, in *Rivista di Biodiritto*, n. 3/2019.

17. Il consenso, quanto meno in certi casi, è divenuto una finzione, ovvero un consenso consapevolmente disinformato. Così M. Barberis, *Separazione dei poteri e giustizia digitale*, Mimesis, 2023.

La seconda parte di tale articolo fissa l'obiettivo di tale *human oversight*, ovvero sia quello di prevenire i rischi per i diritti fondamentali oggetto del trattamento da parte dell'IA. Precisa la norma che tale valutazione, effettuata *ex ante*, terrà conto dell'uso "naturale" di tale sistema, o al più di quello che si poteva ragionevolmente prevedere in base alle caratteristiche del caso concreto – *reasonably foreseeable misuse*. Tale precisazione è foriera di conseguenze, poiché ritorna il tema della trasparenza e dell'esplicabilità, ossia per riuscire ad anticipare un uso ragionevole del sistema, occorre conoscere e comprendere il funzionamento dello stesso. In assenza di un sufficiente livello di tali caratteristiche, tale norma non si potrà rispettare. Già queste prime considerazioni sull'articolo 14 rivelano l'intima connessione tra la *transparency*, la *explainability* e lo *human oversight*: essi sono tutti delle *conditio sine qua non*. Questa triade deve essere considerata unitaria, ovvero un unico "meta-diritto"¹⁸ a una *governance* algoritmica umano-centrica nel contesto dell'IA.

Proseguendo, il comma 3 affronta il tema dell'intensità del controllo umano sul funzionamento del sistema di IA cercando di individuare un concreto contemperamento tra efficienza e diritti. La profondità della sorveglianza umana dipenderà dai diritti in discussione, dai rischi nei loro confronti, dal livello di autonomia del sistema di IA e in ultimo dal contesto di riferimento e di funzionamento del sistema. Quindi, l'effettiva applicazione di tale regola dovrà tenere in considerazione quattro caratteristiche. Da ultimo, opportuno osservare il comma 4. Quest'ultimo elenca le effettive misure che dovrebbero essere offerte all'utilizzatore finale per concretizzare il diritto al controllo umano sull'IA. Queste sono principalmente quelle legate alla possibilità di discostarsi dall'*output* prodotto dal sistema nella consapevolezza che lo stesso può essere errato o comunque impreciso. Purtroppo, non solo l'*AI Act* affronta tali principi, bensì anche il *Digital Service Act* (n. 2065/2022). In particolare, l'articolo 17 impone di offrire una motivazione umana (c.d. *human oversight* rinvenibile anche negli articoli 20 e 21) laddove un determinato contenuto venga moderato nel seno di una piattaforma privata digitale. Questa disposizione si lega strutturalmente, rappresentandone un'altra manifestazione, ai principi di *AI Disclosure* ed *Explainability*. L'esigenza sorge al fine di evitare l'inquinamento del dibattito pubblico e la diffusione di disinformazione (articolo 34 del *Dsa*).

Tale consapevolezza è fondamentale e sarà uno dei principali punti problematici nell'uso dei sistemi di IA ad alto rischio, poiché il rischio di appiattimento della decisione umana su quanto suggerito dal sistema è concreto e attuale.

18. La sua natura implica che l'assenza di un solo componente annulla gli altri, e ciò lo eleva da un mero pacchetto di diritti a un "meta-diritto". Per esempio, il diritto a un ricorso (*oversight*) è inutile se non si sa perché una decisione è stata presa (*explainability*).

In dottrina si è parlato di un *effet moutonnier*¹⁹, ossia l'incapacità dell'Uomo di analizzare in maniera critica la decisione dell'IA in forza del mito dell'infallibilità dell'IA, il quale cattura l'Uomo non consentendogli più una libera scelta. In altre parole, l'opzione individuata dal sistema diverrebbe la *default-option*, la quale godrebbe di un plusvalore di significato, poiché l'individuo considererebbe l'IA superiore. Invero, salvo palesi errori, per discostarsene servirebbe un intenso sforzo²⁰. L'*output* dell'IA assume un simile valore poiché riesce ad alleggerire il peso della motivazione²¹.

Questa tematica sarà una fondamentale perché solamente un vaglio umano della decisione renderà quest'ultima accettabile dalla prospettiva dei consociati. Osservando l'impianto dell'articolo 14 dell'*AI Act* si comprende l'ampiezza potenziale della sua applicazione, poiché i sistemi ad alto rischio occupano una percentuale importante dei sistemi di IA. Il rischio di una interpretazione eccessivamente rigida potrebbe comportare alcuni problemi e il richiamo del comma 3 alle "circostanze" di utilizzo dell'IA potrebbe non essere sufficiente per un equo temperamento tra le esigenze di tutela dei diritti e le necessità di sviluppo tecnologico dell'UE.

I nuovi diritti fondamentali che dovranno essere riconosciuti all'interno del sistema costituzionale per consentire un governo umano-centrico dell'IA sono rinvenibili sia nella legislazione europea sia in quella nazionale. Tali diritti fanno parte di una costellazione di principi intimamente legati tra di loro, ovvero servirà una continua interazione tra gli stessi per consentire all'Europa di ergersi a guida capace di concentrare i valori occidentali volti a un equilibrio tra l'ambizione della *téchne* e la tutela dell'Uomo. Tale temperamento è frutto dell'intervento del diritto nella sua dimensione astratta, ma si evidenzia la necessità che le norme si trasformino da *law in the books* in *law in action*²². In caso contrario, non si potrà che ammettere l'abdicazione assoluta dello *ius* dal suo ruolo di co-protagonista – affianco alla tecnica – e la sua riduzione a mera comparsa nella storia dell'umanità. Un simile scenario dovrà essere assolutamente evitato per riuscire a mantenere l'essere umano *in command*.

19. Così A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Puf, 2018.

20. Tuttavia, viste le competenze richieste, questo scenario appare essere di difficile attuazione.

21. Si veda A. Simoncini, *Amministrazione digitale algoritmica. Il quadro costituzionale, in Il diritto dell'amministrazione pubblica digitale*, Giappichelli, 2020.

22. Così R. Pound, *Law in Books and Law in Action*, in *American Law Review*, vol. XXXIV, 1910. Risulta necessario sia il riconoscimento di tali diritti fondamentali sia concrete normative mirate, ad esempio, all'alfabetizzazione digitale o a interventi di *design* per contrastare *bias* cognitivi.

4. Dai poteri pubblici ai *private governors*: lotta alla disinformazione come banco di prova per i nuovi diritti

L'analisi fin qui svolta ha consentito di definire i contorni giuridici dei nuovi diritti e di vederne un potente riconoscimento nell'*AI Act* e nel *Dsa*. Tuttavia, una delle sfide più pressanti per i sistemi democratici proviene dall'ecosistema informativo dominato da attori privati. La diffusione di *fake news* non è un fenomeno astratto, ma il risultato di precise architetture tecnologiche gestite da quelli che sono stati efficacemente definiti *private governors*. Infatti, le piattaforme *online* che moderano, curano e amplificano i contenuti esercitando, *de facto*, un potere normativo sulla sfera pubblica. Questa condizione va affrontata con urgenza, poiché trascurare tali aspetti rischia di indebolire l'impianto del sistema giuridico.

In questo contesto, l'*AI disclosure*, l'*explainability* e l'*human oversight* diventano strumenti essenziali per una cittadinanza attiva e democratica. In primo luogo, la trasparenza impone che l'utente venga informato dell'utilizzo di determinati sistemi di IA su tali piattaforme per il trattamento dei suoi dati e preferenze. Già tale circostanza è essenziale per segnalare, fin dal principio all'utente, il contesto e le potenzialità critiche. Successivamente, il diritto all'*explainability* si trasforma nel diritto dell'utente-elettore di comprendere perché un contenuto informativo è stato rimosso o la sua visibilità ridotta (*shadow banning*), oppure, e al contrario, perché una notizia falsa viene promossa algoritmicamente raggiungendo milioni di persone. Tale risultato sarà raggiunto laddove l'utente venisse informato, in un linguaggio comprensibile, delle motivazioni che hanno portato il sistema di IA a effettuare certe scelte. Infine, il diritto allo *human oversight* diventa il presidio contro la censura automatizzata garantendo un ricorso umano effettivo contro le decisioni che, nel tentativo di arginare la disinformazione, rischiano di sopprimere il dissenso legittimo o l'informazione veritiera, benché sgradita²³. Lo stesso *effet moutonnier* si manifesta quando milioni di cittadini delegano acriticamente il proprio giudizio agli algoritmi dei *social media* diventando così manipolabili. Questo risultato minerebbe una delle precondizioni fondamentali in una democrazia, per cui il combinato disposto del *Gdpr*, dell'*AI Act* e del *Dsa* si ergono a guida a un utilizzo responsabile dell'IA. Tuttavia, occorre un proficuo impegno per concretizzare tali garanzie, poiché altrimenti il rischio dell'*effet moutonnier* non vorrebbe arginato rimanendo le stesse una mera finzione giuridica.

Riconoscere questi principi come diritti fondamentali significa, pertanto, affrontare il problema della disinformazione alla radice. Ciò come una que-

23. Tale diritto è un presupposto indispensabile per un consenso libero dei cittadini.

stione di equilibrio di potere dotando il cittadino degli anticorpi cognitivi e giuridici necessari per difendere la propria autonomia di giudizio e partecipare consapevolmente al dibattito democratico.

5. Conclusioni

Nel presente elaborato si è cercato di evidenziare l'origine e il significato dei diritti fondamentali e, conseguentemente, di comprenderne il valore e la necessaria presenza. In questa *societas* digitalizzata, il diritto alla conoscibilità, spiegabilità e al controllo umano innanzi all'IA non sono una questione marginale, bensì centrale al fine di garantire le libertà degli individui sia nella loro sfera privata sia in quella pubblica. I sistemi di IA potrebbero essere sia lo scudo sia la spada in relazione alle *fake news*: essi possono aiutare nella lotta alle stesse, ovvero amplificarle. I principi analizzati sono il fondamento di tale discussione, ossia solo se correttamente applicati si potrà giungere a dei risultati positivi. Tale ricostruzione basata sui diritti integra e rafforza l'approccio basato sul rischio del *Dsa* e dell'*AI Act*, offrendo una tutela di rango costituzionale. Tuttavia, l'approccio basato sul rischio di derivazione privatistica è uno consequenzialista, ossia valuta i diritti in termini di "impatti" da mitigare. Invece, l'approccio basato sui diritti fondamentali è uno deontologico: li considera come limiti invalicabili a prescindere dal calcolo dei rischi. La sfida è, come detto, un temperamento tra tali due prospettive con la consapevolezza dell'inesistenza di valori assoluti.

L'ordinamento deve vegliare sul loro rispetto con lo scopo di evitare abusi. A tal fine, tali diritti sono stati positivizzati in appositi testi legislativi, ma gli stessi, ancora più a monte, si ritengono presenti nel quadro delle libertà già sancite nel testo Costituzionale italiano. Questo alla luce di una lettura innovativa dell'articolo 2, il quale viene posto come fondamento per l'elaborazione di nuovi diritti fondamentali legati alla *governance* algoritmica. Di conseguenza, con una simile prospettiva nemmeno laddove determinati testi legislativi fossero modificati-abrogati si potrebbero negare gli stessi, in quanto oramai immanenti e definitivamente affermati.

Sezione III

COMUNICAZIONE ELETTORALE E POLITICA

COMUNICAZIONE ISTITUZIONALE E INTELLIGENZA ARTIFICIALE. APPUNTI

*Massimo Cavino**

SOMMARIO: 1. Definizione normativa di comunicazione istituzionale – 2. Definizione normativa di intelligenza artificiale – 3. Intelligenza artificiale e classificazione della tipologia di rischio per le attività di comunicazione istituzionale – 4. Conclusioni.

1. Definizione normativa di comunicazione istituzionale

La crescente diffusione del ricorso a sistemi di intelligenza artificiale nei più diversi ambiti delle relazioni sociali, e soprattutto la rapidità della evoluzione tecnologica che la caratterizza, pongono al giurista una questione di portata generale: quella della praticabilità delle sue categorie e degli istituti con cui è solito ragionare. Tema senz'altro assai impegnativo che ci pare possa essere affrontato seguendo una impostazione problematica; vale a dire, andando per tentativi: riflettendo su temi circoscritti, perimetrati da concetti usuali, al fine di verificarne la tenuta, la resistenza alle nuove sollecitazioni. In questa prospettiva cercheremo di giungere a qualche conclusione riflettendo sull'impatto che i sistemi di intelligenza artificiale esercitano sulla comunicazione istituzionale.

Per farlo dobbiamo procedere a una, pur rapida, premessa definitoria: cosa intendiamo per comunicazione istituzionale? Cosa per intelligenza artificiale? Sulla base di quali fonti normative?

Per quel che concerne la comunicazione istituzionale il riferimento normativo generale è dato dalla legge 7 giugno 2000, n. 150, recante "Disciplina delle attività di informazione e di comunicazione delle pubbliche amministrazioni". La relazione presentata alla Camera dei deputati il 29 dicembre 1997 sottolinea la necessità di rafforzare il principio di trasparenza e pubblicità dell'azione

* Professore ordinario di Diritto Costituzionale e Pubblico presso l'Università del Piemonte Orientale.

amministrativa, delineato in origine dalla legge n. 241 del 1990, coniugandolo con il diritto “universale” all’informazione. La prospettiva è quella della partecipazione consapevole dei cittadini all’azione amministrativa e del controllo democratico diffuso su di essa. Per comunicazione istituzionale non si deve quindi intendere ogni forma di trasmissione di informazioni da parte di una pubblica amministrazione (si pensi alla comunicazione di avvio del procedimento a vantaggio dei destinatari del provvedimento finale), ma solo quella orientata a tali obiettivi di interesse generale. Pare quindi di poter accogliere la definizione che la qualifica:

come un sistema di comportamenti comunicativi finalizzati al buon funzionamento dell’istituzione al suo interno (ovvero in quanto sistema di interazioni dotato di scopi di livello e complessità diversi), e verso l’esterno (ovvero quale sistema di interazioni tra le unità organizzative interne all’istituzione e il contesto esterno dei portatori di interesse che normalmente interagiscono con detta istituzione)¹.

Nel genere della comunicazione istituzionale si possono così distinguere le specie della comunicazione politica, finalizzata alla illustrazione degli indirizzi perseguiti dall’azione amministrativa²; della comunicazione sociale, orientata

1. E. Zuanelli, *La comunicazione istituzionale: profili teorici, sistemici e normativi*, in *Amministrare*, 1997, n. 3, p. 376. Alcuni autori fanno riferimento al concetto di “comunicazione pubblica”; si tratta di distinzioni soltanto nominalistiche. A titolo di esempio cfr. la definizione di comunicazione pubblica di S. Miconi, *Comunicazione e pubblicità istituzionale: classificazioni e regolamentazione*, in *Il diritto dell’informazione e dell’informatica*, 2016, n. 6, p. 901: «La comunicazione pubblica è lo strumento che consente di esternare e rendere conoscibile ai terzi l’attività delle pubbliche amministrazioni, contribuendo nel contempo al rafforzamento della loro immagine e all’aumento del consenso da parte dei soggetti amministrati. Sulla sua natura vi sono una concezione soggettiva, secondo la quale sarebbe tale ogni forma di comunicazione che semplicemente provenga da una pubblica amministrazione, e una concezione oggettiva, che guarda invece all’oggetto della comunicazione, qualificandola come pubblica qualora assuma interesse e portata generale, ovvero sia idonea a coinvolgere l’intera collettività: in questo senso, sarebbe comunicazione pubblica l’insieme delle informazioni e delle notizie che soggetti, sia pubblici che privati (imprese, enti no profit), forniscono a persone fisiche (cittadini), giuridiche o istituzioni, al fine di favorire la conoscenza, l’informazione e la partecipazione».

2. Da questa deve essere distinta la pubblicità politica quale definita da ultimo dal regolamento (UE) 2024/900 del Parlamento europeo e del Consiglio del 13 marzo 2024, relativo alla trasparenza e al targeting della pubblicità politica. Esso all’articolo 3, paragrafo 2, definisce come pubblicità politica: «la preparazione, collocazione, promozione, pubblicazione, consegna o diffusione, con qualsiasi mezzo, di un messaggio fornito normalmente dietro retribuzione o tramite attività interne o nell’ambito di una campagna di pubblicità politica: a) di, a favore o per conto di un attore politico, salvo se di natura meramente privata o meramente commerciale; oppure b) che possa e sia inteso a influenzare l’esito di un’elezione o referendum, un comportamento di voto o un processo legislativo o regolamentare, a livello dell’Unione, nazionale, regionale o locale». E precisa che non rientrano in questa definizione: «i) ii) i messaggi provenienti

alla promozione della conoscenza di temi rilevanti per la coesione sociale; della comunicazione istruttiva, funzionale alla diffusione della conoscenza delle modalità di funzionamento dell'azione amministrativa e delle forme di partecipazione. Questa classificazione non è esplicitamente accolta dal legislatore che stabilisce genericamente (articolo 1, comma 5, della legge n. 150 del 2000) che è considerata attività di comunicazione istituzionale quella

finalizzata a illustrare e promuovere la conoscenza: a) di disposizioni normative al fine di facilitarne l'applicazione, ovvero di attività, strutture e compiti delle istituzioni, e modalità di funzionamento ed erogazione di servizi di pubblica utilità; b) di temi di carattere sociale o di interesse pubblico; c) dell'utilizzazione dei servizi offerti al pubblico e migliorarne la qualità; d) dell'immagine dell'Italia all'estero, anche dando adeguata visibilità a eventi di importanza nazionale e internazionale.

Si faccia molta attenzione: i diversi oggetti elencati dalla legge non corrispondono in rapporto biunivoco ai diversi tipi di comunicazione istituzionale; ogni specie di comunicazione istituzionale può riguardare ciascuno degli ambiti elencati essendo finalizzata a promuoverne la conoscenza in una diversa prospettiva. Così l'entrata in vigore di una disposizione normativa può essere oggetto di comunicazione politica, illustrandone la coerenza con l'indirizzo politico generale; di comunicazione sociale, sottolineando gli effetti attesi rispetto a specifiche esigenze della collettività; di comunicazione istruttiva, fornendo informazioni utili a una corretta applicazione. Non solo. Si deve anche considerare che, concretamente, una efficace comunicazione istituzionale possa essere orientata a perseguire più finalità, oppure che per raggiungerne una, debba fornire elementi funzionali ad altre, così che un singolo atto comunicativo può avere natura politica, sociale e istruttiva. Alla luce di queste considerazioni si comprende come la scelta del legislatore di non procedere alla defi-

da fonti ufficiali degli Stati membri o dell'Unione strettamente riguardanti l'organizzazione e le modalità di partecipazione a elezioni o referendum, compreso l'annuncio delle candidature o dei quesiti referendari, oppure la promozione della partecipazione a elezioni o referendum; la comunicazione pubblica finalizzata a fornire informazioni ufficiali al pubblico da parte, a favore o per conto di un'autorità pubblica di uno Stato membro o dell'Unione, incluso da parte, a favore o per conto di membri del governo di uno Stato membro, a condizione che non possa e che non sia intesa a influenzare l'esito di un'elezione o di un referendum, un comportamento di voto o un processo legislativo o regolamentare; e iii) la presentazione di candidati in determinati spazi pubblici o nei mezzi di comunicazione, esplicitamente prevista dalla legge, effettuata a titolo gratuito e in condizioni parità di trattamento dei candidati». Chiaro appare che, in concreto, la distinzione tra la pubblicità politica e la comunicazione istituzionale politica è assai ardua poiché assai arduo è distinguere la finalità di condizionare l'orientamento elettorale dei cittadini, che connota la prima, dalla finalità di creare fiducia (e quindi consenso) nell'azione amministrativa che caratterizza la seconda.

nizione delle differenti forme di comunicazione istituzionale fosse obbligata: posto che l'azione amministrativa, anche quella di comunicazione istituzionale, deve essere supportata dalla organizzazione delle strutture, la definizione delle differenti forma di comunicazione avrebbe comportato la necessità di individuare le strutture competenti a gestirle; impresa impossibile in ragione della commistione e sovrapposizione di finalità rispetto ai singoli oggetti di cui abbiamo appena detto. Si è andati pertanto in una direzione diversa. Stabiliti gli ambiti oggetto di comunicazione istituzionale il legislatore ha proceduto a una organizzazione amministrativa³ che distingue le competenze in relazione ai destinatari, disegnando la struttura della comunicazione istituzionale su tre distinte tipologie di ufficio: il portavoce e l'ufficio stampa quali interlocutori dei media professionali; l'ufficio per le relazioni con il pubblico (Urp) come interfaccia tra amministrazione e cittadini⁴.

2. Definizione normativa di intelligenza artificiale

Per una definizione normativa del concetto di intelligenza artificiale dobbiamo invece fare riferimento al regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce le regole armonizzate sul suo utilizzo. Esso all'articolo 3, n. 1, dispone che per sistema di IA si deve intendere:

un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali.

Si tratta però di una definizione non corretta. Se la accettassimo i sistemi di intelligenza artificiale non sarebbero strutturalmente diversi dai sistemi costruiti sulla base delle reazioni di feedback che hanno rappresentato la base dell'informatica a partire dalle riflessioni di Norbert Wiener sulla ciberneti-

3. Cfr. S. Rolando, *Profili organizzativi della comunicazione istituzionale. Bilancio della attuazione della legge 150 e nuove tendenze*, in *Micro & Macro Marketing*, 2005, n. 3, pp. 325 ss.; L. Solito, *Comunicazione istituzionale e società digitale. Un nuovo ruolo per i comunicatori pubblici*, in *Comunicazione politica*, 2018, n. 3, pp. 394 ss.

4. L'Urp era stato istituito dall'art. 12 del decreto legislativo 3 febbraio 1993, n. 29 (ora art.11, del decreto legislativo 30 marzo 2001, n. 165) in attuazione dei principi di partecipazione di cui alle leggi 7 agosto 1990, n. 241, e 8 giugno 1990, n. 142. In argomento A. Rovinetti, *Dall'Urp allo sportello unico della pubblica amministrazione*, *Amministrare*, 1997, n. 3, pp. 393 ss.

ca⁵. In realtà, lo stesso regolamento si corregge, al n. 33 dello stesso articolo 3, ove precisa che i dati di input sono «i dati forniti a un sistema di IA o direttamente acquisiti dallo stesso, in base ai quali il sistema produce un output». La novità dei sistemi di intelligenza artificiale risiede tutta nella possibilità della macchina di darsi autonomamente input ulteriori rispetto a quelli ricevuti, in ragione della sua capacità di calcolo e della disponibilità di informazioni cui la rete le permette di accedere⁶. Questa particolare autonomia della macchina è considerata con grande attenzione dal legislatore europeo che propone *l'uomo come fine* quale canone ermeneutico generale, là dove precisa che intende «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione» (articolo 1, n. 1 del regolamento). La tutela della centralità dell'uomo come fine porta il legislatore comunitario a ispirarsi al principio di precauzione⁷ e a organizzare la disciplina dei sistemi di intelligenza artificiale intorno al concetto di rischio⁸, inteso come «la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso».

In forza di questa impostazione strutturale il regolamento dispone intorno alle pratiche di IA classificandole in ordine decrescente di rischio.

Vieta espressamente quelle che comportano un rischio inaccettabile: poiché utilizzano tecniche subliminali capaci di incidere sulla capacità delle persone di assumere una decisione consapevole; sfruttano la vulnerabilità delle persone per distorcerne materialmente il comportamento; classificano le persone fisiche in base al loro comportamento sociale o alle loro caratteristiche individuali e biometriche a fini discriminatori.

Quindi dispone in ordine alle pratiche ad alto rischio (articolo 6, paragra-

5. N. Wiener, *Introduzione alla cibernetica. L'uso umano degli esseri umani* (1950), Bollati Boringhieri, 2024; Id., *La cibernetica. Controllo e comunicazione nell'animale e nella macchina* (1965), il Saggiatore, 1982. In argomento cfr. M.G. Losano, *Informatica per le scienze sociali. Corso di informatica giuridica*, Einaudi, 1985, pp. 22 ss.

6. Cfr. W. Gasparri, F. Tesi, *Intelligenza artificiale e "AI Act". Il perimetro della difficile sussunzione del soggetto nella macchina calcolante*, in *Diritto pubblico*, 2025, n. 2, pp. 514 ss.

7. Cfr. W. Gasparri, F. Tesi, *Intelligenza artificiale e "AI Act". Il perimetro della difficile sussunzione del soggetto nella macchina calcolante*, cit., pp. 517 ss.

8. Cfr. A. Stiano, *La corsa alla regolamentazione dell'intelligenza artificiale: una riflessione in chiave comparata*, in *Diritto pubblico comparato ed europeo*, 2025, n. 1, pp. 100 ss.; S. Foà, *Intelligenza artificiale e cultura della trasparenza amministrativa. Dalle scatole nere alla casa di vetro*, in *Diritto amministrativo*, 2023, n. 3, pp. 519 ss.

fo 2, e allegato 3) relative a sistemi IA utilizzati nei settori: della biometria (nella misura in cui il diritto nazionale o dell'Unione ne permetta l'uso); delle infrastrutture critiche, vale a dire dei sistemi destinati a essere utilizzati come componenti di sicurezza nella gestione e nel funzionamento delle infrastrutture digitali critiche, del traffico stradale o nella fornitura di acqua, gas, riscaldamento o elettricità; dell'istruzione e della formazione professionale; dell'occupazione, della gestione dei lavoratori e dell'accesso al lavoro autonomo; dell'accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi; delle attività di contrasto (nella misura in cui il diritto nazionale o dell'Unione ne consenta l'uso); della migrazione, del diritto all'asilo e della gestione delle frontiere (nella misura in cui il diritto nazionale o dell'Unione ne consenta l'uso); dell'amministrazione della giustizia e dei processi democratici.

Infine precisa (articolo 6, paragrafo 3) che anche quando trovino applicazione negli ambiti poco sopra elencati, non devono essere considerati ad alto rischio i sistemi di intelligenza artificiale che siano destinati: a eseguire un compito procedurale limitato; a migliorare il risultato di un'attività umana precedentemente completata; a rilevare schemi decisionali o deviazioni da schemi decisionali precedenti e non è finalizzato a sostituire o influenzare la valutazione umana precedentemente completata senza un'adeguata revisione umana; a eseguire un compito preparatorio. Per comprendere a pieno i criteri di esclusione dalle attività ad alto rischio è particolarmente utile la lettura del *considerando* n. 53 del regolamento che introduce come principio generale quello in forza del quale non comportano un rischio significativo i sistemi di IA che non influenzano materialmente il processo decisionale. Nello stesso *considerando* si precisa che:

Ai fini del presente regolamento, un sistema di IA che non influenza materialmente l'esito del processo decisionale dovrebbe essere inteso come un sistema di IA che non ha un impatto sulla sostanza, e quindi sull'esito, del processo decisionale, sia esso umano o automatizzato. Un sistema di IA che non influenza materialmente l'esito del processo decisionale potrebbe includere situazioni in cui sono soddisfatte una o più delle seguenti condizioni. La prima di tali condizioni dovrebbe essere che il sistema di IA sia destinato a svolgere un compito procedurale ristretto, come un sistema di IA che trasforma dati non strutturati in dati strutturati, un sistema di IA che classifica i documenti in entrata per categorie o un sistema di IA utilizzato per rilevare duplicati tra un gran numero di applicazioni. Tali compiti sono di natura così ristretta e limitata da comportare solo rischi limitati che non aumentano con l'uso di un sistema di IA in un contesto elencato come uso ad alto rischio in un allegato del presente regolamento. La seconda condizione dovrebbe essere che il compito svolto dal sistema di IA sia inteso a migliorare il risultato di un'attività umana precedentemente completata che può essere pertinente ai fini degli usi ad alto rischio elencati nell'allegato del

presente regolamento. Tenuto conto di tali caratteristiche, il sistema di IA fornisce solo un livello aggiuntivo a un'attività umana con conseguente riduzione del rischio. Tale condizione si applicherebbe, ad esempio, ai sistemi di IA destinati a migliorare il linguaggio utilizzato in documenti redatti in precedenza, ad esempio in relazione al tono professionale, allo stile accademico del linguaggio o allineando il testo a una determinata comunicazione di marchio. La terza condizione dovrebbe essere che il sistema di IA sia inteso a individuare modelli decisionali o deviazioni da modelli decisionali precedenti. Il rischio sarebbe ridotto in quanto l'uso del sistema di IA segue una valutazione umana precedentemente completata che non è destinato a sostituire o influenzare, senza un'adeguata revisione umana. Tali sistemi di IA comprendono, ad esempio, quelli che, dato un determinato modello di valutazione di un insegnante, possono essere utilizzati per verificare *ex post* se l'insegnante possa essersi discostato dal modello di valutazione in modo da segnalare potenziali incongruenze o anomalie. La quarta condizione dovrebbe essere che il sistema di IA sia destinato a svolgere un compito che è solo preparatorio rispetto a una valutazione pertinente ai fini dei sistemi di IA elencati in un allegato del presente regolamento, e pertanto la probabilità che l'output del sistema presenti un rischio per la valutazione posteriore è molto ridotto. Tale condizione riguarda, in particolare, soluzioni intelligenti per la gestione dei fascicoli, che comprendono varie funzioni quali l'indicizzazione, la ricerca, l'elaborazione testuale e vocale o il collegamento dei dati ad altre fonti di dati, o i sistemi di IA utilizzati per la traduzione di documenti iniziali.

3. Intelligenza artificiale e classificazione della tipologia di rischio per le attività di comunicazione istituzionale

Alla luce di queste premesse normative si potrebbe tentare una classificazione per tipologia di rischio delle azioni degli organi titolari delle mansioni di comunicazione istituzionale.

Una prima considerazione riguarda il rapporto tra comunicazione e processo decisionale: è possibile distinguere nettamente il secondo dalla prima, e quindi stabilire, in via generale, se e quanto l'una possa condizionare l'altro? Pare di non poter rispondere che negativamente: la complessità delle operazioni che, come abbiamo detto, caratterizza la comunicazione istituzionale impedisce di accogliere soluzioni generali e impone di considerare le singole operazioni affidate al portavoce, all'ufficio stampa e all'ufficio per le relazioni con il pubblico⁹.

9. Si vedano le osservazioni di F. Faccioli, *Comunicazione pubblica e media digitali: la prospettiva del public engagement*, in *Problemi dell'informazione*, 2016, n. 1, pp. 20 e s. che sottolinea come la diffusione dei media digitali porti tutti gli apparati amministrativi a condividere le esigenze della comunicazione istituzionale, non solo gli uffici a essa normativamente preposti. Pare però di poter affermare che, fino a ora, alla scelta organizzativa operata dal legislatore del 2000, non siano state proposte alternative credibili.

La comunicazione affidata al portavoce è sicuramente quella più vicina alla formazione del processo decisionale posto che il portavoce, sul piano organizzativo, è soggetto di stretta collaborazione in staff all'organo titolare dell'indirizzo politico. In ragione del suo compito istituzionale, che consiste nell'intrattenere "rapporti diplomatici"¹⁰ con i giornalisti, in molte occasioni egli è chiamato ad assolvere a una funzione di consulente intorno alle ricadute che la comunicazione di una determinata presa di posizione potrebbero generare sui media; benché non si possa immaginare che la sua azione giunga alle soglie di un'autentica co-decisione, certamente essa può incidere, anche pesantemente, sul processo decisionale.

Assai vicina al processo decisionale è anche l'attività dell'ufficio stampa. In linea di principio essa si distingue da quella del portavoce in quanto non orientata a mantenere un rapporto dinamico con i media dall'esterno, ma piuttosto dall'interno. Mentre il portavoce fornisce elementi per orientare i media alla diffusione di notizie e informazioni, l'ufficio stampa diffonde direttamente notizie e informazioni. In linea di principio il portavoce può contattare i giornalisti, come da questi essere contattato, e per questo deve intrattenere un rapporto costante col vertice politico; mentre l'ufficio stampa contatta i giornalisti inviando loro dei comunicati e per questo si deve ritenere che, in linea generale, si limiti a ricevere gli input che il vertice politico darà occasionalmente; mentre l'azione del portavoce può collocarsi anche a monte del processo decisionale, quella dell'ufficio stampa si colloca sempre a valle. In tal senso è di particolare interesse la decisione della Corte di cassazione, sez. lavoro, del 15 giugno 2020, n. 11543, che ha stabilito che l'addetto all'ufficio stampa pubblico, pur dovendosi iscrivere all'albo dei giornalisti, non può essere assimilato alla figura del giornalista, in quanto sottoposto a direttive e privo di quei tratti di spiccata autonomia nell'acquisizione delle notizie e nell'esercizio del diritto di critica che caratterizzano l'attività giornalistica.

L'ufficio più lontano dal vertice politico è l'ufficio per le relazioni con il pubblico. In linea di massima esso è chiamato non già a dare notizie, a diffondere cioè elementi di novità che caratterizzano la vita dell'istituzione, ma piuttosto a fornire informazioni di pubblica utilità, presentandosi come punto di accesso

10. Cfr. F. Martini, *Il campione nazionale. Primo e quarto potere, vent'anni dopo mani pulite*, in *Problemi dell'informazione*, 2011, n. 1, pp. 66 ss. «Durante la fase declinante della Prima Repubblica i capiufficio stampa dei partiti erano distinti signori che tenevano i rapporti diplomatici con i direttori dei giornali, consigliavano sottovoce, quasi mai si sporcavano le mani. Con la Seconda Repubblica i portavoce interferiscono con efficacia pari alla loro bravura nella fattura quotidiana degli articoli, facendo trapelare frasi «segrete», autorizzando virgolettati, riferendo sintesi di summit, in altre parole diffondendo indiscrezioni che stanno a cuore al loro datore di lavoro».

dei cittadini alla pubblica amministrazione¹¹: vale a dire a rendere più accessibili dati e procedure anche già consolidati ma che l'utenza potrebbe faticare a gestire in autonomia. L'applicazione di sistemi di IA al suo funzionamento merita particolare attenzione. Per il portavoce e l'ufficio stampa il ricorso a sistemi di IA esaurisce la sua efficacia all'interno dell'organizzazione amministrativa posto che, in linea di principio, esso è funzionale a produrre l'informazione che poi sarà trasmessa. Per l'Urp esso può invece avere rilevanza esterna, dal momento che sistemi di IA ormai comuni come i *chatbot* sono utilizzati non già per produrre informazioni, successivamente "esternate" da un operatore umano, ma per trasmetterle direttamente al pubblico senza interazione umana: rispetto a questo specifico utilizzo rileva in modo significativo il principio di trasparenza sancito dall'articolo 50 del regolamento europeo, in forza del quale i sistemi di intelligenza artificiale destinati a una interazione diretta con gli esseri umani devono essere progettati in modo tale da consentire loro di riconoscere la natura non umana del proprio interlocutore; e in particolare quanto dispone il paragrafo 4, comma 2, dello stesso articolo 50, per il quale:

I *deployer* di un sistema di IA che genera o manipola testo pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico rendono noto che il testo è stato generato o manipolato artificialmente. Tale obbligo non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati o se il contenuto generato dall'IA è stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione del contenuto.

Va da sé che lo spazio più o meno ampio che l'azione umana conserva nella produzione e nella diffusione di contenuti informativi presso il pubblico determina poi una serie di questioni, in larga misura non ancora definite, in termini di responsabilità giuridica¹².

11. Per una riflessione sul ruolo effettivamente svolto dagli Urp e sulla percezione da parte dei cittadini cfr. L. D'Ambrosi, M. Laudonio, P. Brescia, G. Radano, *I cittadini al centro. L'ufficio relazioni con il pubblico tra ascolto, partecipazione e intelligenza artificiale*, in *Problemi dell'informazione*, 2024, n. 2, pp. 161 ss.

12. Per una rassegna delle diverse problematiche connesse a questo tema cfr. G. Comandé, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi Giuridica dell'Economia*, 2019, n. 1, pp. 169 ss.; L.M. Lucarelli Tonini, *L'IA tra trasparenza e nuovi profili di responsabilità: la nuova proposta di "AI liability directive"*, in *Il diritto dell'informazione e dell'informatica*, 2023, n. 2, pp. 327 ss.; S. Corso, *Responsabilità per causalità presunta. Intelligenza artificiale, responsabilità civile e diritto*, in *Responsabilità civile e previdenza*, 2025, n. 2, pp. 640 ss.; C. Amato, *Responsabilità civile da "intelligenza" artificiale: prospettive di disciplina europea*, in *Rivista trimestrale di diritto e procedura civile*, 2025, n. 2, pp. 373 ss.

Le differenti tipologie di comunicazione, funzionali alle diverse specie di destinatari, porterà ciascun ufficio a chiedere specifiche prestazioni ai sistemi di intelligenza artificiale il cui livello di rischio potrà essere calcolato rispetto alla concreta possibilità di incidere sulla decisione umana. Da ciò discende la conseguenza che la valutazione del rischio non può essere condotta in astratto rispetto al sistema di IA, ma soltanto in relazione all'uso comunicativo che se vorrà fare: l'uso di un sistema di intelligenza artificiale capace di classificare le abitudini sociali degli abitanti di una città può essere considerato a basso rischio se finalizzato a predisporre gli strumenti comunicativi di un Urp, e a rischio elevato se pensato per valutare il livello di gradimento che un determinato provvedimento potrebbe generare, indipendentemente dalla sua utilità.

4. Conclusioni

Le conclusioni cui permette di giungere questa rapida rassegna sono al contempo parziali e provvisorie. Parziali, perché declinano le questioni sottese all'utilizzo dei sistemi di intelligenza artificiale in relazione a un ambito limitato – per quanto trasversale – dell'azione delle pubbliche amministrazioni; provvisorie, perché la maggior parte delle problematiche affrontate sono in attesa di soluzioni normative soddisfacenti. Ma forse quel che si può constatare, guardando dallo specifico punto di vista della disciplina della comunicazione istituzionale, può essere l'indice di una generale difficoltà del legislatore europeo a stabilire regole operative: il regolamento 2024/1689 del 13 giugno 2024, a dispetto della sua rubrica, non introduce regole, ma principi che, in ragione della propria amplissima portata, richiedono un'azione di continuo bilanciamento da parte di chi sia chiamato ad applicarli¹³. Né ha prodotto effetti di maggiore rilievo l'intervento del legislatore italiano a opera della legge 23 settembre 2025, n. 132 recante "Disposizioni e deleghe al Governo in materia di intelligenza artificiale". La legge muovendo dalle definizioni fornite dal regolamento europeo stabilisce, al capo secondo (artt. 7 ss.) disposizioni di settore, con particolare riferimento all'ambito sanitario e della disabilità, al trattamento dei dati personali, agli ambiti lavorativi e professionali, alla organizzazione dell'amministrazione pubblica e dell'attività giudiziaria¹⁴. Complessivamente

13. Quel che, in qualche misura, ci suggerisce di rivedere in chiave problematica la schematica partizione tra regolamenti e direttive che insiste sulla immediata e uniforme applicabilità dei primi. Se essa era già stata messa in discussione dalle direttive particolareggiate, oggi è ancor più sollecitata dalla vaghezza dei regolamenti.

14. Per quel che concerne l'oggetto specifico delle nostre riflessioni dobbiamo rimarcare che la comunicazione istituzionale non viene presa in considerazione; né una specifica discipli-

considerata, non solo quindi in relazione alle deleghe, ma anche per quanto concerne le disposizioni che si dovrebbero intendere come immediatamente applicabili, la legge dispone per principi.

A caratterizzare la normativa sull'uso dei sistemi di intelligenza artificiale è dunque la vaghezza, e talvolta l'ambiguità; ma forse, in ragione delle caratteristiche del suo oggetto, una disciplina efficace non può che essere vaga e ambigua, se non confusa. Certo, poiché il tema della intelligenza artificiale stabilisce una connessione tra l'informatica e gli ambiti umanistico-sociali, si potrebbe essere indotti a credere che il legislatore debba trattarlo con un linguaggio quanto mai formalizzato, adattando alle strutture della prima le attese normative dei secondi. Ma proprio rispetto a questo tema sono ancora attuali le riflessioni di Perelman, che sottolineava come il diritto, utilizzando una lingua naturale, renda inutili i tentativi di ridurlo alle formalizzazioni tipiche dei linguaggi matematici:

L'inconvénient de cette tentative de réduire la langue naturelle à une langue parfaite suppose que la langue naturelle n'a qu'un seul usage, celui d'être un instrument de communication parfait, ne donnant lieu à aucune ambiguïté, à aucune controverse. Mais, peut-on dire que le verre constitue un matériau parfait parce qu'il est transparent et indéformable ? Qui voudrait confectionner dans ce matériau des chemises et des pantalons ? N'oublions pas, en effet, que la langue naturelle sert à plus d'un usage et que certains de ceux-ci nous obligent de nous écarter des conditions qu'imposent à une langue artificielle, comme celle de la logique formelle, les logiciens et les mathématiciens.

E ricordava che in molte situazioni concrete:

pour promouvoir des objectifs divers, l'on va être obligé soit d'introduire des ambiguïtés et des confusions, soit d'utiliser des notions confuses, soit de tenter de préciser une notion confuse dans un contexte déterminé, ce qui lui ajoutera un nouvel usage, qui viendra augmenter la confusion de cette notion envisagée en dehors du contexte spécifique où elle a été clarifiée¹⁵.

na è prevista nell'ambito della "Strategia nazionale per l'intelligenza artificiale" prevista all'articolo 19 della legge che, per altro, al paragrafo 2, indica quale obiettivo primario la diffusione della conoscenza in materia di IA.

15. Ch. Perelman, *L'usage et l'abus des notions confuses*, in *Logique et Analyse*, 1978, vol. 21, n. 81, p. 5.

IL REGOLAMENTO EUROPEO SULLA PUBBLICITÀ POLITICA, OVVERO DEL DIFFICILE TENTATIVO DI REGOLAMENTARE LA PROPAGANDA ONLINE (E NON SOLO)

*Anna Papa**

SOMMARIO: 1. Premessa – 2. La formazione dell'opinione pubblica in una società disintermediata. Il primo pilastro: l'informazione professionale e la tutela del diritto a essere informati – 3. Il secondo pilastro della formazione dell'opinione pubblica: la comunicazione e propaganda politica – 4. La disciplina europea della propaganda politica – 5. Regolamento europeo e spazi bianchi nella normativa nazionale in materia di comunicazione e propaganda politica.

1. Premessa

L'entrata in vigore, a ottobre 2025, del Regolamento (UE) 2024/900 del Parlamento europeo e del Consiglio, relativo alla trasparenza e al targeting della pubblicità politica¹, e la decisione di due grandi piattaforme digitali, come Meta e Google, di sospendere, almeno per il momento, ogni attività di propaganda sponsorizzata, evidenzia come il legislatore europeo abbia affrontato un tema sensibile sul piano economico e politico, sia perché va a impat-

* Professoressa ordinaria di Diritto costituzionale e pubblico, Università degli Studi di Napoli Parthenope.

1. Il Regolamento, come previsto dalla relativa Proposta presentata dalla Commissione, costituisce una integrazione della normativa sui servizi digitali di cui al Regolamento (EU) 2022/2065 (Dsa) che pone in capo agli intermediari *online* alcuni obblighi generali in relazione alla trasparenza della pubblicità *online*, e delle disposizioni sul trattamento dei dati di cui ai regolamenti (EU) 2016/679 e 2018/1725. Tale disciplina, inoltre, tiene conto delle previsioni del Codice di buone pratiche rafforzato sulla disinformazione del 2022, cui hanno aderito le principali piattaforme *online*, che prescrive a queste ultime determinati impegni per la trasparenza della pubblicità politica e delle campagne di sensibilizzazione. Su questi atti europei, nella linea prospettica che qui rileva, si consenta il rinvio ad A. Papa, *Disintermediazione dell'informazione e della comunicazione politica nelle democrazie digitali*, in *federalismi.it*, 30 maggio 2025.

tare sull'attività e gli interessi dei sovrani digitali, le *big tech*, sia, per la parte che qui maggiormente rileva, perché certifica, semmai ve ne sia bisogno, il rilevante impatto che la circolazione delle opinioni politiche nell'ecosistema digitale ha da tempo sulla formazione dell'opinione pubblica. Sul punto appare, infatti, importante sottolineare sin da subito che, sebbene il Regolamento non faccia esplicito riferimento nel titolo alla dimensione digitale della comunicazione, è proprio su quest'ultima che l'Unione ha deciso di intervenire, proseguendo l'azione di regolamentazione di questo ambito dello spazio comune europeo.

Appare quindi evidente che la volontà di definire regole riguardanti la pubblicità politica non tocchi solo la materia della concorrenza e la tutela dei dati personali – obiettivi comunque dichiarati dal Regolamento² – ma investa le regole stesse della formazione dell'opinione pubblica nello spazio digitale europeo, nella sua dimensione comunitaria ma anche in quella dei singoli Stati membri.

La circolazione di opinioni e notizie rilevanti per il dibattito pubblico avviene, infatti, in modo crescente nello spazio digitale, in particolar modo sulle piattaforme, che hanno mostrato nel tempo atteggiamenti non sempre lineari – talvolta influenzati anche da ragioni proprietarie – nei confronti della selezione e personalizzazione dei contenuti (targeting) e del contrasto alla disinformazione. Al tempo stesso, crescente è la consapevolezza che un importante flusso di notizie e opinioni che circolano in Rete rappresentano il precipitato di processi ideati per influenzare l'opinione pubblica, spesso celati dietro lo schermo dell'espressione del pensiero individuale, basati non di rado su tecniche di disintermediazione e manipolazione dei contenuti³.

2. Su questi aspetti cfr. C. Lobascio, *Micro-targeting politico e protezione dei dati personali*, in *Rivista italiana di informatica e diritto*, 1, 2025; E.A. Rossi, *Problemi di coordinamento normativo nel contrasto alla disinformazione politica e commerciale*, in *federalismi.it*, 30 maggio 2025.

3. La disintermediazione dei processi e delle relazioni caratterizza ormai numerosi ambiti delle odierne società occidentali. Sviluppatisi a partire dagli anni '80, soprattutto in ambito finanziario, ha certamente ricevuto slancio dal diffondersi della Rete e dei *social network* ma, come viene da più parti sottolineato, il suo fondamento può rinvenirsi nella crescente convinzione, individuale e collettiva, di poter svolgere azioni o assumere decisioni senza il ricorso all'intermediazione di esperti e professionisti del settore. Anche la comunicazione e l'informazione nello spazio pubblico sono interessate da tempo da processi di disintermediazione, che hanno preso avvio negli anni '90 nel settore televisivo e si sono poi nel tempo consolidati con la diffusione della Rete e dei *social network*, producendo profondi cambiamenti nelle dinamiche di formazione dell'opinione pubblica e nell'azione dei suoi principali attori (partiti politici, mass media), oltre a favorire l'ingresso nell'arena pubblica di soggetti privati, proprietari delle piattaforme social, che svolgono oggi un ruolo primario nella selezione dei contenuti. Come è stato autorevolmente sottolineato, «gli attori politici e culturali tradizionali non si sono accorti che, nell'apparente vuoto sociale di una comunicazione pubblica disin-

Su questo, l'utilizzo dell'intelligenza artificiale sta peraltro avendo un impatto significativo, anche per la facilità con la quale è possibile realizzare fotografie e video molto verosimili, come nel caso dei deepfake o, in generale, dei contenuti sintetici, al punto che molte piattaforme digitali hanno chiesto ai creator di dichiarare tale natura, qualora presente⁴.

Il quadro che emerge si presenta, quindi, caratterizzato da una significativa complessità, alla quale si aggiunge l'innegabile permeabilità dell'ecosistema digitale rispetto alla condivisione e diffusione di contenuti politici, collegabili a posizioni di post verità o comunque di influenza⁵, provenienti da soggetti extra europei. In parte ciò è dovuto alle caratteristiche proprie della Rete, globalità e atemporalità, in parte tuttavia deriva dall'essere, quello europeo, uno spazio di libertà che produce regole di utilizzo della tecnologia digitale ma non la progetta e produce, trovandosi così a dover mediare con i grandi colossi digitali e i loro Paesi di riferimento. Come si legge nel regolamento

termediata, si inserivano nuovi operatori in grado di creare meccanismi di consenso nei social e attraverso di essi, sfruttando maglie e connessioni che solo qualche anno fa sembravano innocue, veicolando messaggi semplici, schematici, di immediata comprensione, a prescindere dalla loro fondatezza». Il risultato attuale è, come si avrà modo di analizzare, non solo l'opacità di una parte importante del processo di formazione dell'opinione pubblica, nella misura in cui gli algoritmi e i loro proprietari sfuggono alle regole democratiche della trasparenza, ma anche un profondo cambiamento della relazione tra rappresentanti e rappresentati che si manifesta sia nel linguaggio utilizzato sia nel complessivo modo di interagire che porta a una crescente valorizzazione della componente decidente della democrazia rispetto agli elementi propri del parlamentarismo. Su questi aspetti cfr. P. Bilancia, *Crisi della democrazia rappresentativa e aperture a nuove istanze di partecipazione democratica*, in *Federalismi.it*, 2017, 1; B. Caravita, *Social network, formazione del consenso, istituzioni politiche: quale regolamentazione possibile?*, in *Federalismi.it*, 2019, 2; A. Papa, *L'impatto della disintermediazione informativa nei processi di formazione dell'opinione pubblica. Quale (problematica) prospettiva nella "democrazia della comunicazione"*, in questa *Rivista*, 2022, 3; A. Morrone, *Digitale e "governo" della società*, in M. Ladu, N. Maccabiani (a cura di), *L'individuo e la realtà digitale*, Napoli, 2023, 139.

4. Sul tema cfr. M.E. Bucalo, *Intelligenza artificiale e deepfake: le nuove frontiere della disinformazione e i possibili rimedi giuridici*, in *MediaLaws*, n. 3/2024, pp. 55 ss.

5. Su questi aspetti cfr., tra gli altri, F. Pizzetti, *Fake news e allarme sociale: responsabilità, non censura*, in *MediaLaws*, n. 1/2017; C. Pinelli, *"Postverità", verità e libertà di manifestazione del pensiero*, ivi. Su questi aspetti cfr., tra gli altri, G. Marchetti, *Le fake news e il ruolo degli algoritmi*, in *MediaLaws*, n. 1/2020. In generale sul tema del ruolo degli aggregatori di contenuti da parte delle piattaforme, l'Unione europea e a cascata i legislatori nazionali hanno definito regole di tutela degli informatori professionali (giornali in primis), anche con lo scopo di ridurre la distanza, economica, che li separa dai siti di informazione non professionale. Per una analisi della recente normativa in materia si rinvia ai contributi presenti sul numero monografico della rivista *Diritto ed economia dei mezzi di comunicazione*, 2023, dedicato a *Le nuove prospettive del diritto d'autore*.

la pubblicità politica può essere diffusa o pubblicata con vari mezzi e media a livello transfrontaliero, sia online che offline. Sta aumentando rapidamente e può essere diffusa o pubblicata nei media tradizionali offline, come giornali, televisione e radio, ma anche sempre più spesso su piattaforme online, siti web, applicazioni mobili, videogiochi e altre interfacce digitali. Queste ultime non solo si prestano particolarmente a un'offerta di tipo transfrontaliero, ma pongono altresì sfide nuove e difficili sul fronte regolamentare e applicativo. La pubblicità politica online è in forte aumento e alcune forme lineari di pubblicità politica offline, come la radio e la televisione, sono offerte anche online come servizi a richiesta. Nell'organizzare campagne di pubblicità politica, la tendenza è ad avvalersi di un'ampia scelta di media e forme (Considerando 2).

Ne deriva, nella prospettiva del legislatore europeo, «la necessità di garantire la trasparenza che rappresenta un fine pubblico legittimo, conformemente ai valori condivisi dell'Unione e dei suoi Stati membri a norma dell'articolo 2 del trattato sull'Unione europea (Tue)», anche in considerazione del fatto che, da un lato, «non è sempre facile per i cittadini riconoscere i messaggi di pubblicità politica ed esercitare i diritti democratici in maniera consapevole», dall'altro «la crescente sofisticatezza della disinformazione, la diversificazione degli attori, la rapida evoluzione delle nuove tecnologie e l'accresciuta diffusione della manipolazione delle informazioni e delle interferenze nei nostri processi normativi ed elettorali democratici costituiscono sfide importanti per l'Unione e per gli Stati membri» (Considerando 3).

Da qui, come appena accennato, l'esigenza di disciplinare la sponsorizzazione di contenuti immessi nel circuito comunicativo europeo, con particolare attenzione qualora si tratti di finanziatori stranieri, al fine non solo di tutelare la sfera pubblica europea da interferenza esterne, anche se appare legittimo il dubbio sull'effettiva possibilità di tutelare i confini digitali europei, ma anche di limitarne l'impatto, o quanto meno di ridurre l'opacità prodotta sull'opinione pubblica dei singoli Stati membri, nei quali – e l'Italia ne è un esempio – manca una disciplina specifica della comunicazione politica digitale.

2. La formazione dell'opinione pubblica in una società disintermediata. Il primo pilastro: l'informazione professionale e la tutela del diritto a essere informati

Come è noto, i processi di formazione dell'opinione pubblica sono un elemento qualificante dei regimi politici, in particolar modo di quelli democratici, e si basano su una pluralità di elementi: l'azione di partiti e associazioni, la libe-

ra circolazione delle opinioni⁶, l'informazione professionale⁷, e, non da ultima, la comunicazione politica. Una combinazione di elementi, tutti indispensabili, nei confronti dei quali ciascun regolatore (la disciplina è stata sino a oggi essenzialmente nazionale, stante le peculiarità di ciascun sistema) è chiamato a porre regole solo qualora queste ultime si presentino strettamente necessarie per la salvaguardia di diritti fondamentali o dei valori condivisi della comunità. In questa ottica, in Italia, sono presenti regole sull'informazione professionale, affinché essa risponda a canoni complessivi di imparzialità, pluralità e obiettività. Del pari sono presenti regole che limitano la libertà di espressione individuale a tutela della dignità altrui, mentre sino a oggi poche disposizioni sono state introdotte nei confronti dei partiti politici (per i quali sovente si auspicano regole a garanzia della democraticità interna) e della comunicazione politica.

Se dalla dimensione solo analogica si passa a quella digitale, prendendo in considerazione i due pilastri comunicativi, informazione professionale e comunicazione e propaganda politica, la presenza di norme diminuisce in modo significativo: il legislatore è intervenuto, con non poche difficoltà, sulla stampa online in modo molto limitato; per quanto riguarda, invece, la comunicazione e propaganda politica le regole sono sostanzialmente ferme ai manifesti stradali e alla par condicio televisiva.

Eppure, appare evidente che l'informazione professionale (il quarto potere come lo definiva Edmund Burke) e la comunicazione dei soggetti politici nella sfera digitale sono oggi sempre più rilevanti nei processi di formazione dell'opinione pubblica: un numero crescente di persone si informa sul web e mediante le piattaforme digitali, in una modalità disintermediata⁸, o almeno

6. È patrimonio comune la definizione della libertà di espressione come pietra angolare dell'ordine democratico, utilizzata per la prima volta dalla Corte costituzionale nella sentenza n. 84 del 1969 e poi ripresa in numerose altre, nelle quali il Giudice delle Leggi ha sottolineato come tale libertà caratterizzi la forma di Stato vigente in Italia e, pur non potendo parlare di una gerarchia tra i diritti costituzionalmente previsti, ha riconosciuto che la disciplina più garantistica prevista nell'art.21 Cost. si giustifica sotto un profilo di sistema poiché il mantenimento della libertà di manifestazione del pensiero costituisce di per sé la garanzia di tutti gli altri diritti costituzionali. Tra le altre, cfr., le sentenze nn. 9 e 25 del 1965; n. 126 del 1985; n. 293 del 2000; n. 243 del 2001; n. 129 del 2009.

7. Il Costituente e il legislatore repubblicano hanno scritto un sistema peculiare di garanzie e responsabilità riguardante l'informazione professionale al fine di garantire che i cittadini potessero (e possano) formarsi una opinione attingendo a un contesto informativo il più possibile plurale, obiettivo e imparziale. Per una ricostruzione di questo quadro si consenta il rinvio, tra gli altri, ad A. Papa, *Il diritto dell'informazione e della comunicazione nell'era digitale*, Giappichelli, 2021.

8. Come è stato sottolineato, «la ricerca spasmodica da parte degli esponenti politici di ogni colore (ma soprattutto di alcuni) di un contatto diretto con la massa informe dei cybernauti fa sì che proprio il web si ponga quale luogo elettivo nel quale vengono a maturazione o,

apparentemente tale, nella quale sovente ciò che conta è la notizia e non la sua contestualizzazione e verifica, con il rischio di trasformare in informazione ciò che invece non è altro che opinione personale assunta a fatto e di diffondere la convinzione che l'analisi dei fatti e delle opinioni non sia importante quanto descrivere (i primi) e formulare (le seconde)⁹.

Non si tratta di questione di secondaria rilevanza. Infatti, un ecosistema informativo plurale e trasparente – per usare la terminologia del legislatore europeo – richiede non solo una libera informazione ma anche la tutela del diritto di ciascuno di ricevere informazioni corrette e plurali; convinzione che, in Italia, ha portato la giurisprudenza costituzionale a sancire l'esistenza di una specifica situazione giuridica soggettiva, ricavabile dall'art. 21 Cost., che si sostanzia nel "diritto a essere informati". Come appare evidente, fonte di questo diritto non è la disposizione costituzionale di tutela della manifestazione del pensiero, *lato sensu* intesa, dal momento che quest'ultima non appare suscettibile di produrre alcuna pretesa in chi legge o ascolta tale esternazione, stante la sua natura squisitamente individuale. A dare fondamento a questo diritto è l'altra situazione giuridica soggettiva tutelata dall'art. 21 Cost., ossia il diritto di informare; infatti, pur non essendo configurabile un obbligo di informare, è altrettanto vero – come più volte ricordato dalla Corte costituzionale – che qualora un individuo, in particolar modo un giornalista, decida di farlo in modo professionale, allora è chiamato a impegnarsi a produrre una informazione corretta e imparziale, mentre è compito dell'ordinamento giuridico garantire che la stessa sia anche "plurale"¹⁰.

come che sia, sono sollecitate a formarsi le decisioni di maggior rilievo, con riflessi immediati a carico dei diritti fondamentali e per l'intero spettro dei beni della vita costituzionalmente protetti». Così A. Ruggeri, *"Forma di governo" e "sistema dei partiti": due categorie ormai inservibili per la teoria costituzionale?*, in *Consultaonline*, n. 2/2018, p. 7.

9. Si consenta il rinvio ad A. Papa, *Democrazia della comunicazione e formazione dell'opinione pubblica*, in *federalismi.it*, n. 1/2017.

10. La Corte costituzionale, nella definizione del contenuto della libertà di informare, ha sottolineato come quest'ultima presenti una valenza non solo "individuale", ma anche "strumentalmente orientata" al buon funzionamento della vita democratica, con la conseguente riconduzione di tutti i mezzi di comunicazione di massa nell'alveo dei "servizi pubblici o comunque di pubblico interesse", in quanto destinati a soddisfare un "interesse generale" (Corte costituzionale, sentenza n. 112 del 1993). In questo duplice contesto interpretativo si inseriscono le ulteriori numerose sentenze del giudice costituzionale in tema di informazione, nelle quali viene sancito quale imperativo costituzionale che tale diritto venga «qualificato e caratterizzato, tra l'altro, sia dal pluralismo delle fonti cui attingere conoscenze e notizie – così da porre il cittadino in condizione di compiere le proprie valutazioni avendo presenti punti di vista e orientamenti culturali e politici differenti – sia dall'obiettività e dall'imparzialità dei dati forniti, sia infine dalla completezza, dalla correttezza e dalla continuità dell'attività di informazione erogata» (Corte costituzionale, sentenze n. 112 del 1993 e n. 69 del 2009). In particolare, nella sentenza n. 155 del 2002, la Corte sottolinea che «il diritto alla completa e obiettiva informazione del cittadino appare [...] tutela-

Se, quindi, la libertà di informare contiene due distinte situazioni giuridiche soggettive, una finalizzata a informare e l'altra a informarsi¹¹, e se l'ecosistema informativo è composto non solo da stampa e televisione ma anche dalla dimensione digitale, appare consequenziale che senza la garanzia, in tutti e i tre mass media, di regole a tutela della qualità – in termini di oggettività, imparzialità e pluralismo¹² – dell'informazione prodotta, non può parlarsi di un sistema trasparente al servizio di una democratica formazione dell'opinione pubblica.

3. Il secondo pilastro della formazione dell'opinione pubblica: la comunicazione e propaganda politica

Considerazioni analoghe possono farsi in relazione alla comunicazione politica, intesa come scambio e confronto di contenuti di interesse pubblico, prodotti dal sistema politico-istituzionale, dai mass media e dal cittadino, quale individuo e quale elettore.

to in via prioritaria soprattutto in riferimento a valori costituzionali primari, che non sono tanto quelli della pari visibilità dei partiti, quanto piuttosto quelli connessi al corretto svolgimento del confronto politico su cui *sempre* (e quindi anche al di fuori dei periodi di competizione elettorale) si fonda il sistema democratico». Posto quindi l'obiettivo, ossia la realizzazione di una democrazia critica, basata sulla partecipazione attiva e consapevole del popolo alle scelte politiche, la Corte individua tre requisiti necessari: il pluralismo interno ed esterno del sistema dei media, una sostanziale parità di chances nell'accesso e una informazione corretta e imparziale.

11. Cfr., tra le altre, le sentenze della Corte costituzionale n. 112/93 e, in precedenza, nn. 122/70; 105/72; 1/81; 194/87. La Corte correla quindi in modo stringente la possibilità per il cittadino di informarsi e la partecipazione "consapevole" alla formazione della volontà generale. Appare infatti evidente che la Corte voglia sottolineare l'esigenza di offrire all'elettore la possibilità di potersi formare una opinione prima del voto, se vuole, fermo restando che la partecipazione del cittadino alla formazione della volontà popolare mediante il voto è, in un sistema istituzionale basato sul suffragio universale, un diritto assoluto, non limitabile sul piano generale, con la conseguenza che chiunque può votare senza prima formarsi una opinione e senza doverne dare giuridicamente conto ad alcuno. In particolare, nella sentenza n. 112/1993, la Corte costituzionale ha sottolineato come «il "diritto all'informazione" va determinato e qualificato in riferimento ai principi fondanti della forma di Stato delineata dalla Costituzione, i quali esigono che la nostra democrazia sia basata su una libera opinione pubblica e sia in grado di svilupparsi attraverso la pari concorrenza di tutti alla formazione della volontà generale» (su tale decisione, P. Costanzo, *"Diritto all'informazione" e "diritto d'informazione" nei più recenti sviluppi della giurisprudenza costituzionale in materia radiotelevisiva*, in P. Barile, R. Zaccaria (a cura di), *Rapporto '93 sui problemi giuridici della radiotelevisione in Italia*, Giappichelli, 1994, pp. 479 ss.

12. Cfr. F. Donati, *Unione europea, libertà e pluralismo dei mezzi di informazione*, in *MediaLaws*, n. 2/2022; M. Gozzi, *Internet e "standard della community" nella regolamentazione delle piattaforme online: sulla non vessatorietà delle clausole in materia di disinformazione*, in *MediaLaws*, n. 3/2022; L. Scaffardi, *Internet tra autolimitazione e controllo pubblico*, in *Rivista AIC*, n. 4/2023; M. Luciani, *Ogni cosa al suo posto*, Giuffrè, 2023.

Si tratta, in questo caso, di un flusso di informazioni, notizie, opinioni, che si muove nella sfera pubblica, su temi diversi e con contenuti plurimi, e che – a differenza dell'attività giornalistica in senso stretto – persegue l'obiettivo di portare a conoscenza dei cittadini i programmi, i punti di vista e le differenze tra i diversi soggetti presenti sulla scena politica. Appare evidente che si tratti di una azione comunicativa “di parte”, promossa e realizzata dal singolo soggetto politico (esponente politico o partito) e anche per questo si presenta particolarmente reattiva, per forme, luoghi e linguaggio, ai cambiamenti del sistema politico¹³.

La legislazione nazionale fornisce una definizione molto ampia di comunicazione politica, ricomprendendovi sostanzialmente ogni «esposizione di opinioni e valutazioni politiche»¹⁴, anche nella forma dello scambio e confronto di contenuti di interesse pubblico, prodotti dal sistema politico-istituzionale, dai mass media e dal cittadino, quale individuo e quale elettore. In questa accezione, essa può quindi essere rappresentata come un flusso di informazioni, notizie, opinioni, che si muove nella sfera pubblica, su temi diversi e con contenuti plurimi, e che – a differenza dell'attività giornalistica in senso stretto – persegue l'obiettivo di portare a conoscenza dei cittadini i programmi, i punti di vista e le differenze tra i diversi soggetti presenti sulla scena politica¹⁵.

13. Come è stato sottolineato, la crisi partitica degli anni '80 ha coinciso in modo non casuale con l'affermarsi di una nuova strategia comunicativa di tipo politico: quella emozionale, rientrando in una logica di marketing politico che, facendo leva su elementi estranei alla logica del dibattito, mira a soddisfare l'esigenza dell'elettore, il quale ha un tempo limitato da dedicare alla formazione della propria idea, di poter fruire di scorciatoie politiche, ossia di fonti semplificate capaci di agevolare la sua scelta di voto, elettorale o non elettorale. Sul punto cfr. A. Pirozzoli, *La comunicazione politica*, in M. Aini (a cura di), *Informazione, potere, libertà*, Giappichelli, 2005, p. 384; C. Biancalana (a cura di), *Disintermediazione e nuove forme di mediazione. Verso una democrazia post-rappresentativa*, Fondazione Feltrinelli, 2018.

14. L'art. 2, l. 22 febbraio 2000, n. 28, *Disposizioni per la parità di accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica*, definisce comunicazione politica radiotelevisiva «la diffusione sui mezzi radiotelevisivi di programmi contenenti opinioni e valutazioni politiche»; la stessa legge, all'art. 11-ter, lett. c, ritiene programma di comunicazione politica «ogni programma in cui assuma carattere rilevante l'esposizione di opinioni e valutazioni politiche manifestate attraverso tipologie di programmazione che comunque consentano un confronto dialettico tra più opinioni, anche se conseguito nel corso di più trasmissioni».

15. In un sistema democratico la presenza di spazi dedicati alla comunicazione politica, e alla parità di accesso agli stessi, si presenta di particolare rilevanza. In Italia, tuttavia, solo in tempi relativamente recenti si è avvertita l'esigenza di disciplinarne il contesto di riferimento e il funzionamento. Infatti, fatte salve le disposizioni contenute nella legge n. 212 del 1956 (in materia di stampati e volantini elettorali e di convocazione di comizi e riunioni elettorali) e quelle introdotte con la legge n. 103 del 1975 in materia di spazi televisivi autogestiti da associazioni (latamente intese, ricomprendendovi quindi anche e soprattutto partiti e sindacati)

Diversa, ma non nettamente distinguibile, dalla comunicazione si presenta la propaganda politica, intesa come «qualunque attività volta a diffondere in modo deliberato e sistematico messaggi a un determinato uditorio al fine di creare un'immagine (positiva o negativa) di determinati fenomeni, stimolando altresì comportamenti adeguati»¹⁶. La propaganda politica si distingue quindi dalla comunicazione per il fine specifico che persegue, ossia la “persuasione” dei destinatari, il «cercar di acquistare proseliti alle proprie idee»¹⁷ e, da un punto di vista soggettivo, l'essere realizzata da soggetti politici. Essa è quindi una azione comunicativa “di parte”, promossa e realizzata dal singolo soggetto politico (esponente politico o partito), particolarmente reattiva, per forme, luoghi e linguaggio, ai cambiamenti del sistema politico¹⁸, pronta a sperimentare nuove forme di interazione, stili comunicativi e dinamiche del linguaggio¹⁹,

espressive del pluralismo, solo negli anni '90 si è giunti all'approvazione della legge n. 515 del 1993 di disciplina della comunicazione politica nel periodo che precede le consultazioni elettorali (escludendo solo quelle referendarie, poi inserite nella legge successiva) con riferimento, tra l'altro, all'accesso ai mezzi di informazione, al regime economico-finanziario delle campagne elettorali, ai sondaggi, ai divieti di comunicazione istituzionale e pubblica. In particolare, la legge 515/1993 ha rappresentato, come è noto, il primo tentativo posto in essere dal legislatore di dare un contenuto ai concetti di “propaganda”, “informazione” e “pubblicità” elettorale. Tale sforzo è stato poi ripetuto e perfezionato con la legge n. 28 del 2000 (anche nota come legge sulla *par condicio*), di disciplina della comunicazione politica e dell'accesso ai mezzi di informazione durante le campagne elettorali e referendarie. La nuova disciplina ha in parte introdotto modifiche alla legge del 1993 – che comunque resta vigente ancora oggi avendo un contenuto regolatorio che riguarda anche temi diversi dalla comunicazione politica, come il finanziamento delle campagne elettorali – in altra parte ha introdotto elementi nuovi. Come è stato sottolineato, per lungo tempo in Italia il processo di formazione della volontà dei cittadini era limitato al periodo elettorale, che veniva considerato «una sorta di spazio partitico esclusivo, impermeabile a qualsiasi intrusione da parte del potere legislativo, esecutivo e giudiziario». Così G. Gardini, *Le regole dell'informazione*, Bruno Mondadori, 2009, p. 278.

16. G. Mazzoleni, *Comunicazione pubblica*, in *Rassegna italiana di sociologia*, n. 3/1997.

17. F. Lancaster, voce *Propaganda elettorale*, in *Enciclopedia del diritto*, Giuffrè, 1988, p. 126.

18. Come è stato sottolineato, la crisi partitica degli anni '80 ha coinciso in modo non casuale con l'affermarsi di una nuova strategia comunicativa di tipo politico: quella emozionale, rientrando in una logica di marketing politico. Sul punto cfr. A. Pirozzoli, *La comunicazione politica*, cit., p. 384; C. Biancalana (a cura di), *Disintermediazione e nuove forme di mediazione. Verso una democrazia post-rappresentativa*, cit. Sul complesso tema del rapporto tra rappresentanza politica e Rete internet cfr., tra gli altri, S. Staiano, *La rappresentanza*, in *Rivista AIC*, n. 3/2017; A. Ruggeri, “Forma di governo” e “sistema dei partiti”: due categorie ormai inservibili per la teoria costituzionale?, cit. Su questi aspetti si consenta anche il rinvio ad A. Papa, *Rappresentanza politica versus democrazia elettronica tra molteplicità di prospettazioni e problematicità di rapporti*, in L. Chieffi (a cura di) *Rappresentanza politica, gruppi di pressione, élites al potere*, Giappichelli, 2006.

19. Su questi aspetti cfr. P. Marsocci, A. Papa, *Lingua e linguaggio della comunicazione pubblica*, in *Rivista AIC*, n. 1/2023, p. 139.

grazie alla libertà (di scelta dei destinatari, del dove, del quando e delle finalità) che quest'ultimo può assumere, seppure nei limiti della tutela della dignità delle persone, del decoro e del prestigio delle istituzioni e pur in presenza di fenomeni di particolare criticità come i discorsi d'odio²⁰ o la disinformazione²¹.

Sul piano teorico questo tipo di propaganda presenta caratteri più generali rispetto a quella elettorale, intesa quale «attività inerente direttamente o indirettamente alla competizione elettorale²² e tendente, come scopo mediato o immediato, ad acquistare voti o a sottrarli agli avversari»²³, anche se appare un dato condiviso che, in una sistema di governance politica multilivello, le campagne elettorali, alle quali si affiancano sovente anche quelle referendarie, si susseguono senza soluzione di continuità, tra scadenze europee e nazionali, rinnovi delle rappresentanze regionali e di quelle delle principali città capoluogo. Una campagna elettorale permanente dove quindi i periodi della propaganda elettorale e di quella politica finiscono per lo più per coincidere, rendendo quindi non sempre agevole comprendere in che modo applicare la normativa di riferimento²⁴.

20. Sui discorsi d'odio, cfr. P. Falletta, *Controlli e responsabilità dei social network sui discorsi d'odio online*, in *MediaLaws*, n. 1/2020, pp. 146 ss.; L. Califano, *La libertà di manifestazione del pensiero... in rete; nuove frontiere di esercizio di un diritto antico. Fake news, hate speech e profili di responsabilità dei social network*, in *federalismi.it*, 2021, 26.

21. Sul tema della disinformazione, nella prospettiva di un sistema economico in trasformazione, cfr. A. Nicita, *Il mercato della verità. Come la disinformazione minaccia la democrazia*, il Mulino, 2021; F. Fabrizzi, *La cura contro la disinformazione. Rimedi giuridici (e non) per contrastarne la diffusione*, in questo volume.

22. E. Bettinelli, voce *Propaganda elettorale*, in *Dig. disc. pubbl.*, II, Utet, 1997, p. 61, definisce la propaganda elettorale come «l'insieme di variegate attività realizzate dai soggetti competitori di qualsiasi gara elettorale per acquisire il consenso dei cittadini titolari del diritto di voto» e aggiunge che «questi vengono presi in considerazione non solo in quanto "massa" (o platea) indifferenziata, ma anche e (più efficacemente) in quanto appartenenti a differenti categorie o situazioni. Cosicché essi possono essere raggiunti da messaggi *personalizzati*, tali cioè che tengano conto delle particolari condizioni di vita e degli specifici interessi addirittura dei singoli»; parla, invece, di «forma di induzione della volontà degli elettori da qualificare come *dolus bonus*» U. Prosperetti, *L'elettorato politico attivo*, Giuffrè, 1954, p. 149. Sulla propaganda dal punto di vista costituzionalistico, cfr. anche: S. Furlani, *Propaganda elettorale*, in *Noviss. dig. it.*, XIV, Utet, 1967, pp. 92 ss.; A.A. Romano, voce *Propaganda elettorale*, in *Enc. giur. Treccani*, XXIV, Istituto Treccani, 1991; M. Gobbo, *La propaganda politica nell'ordinamento costituzionale*, il Mulino, 1997, p. 69; G. Sani, voce *Propaganda*, in N. Bobbio, N. Matteucci e G. Pasquino, *Dizionario di politica*, Utet, 2016; F. Meola, *Tecnologie digitali e neuro-marketing elettorale. A proposito di una possibile regolamentazione delle nuove forme di propaganda politica*, in *Costituzionalismo.it*, n. 1/2020; M. Monti, *Le categorie del costituzionalismo e le fonti alla prova della propaganda politica online: utente-consumatore o utente-cittadino*, in *DPCE Online*, n. 50/2021, pp. 1269 ss.

23. E. Caterina, *La comunicazione elettorale sui social media tra autoregolazione e profili di diritto costituzionale*, in *Osservatorio sulle fonti*, n. 3/2021, p. 1395.

24. Diversificate sono le definizioni di propaganda elettorale presenti nella legislazione na-

Come prima ricordato, una parte significativa della propaganda politica odierna si sviluppa sulle piattaforme digitali, che consentono una interlocuzione diretta tra politici ed elettori, bypassando non solo il sistema dei media ma le stesse strutture di partito. Sono soprattutto i leader di partito a ritenere particolarmente proficua questa forma di comunicazione disintermediata, maggiormente rispondente alle attuali dinamiche relazionali ma anche alla forma dei movimenti contemporanei, poiché consente loro un dialogo diretto con i cittadini²⁵.

Sul piano sistemico questa forma di propaganda presenta, agli occhi del giurista, due problematiche fondamentali. La prima si sostanzia nella crescente difficoltà di segnare una linea di demarcazione con la comunicazione in generale e con quella istituzionale in particolare²⁶, soprattutto nei casi in cui chi ri-

zionale. L'art. 1, l. 4 aprile 1956, n. 212, "Norme per la disciplina della propaganda elettorale", si riferisce all'«affissione di stampati, giornali murali o altri e di manifesti di propaganda, da parte di partiti o gruppi politici che partecipano alla competizione elettorale»; l'art. 6 fa riferimento anche a «propaganda elettorale luminosa o figurativa, a carattere fisso in luogo pubblico» (come le insegne luminose e a esclusione di quelle indicanti le sedi dei partiti) e al «lancio o il getto di volantini in luogo pubblico o aperto al pubblico e ogni forma di propaganda luminosa mobile». Dalla l. 10 dicembre 1993, n. 515, "Disciplina delle campagne elettorali per l'elezione alla Camera dei deputati e al Senato della Repubblica", si evince la medesima nozione di propaganda elettorale, in questo caso declinata soprattutto con riferimento all'accesso agli spazi televisivi oltre che a «le pubblicazioni di propaganda elettorali a mezzo di scritti, stampa o foto stampa, radio, televisione, incisione magnetica e ogni altro mezzo di divulgazione».

25. Il tema del consenso e del suo linguaggio è stato indagato da diverse prospettive. Cfr., tra gli altri e in una prospettiva multidisciplinare, S. Rodotà, *Tecnopolitica, La democrazia e le nuove tecnologie della comunicazione*, Laterza, 2004; B. Caravita, *Il consenso ai tempi del social, in federalismi.it*, n. 13, 2018; C. Cepernich, R. Bracciale, *Sfumature del razionale. La comunicazione politica emozionale nell'ecosistema ibrido dei media*, in *Comunicazione politica*, n. 1/2019.

26. Un tentativo di definire una linea di demarcazione tra le due forme di comunicazione è presente nella legge n. 28 del 2000 ("Disposizioni per la parità d'accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica"), che prevede il divieto di svolgere attività di comunicazione e di informazione istituzionale nel periodo delle campagne elettorali. Il divieto era già stato introdotto dalla legge n. 515 del 10 dicembre 1993, primo intervento volto a disciplinare gli spazi e i tempi della campagna elettorale, che imponeva a tutti gli enti e le amministrazioni pubbliche di astenersi da iniziative di promozione e pubblicità della propria attività nei trenta giorni precedenti l'inizio della campagna elettorale e per la durata della stessa, al fine di evitare un uso scorretto di questi strumenti, una propaganda elettorale velata, nonché una distorsione della competizione elettorale tra amministratori uscenti e nuovi soggetti politici. La disciplina del 2000 muove proprio dall'esigenza di risolvere le ambiguità del precedente regime. Il periodo relativo al divieto, viene ora chiarito, va dalla data di convocazione dei comizi elettorali, fino alla chiusura delle operazioni di voto, e gli obblighi riguardano ogni forma o strumento di propaganda, con qualsiasi tecnica e a qualsiasi scopo effettuata e anche se finalizzata alla promozione dell'immagine politica e dell'attività istituzionale dell'ente e non del singolo rappresentante. Tali disposizioni valgono per le elezioni europee, nazionali, regionali, mentre non abrogano la disciplina "speciale" dettata in tema di campagna per l'elezione diretta

copre i vertici di un organo o ente pubblico – ministro, presidente di regione, sindaco ecc. – si trova nella contemporanea legittima posizione (e conseguente responsabilità) di esponente di un partito e di vertice dell'amministrazione. Mentre sono previste, nelle dinamiche tradizionali, alcuni, seppur blandi, strumenti per evitare che la comunicazione istituzionale, che è comunicazione dell'ente e non del politico, venga utilizzata da quest'ultimo con finalità di propaganda, questa linea di demarcazione in Rete non è neppure immaginata²⁷.

La seconda criticità è l'asimmetria normativa che caratterizza la propaganda politica nello spazio fisico, nei due mass media tradizionali e in Rete. Sul punto occorre premettere che in Italia solo dopo diversi decenni di vita repubblicana si è avvertita l'esigenza di disciplinarne il contesto di riferimento e il funzionamento; infatti, fatte salve le disposizioni degli anni '50 e '70, solo con le leggi del 1993 e del 2000 si è giunti alla definizione di una articolata disciplina della comunicazione e propaganda politica, con particolare riferimento all'ambito televisivo²⁸.

del sindaco (legge n. 81, del 25 marzo 1993, artt. 28 e 29), dove il divieto di propaganda è assoluto, non limitabile nemmeno a tutela dell'efficace svolgimento delle funzioni degli enti.

27. Non di rado si assiste, nella comunicazione politica sulle piattaforme digitali, alla situazione in cui il titolare di una carica istituzionale, che in autonomia o per il tramite del proprio portavoce abbia espresso valutazioni su una vicenda in un comunicato stampa o con un commento su un proprio sito o profilo social istituzionale, decida di commentare la medesima vicenda anche sul proprio profilo social personale, utilizzando un linguaggio diverso, meno contenente e anche per questo poco istituzionale; o ancora può capitare che un *social network* decida di ripubblicare nella sfera pubblica post e commenti che un soggetto pubblico aveva espresso mesi o anni prima, quindi fuori da una corretta contestualizzazione: o ancora che un politico dichiari di non aver personalmente scritto un determinato post sul proprio profilo istituzionale, ma che a farlo sia stato non il portavoce bensì il proprio *social media manager*, dando rilievo a una figura professionale, ancora oggi poco nota, alla quale in ogni caso non si applicano le incompatibilità normativamente previste per il portavoce e che, soprattutto, rende legittimo anche il quesito su chi stia effettivamente comunicando nella sfera pubblica mediante la diffusione di un contenuto politico su un social network. Su questi temi cfr., L. Solito, *Comunicazione istituzionale e società digitale. Un nuovo ruolo per i comunicatori pubblici*, in *Comunicazione politica*, n. 3/2018; B. Baldi, *Il linguaggio del potere. Comunicazione politica e società*, Carocci, 2021.

28. L. 10 dicembre 1993, n. 515 ("Disciplina delle campagne elettorali per l'elezione alla Camera dei deputati e il Senato della Repubblica") e L. 22 febbraio 2000, n. 28 ("Disposizioni per la parità di accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica") e successive modificazioni e integrazioni. Al combinato disposto di queste due leggi si deve la disciplina, che si applica sia nei periodi elettorali e referendari, sia durante la legislatura, dell'accesso ai mezzi di informazione, del regime economico-finanziario delle campagne elettorali, dei sondaggi, dei divieti di comunicazione istituzionale e pubblica, oltre al tentativo di dare un contenuto ai concetti di "propaganda", "informazione" e "pubblicità" elettorale. Restano tuttavia alcune importanti criticità, dal momento che la normativa attuale non fornisce alcuna indicazione né sui soggetti legittimati ad accedere agli spazi televisivi, attribuendo così ampia discrezionalità alle emittenti, né sui criteri di ripartizione degli spa-

Questa normativa, pur essendo stata approvata in un momento storico nel quale erano già presenti segnali importanti di disintermediazione della comunicazione politica televisiva²⁹ e di crescente rilevanza della Rete e dei social network, non ha introdotto alcuna disposizione riguardante l'ecosistema digitale. Ne deriva che la propaganda che esso veicola è esclusa da ogni forma di regolamentazione, sia sul piano delle garanzie del contraddittorio, che si ritengono forse superate dall'indubbio pluralismo esterno che Internet consente e che invece rappresenta un argine importante alla diffusione di notizie volutamente false³⁰, sia su aspetti solo apparentemente secondari, quali il silenzio elettorale – in Rete praticamente inesistente – o il divieto di pubblicare sondaggi o orientamenti di voto nei quindici giorni che precedono la giornata elettorale e fino alla chiusura delle operazioni di voto³¹, anch'esso frequentemente violato sul *web*³².

4. La disciplina europea della propaganda politica

Come sottolineato in premessa, l'Unione europea rappresenta ormai in modo consolidato il livello normativo dell'ecosistema digitale, in tutte le sue sfaccettature e dimensioni. Negli ultimi anni, il legislatore comunitario ha intrapreso anche la strada della regolamentazione dell'informazione e della comu-

zi stessi, compito quest'ultimo che è affidato alla Commissione parlamentare di vigilanza per quanto riguarda la Rai e all'Agcom per le emittenti private.

29. Come è stato sottolineato, nelle trasmissioni informative in senso lato si pone sempre l'esigenza di mantenere un necessario distacco e un equilibrio rispetto all'attualità politica. Ciò è però particolarmente difficile in quei programmi per i quali non è sempre agevole tracciare una «valida demarcazione tra informazione e comunicazione politica». Cfr. G. Gardini, *Le regole dell'informazione. Dal cartaceo al bit*, Giappichelli, 2014, p. 102.

30. Su questi aspetti cfr. F. Donati, *Internet e campagne elettorali*, in *federalismi.it*, n. 16, 2019.

31. Come previsto nel Regolamento Agcom «in materia di pubblicazione e diffusione dei sondaggi sui mezzi di comunicazione di massa» (Allegato A alla delibera n. 256/10/CSP del 9 dicembre 2010) si ha inosservanza del divieto citato nel testo «quando vengono riportate nel circuito dell'informazione radiotelevisiva, delle edizioni cartacee o elettroniche di quotidiani o periodici, o della diffusione di notizie mediante agenzia, dichiarazioni concernenti i risultati di sondaggi politici ed elettorali rilasciate da esponenti politici o da qualunque altro soggetto in qualsiasi sede, a meno che i sondaggi cui tali dichiarazioni si riferiscono non siano già stati resi pubblici, secondo le forme stabilite dagli articoli 4 e 5, nel periodo precedente», ossia più di quindici giorni prima del giorno delle elezioni.

32. Sull'influenza dei sondaggi sulle scelte degli elettori cfr., di recente, L. Bartolucci, *Quando i seggi parlamentari sono attribuiti anche in base ai sondaggi*, in *Forum di Quaderni costituzionali*, 2023, pp. 87 ss.; sul silenzio elettorale cfr., tra gli altri, F. Fabrizzi, *Campagne elettorali, formazione del consenso, astensionismo. Nel silenzio assordante sui referendum, ha ancora senso il silenzio elettorale?*, in *federalismi.it*, n. 16, 2022.

nicazione in Rete e sulle piattaforme con strumenti di diverso grado di armonizzazione ma dando comunque il segnale che la dimensione online in Europa, qualunque sia l'oggetto, richiede regole applicabili senza barriere nazionali.

In questo contesto, il Regolamento (EU)2024/900 interviene, come anticipato, sulla circolazione dei contenuti di quella che viene definita pubblicità politica, che nell'ordinamento italiano, si sostanzia nella propaganda politica, dal momento che il termine pubblicità viene usato dal legislatore nazionale solo con l'aggettivazione "elettorale" e avendo di conseguenza un contenuto definito e diverso. Del pari, considerata la ricordata sottile linea di demarcazione tra comunicazione e propaganda politica, può dirsi che esse finiscono, nel momento in cui svolte da un soggetto politico, nel confluire nella medesima definizione che il regolamento europeo dà della seconda, definendola come un messaggio fornito normalmente dietro retribuzione o tramite attività interne o nell'ambito di una campagna di pubblicità politica: a) di, a favore o per conto di un attore politico, salvo se di natura meramente privata o meramente commerciale; oppure b) che possa e sia inteso a influenzare l'esito di un'elezione o referendum, un comportamento di voto o un processo legislativo o regolamentare, a livello dell'Unione, nazionale, regionale o locale.

Il primo punto che si può fissare è, quindi, che il legislatore europeo (composto – giova ricordarlo – dal Parlamento europeo e dai rappresentanti dei governi nazionali nel Consiglio) ha disciplinato uno strumento proprio del dibattito democratico, la propaganda politica. E, pur essendo evidente che l'obiettivo principale sia la dimensione digitale, il Regolamento non dichiara di volersi limitare a questo, risultando suscettibile di applicazione in tutti gli spazi della comunicazione politica.

Ciò emerge anche dalla definizione di messaggio, che pone il baricentro di attenzione sull'obiettivo di quest'ultimo, che è quello di influenzare i cittadini o i decisori pubblici sia nella fase delle elezioni sia durante la legislatura (di qualsiasi livello della multilevel governance europea). Ciò che si vuole regolamentare è quindi l'attività di creazione di messaggi politici di tipo "professionale", in grado di stimolare persuasione nei confronti dei destinatari, ancor più se questi ultimi vengono preliminarmente selezionati mediante attività di targeting.

Di conseguenza, il regolamento esclude dal proprio ambito di applicazione l'opinione espressa a titolo personale dal soggetto politico³³. Come appare evi-

33. Sulla base di quanto previsto dal Considerando 30, «Le opinioni politiche espresse a titolo personale costituiscono una particolare manifestazione del diritto alla libertà di espressione e di informazione. Non si tratta di pubblicità politica ed è necessario che tale distinzione sia operata in modo chiaro. A tal fine, non dovrebbero rientrare nel campo di applicazione del presente regolamento le opinioni politiche espresse a titolo personale. L'individuazione delle opinioni politiche espresse a titolo personale dovrebbe di norma derivare dall'autodetermina-

dente, si tratta di una specificazione da un lato importante, volta a chiarire che non si vogliono porre le premesse per attività censorie o limitatrici della libera manifestazione del pensiero politico, dall'altra rivelatrice di un dato non sempre percepito dagli utenti, ossia che il dialogo dei politici con i cittadini, sebbene realizzato in una modalità informale e diretta tipica dei social, difficilmente è il frutto di un esercizio personale del soggetto che comunica. Anche questo tipo di comunicazione, come già emerso in passato con l'attività dei portavoce, rappresenta il prodotto dell'attività di professionisti, molto spesso strutturati in società di servizi, dei quali si avverte ora l'esigenza di conoscere le modalità di funzionamento e il finanziamento delle campagne: in altri termini, nelle parole del Regolamento, gli sponsor e il valore delle sponsorizzazioni, soprattutto qualora provenienti da soggetti esterni all'Unione.

Del pari, il Regolamento, pur affermando che le sue disposizioni non si applicano in via generale all'informazione professionale, già oggetto di regolamentazione negli Stati membri di stabilimento con riferimento alla responsabilità editoriale, estende anche a quest'ultima gli obblighi di trasparenza relativamente alla pubblicazione di contenuti politici sponsorizzati³⁴. Si tratta di una previsione di non poco conto perché da un lato accende un faro su una pratica che, sul piano del diritto a essere informati, si auspica sia molto limitata, ossia quella delle sponsorizzazioni non dichiarate, e dall'altro conferma che il focus della regolamentazione europea è la natura del messaggio veicolato, ovunque e da chiunque.

zione individuale, ma possono essere presi in considerazione elementi contestuali. Tra i fattori pertinenti potrebbero rientrare il fatto che il parere sia espresso per conto di un altro soggetto, che il messaggio intenda autopromuovere una candidatura o una campagna nell'ambito di un processo elettorale, referendario, legislativo o regolamentare, che sia espresso da un individuo generalmente attivo nel partecipare a campagne o nell'intraprendere azioni a favore del cambiamento su questioni politiche o sociali e che il messaggio sia diffuso a un numero indefinito di persone. Un'opinione politica non dovrebbe essere considerata espressa a titolo personale se è prevista una remunerazione specifica da parte di terzi, comprese prestazioni in natura, a fronte dell'espressione di tale opinione o in relazione alla stessa».

34. Come si legge nel Considerando 29, «i media contribuiscono al buon funzionamento dei processi democratici e svolgono un ruolo essenziale nel garantire la libertà di espressione e di informazione, in particolare nei periodi immediatamente precedenti le elezioni. Essi offrono uno spazio di dibattito pubblico e contribuiscono alla formazione delle opinioni. Il presente regolamento non dovrebbe pertanto pregiudicare la libertà editoriale dei media. Le opinioni politiche espresse su qualsiasi supporto sotto la responsabilità editoriale non dovrebbero essere contemplate, a meno che non sia stato fornito da terzi un pagamento specifico o altra remunerazione per la loro preparazione, collocazione, promozione, pubblicazione, consegna o diffusione o in relazione a tali attività. Tuttavia, laddove siano successivamente promosse, pubblicate o diffuse da prestatori di servizi di pubblicità politica, tali opinioni politiche potrebbero essere considerate pubblicità politica».

Da qui l'importanza di questo Regolamento, il cui impatto è stato immediatamente ben compreso dalle *big tech*, che da tempo svolgono anche il ruolo di fornitori di servizi pubblicitari, sui quali la nuova normativa fa gravare una serie di obblighi che vanno dalla raccolta delle dichiarazioni degli sponsor in merito alla natura politica dei messaggi che intendono veicolare, anche al fine di poterli etichettare come tali al momento della diffusione, sino alla pubblicazione dei relativi avvisi di trasparenza³⁵.

Tuttavia, sembra potersi esprimere la convinzione che non siano solo gli oneri burocratici a produrre resistenza nei fornitori di servizi, quanto piuttosto la consapevolezza che l'atto europeo ha l'obiettivo di ridurre quell'area di opacità che ormai da tempo sta caratterizzando l'informazione e la comunicazione in Rete, sempre meno luogo di equali e sempre più spazio nel quale l'individuo digitale rischia di involuparsi nella "illusione della scelta", convinto di partecipare in modo trasparente e paritario al processo decisionale e ritrovandosi, invece, inconsapevolmente guidato in scelte predeterminate, grazie a processi di *news engagement*, di paternalismo sociale e politico³⁶. Per questo, per combattere la manipolazione informativa occorre riconoscerne che essa è presente anche, come in passato, nei sistemi contemporanei ed è in grado di influenzarli in modo significativo³⁷. Ha solo mutato forma e strumenti, in un'ottica globale che produce cambiamenti negli equilibri geopolitici e rende l'Europa uno spazio di libertà ma anche un'isola digitale troppo dipendente dalla tecnologia altrui.

5. Regolamento europeo e spazi bianchi nella normativa nazionale in materia di comunicazione e propaganda politica

Come prima ricordato, la normativa italiana in materia di comunicazione e propaganda politica si presenta alquanto datata e lacunosa, se letta nell'ottica della società digitale. Questo spazio comunicativo è praticamente non rego-

35. Gli editori, in base agli artt. 11 e 12 del Regolamento sono tenuti a effettuare una serie di adempimenti idonei a fornire quante più informazioni possibili sulla natura della sponsorizzazione. Si tratta di un onere di non poco conto che sta già suscitando, oltre – come già accennato – alla autosospensione di Meta e Google dalla pubblicazione sulle loro piattaforme di questo tipo di messaggio, notevoli perplessità degli operatori del settore. Su questo punto si rinvia ai siti di Agenda digitale e dell'Associazione Nazionale Social Media Manager.

36. Cfr. L. Morlino, M. Sorice (a cura di), *L'illusione della scelta*, Luiss University Press, 2021.

37. Sui tanti aspetti, sociali, culturali, linguistici e giuridici, della manipolazione cfr. C. Gravet, M.G. Petrillo, V. Sperti (a cura di), *Manipulation*, Centre International de Phonétique Appliquée CIPA, 2026.

lamentato e ciò ha favorito l'utilizzo del web e delle piattaforme digitali per forme innovative di dialogo tra politici e cittadini. Di questo il sistema nel suo complesso ha certamente beneficiato, anche grazie alla bidirezionalità della comunicazione, pur nella consapevolezza che queste forme di interazione sono diventate nel tempo sempre più il frutto di strategie comunicative di social media manager³⁸.

Un quesito che l'approvazione del Regolamento europeo porta con sé riguarda la possibilità che esso possa trovare applicazione negli spazi vuoti della disciplina nazionale in materia. L'atto europeo dichiara espressamente di non voler incidere sulle regolamentazioni nazionali in materia di pubblicità politica. Tuttavia, appare necessario chiedersi se questa volontà presenti un valore assoluto, quindi volta a ritenere che esista una riserva di legge nazionale in materia, o se solo in presenza di una regolamentazione nazionale il regolamento europeo sia destinato ad arretrare.

Alcune disposizioni portano a ritenere prevalente questa seconda ipotesi, a partire dall'utilizzo del verbo utilizzato. Il regolamento chiarisce, infatti, di non voler *pregiudicare* «il contenuto dei messaggi di pubblicità politica o le norme dell'Unione o nazionali che disciplinano aspetti relativi alla pubblicità politica diversi da quelli contemplati dal presente regolamento, comprese le norme sull'organizzazione, il finanziamento e lo svolgimento di campagne politiche, le norme sui divieti o le restrizioni generali imposti alla pubblicità politica durante determinati periodi e, se del caso, le norme sui periodi elettorali». La disposizione sembra quindi riferirsi a regole esistenti, non a generali riserve.

Peraltro, quando fa riferimento ai periodi elettorali nei quali si applica una normativa più stringente a tutela dell'autodeterminazione politica dei cittadini, il Regolamento non cita solo le elezioni per il rinnovo del Parlamento europeo ma anche tutte le elezioni o referendum organizzati a livello nazionale, regionale e locale negli Stati membri, spingendosi a considerare come momento elettorale anche elezioni per stabilire la leadership dei partiti politici, le primarie nel vocabolario politico italiano (Considerando 31). Se è così ap-

38. Come viene sottolineato (G. Palombino, *L'insindacabilità digitale al tempo della democrazia digitale*, in *federalismi.it*, 21 settembre 2022), non appare immotivato chiedersi chi sia in realtà a esprimersi quando il rappresentante posta un'opinione su un social network, chi sia a formulare il pensiero e materialmente a condividerlo in rete. È ormai noto, infatti, che le piattaforme generano l'illusione, in capo all'utente, di entrare in contatto diretto con chi condivide un'opinione, ma in realtà, all'apparente personalizzazione della comunicazione, corrisponde una sua spersonalizzazione, dovuta al fatto che a parlare è una figura professionale nuova e che risponde al nome di social media manager. Su questi aspetti cfr., tra gli altri, I. Spadaro, *Comunicazione politica e democrazia digitale*, in *DPE. Rassegna online*, n. 1, 2020; A. Lauro, *Siamo tutti giornalisti? Appunti sulla libertà di informazione nell'era social*, in *MediaLaws*, n. 2/2021.

pare difficile immaginare una autolimitazione del regolamento in presenza di lacune nell'ordinamento nazionale – come accade per quello italiano; è invece espressamente previsto che gli Stati membri possano prevedere periodi di tempo più lunghi per la durata delle restrizioni alla sponsorizzazione da parte di entità o cittadini di paesi terzi o norme nazionali più rigorose *nel rispetto del diritto dell'Unione* ma aprendo alla possibilità che il regolamento, pur essendo un atto di armonizzazione forte, non trovi applicazione.

Restano tuttavia, in conclusione, un dato e un dubbio. Il primo è la conferma, se ancora fosse necessario, che la regolamentazione dei processi legati alla tecnologia digitale è sempre più europea: tutela dei dati e del mercato digitale, intelligenza artificiale³⁹, contrasto alla disinformazione, ora regolamentazione della pubblicità politica. Non è sempre scontato e scritto nei Trattati e anche questa volta si è reso necessario un accordo politico tra consiglio europeo e parlamento per rendere possibile questa approvazione.

Il dubbio riguarda, invece, la reale possibilità che questo Regolamento trovi ampia applicazione e soprattutto condivisione. L'opposizione manifestata dalle piattaforme rappresenta certamente un segnale importante, anche in un'ottica di rapporti di forza tra zone geografiche (Europa e Stati Uniti in particolare). La strada scelta, tuttavia, appare quella giusta, se capace di avvicinare forma e sostanza, di aumentare la consapevolezza comune della significativa quota di contenuti sponsorizzati che circolano in Rete senza che sino a oggi dovessero apparire come tali. A giudicare dalla drastica riduzione dei messaggi di pubblicità politica ricevuti dagli utenti nei primi mesi di applicazione del regolamento, erano davvero molti! E allora, piace pensare che si possa associare alla parola pluralismo quella di trasparenza dell'agire comunicativo di coloro che, in virtù del rapporto rappresentativo, sono chiamati ad assumere decisioni nell'interesse generale.

39. L'approvazione della recente legge 23 settembre 2025, n. 132, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, non modifica questa considerazione, dal momento che il testo in vigore si inserisce in quegli inevitabili spazi bianchi della regolamentazione europea ma non può che porsi in una posizione integrativa, come in effetti appare, rispetto al regolamento EU 2024/1689 del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale.

LA COMUNICAZIONE POLITICA NELLA DIGITAL ERA

*Giuliaserena Stegher**

SOMMARIO: 1. Considerazioni introduttive – 2. La comunicazione politica online nel contesto euro-unionale – 3. La comunicazione politica alla prova del nuovo regolamento europeo 2024/900 – 4. L'armonizzazione europea della pubblicità politica digitale: i timori di alcuni Parlamenti nazionali – 5. La democrazia nell'era algoritmica: tra sfide e prospettive.

1. Considerazioni introduttive

Negli ultimi anni abbiamo assistito alla proliferazione di profili sulle piattaforme e all'adozione di modalità comunicative basate su applicazioni di messaggistica istantanea centralizzata. A ciò ha fatto seguito un duplice fenomeno: da un lato, l'instaurazione di un rapporto di affiliazione virtuale tra candidati e rispettivi elettorati; dall'altro, una progressiva mercificazione dell'immagine dei vari candidati, concepiti e proposti secondo logiche di marketing. Questa evoluzione ha profondamente ridisegnato le strategie di comunicazione politica.

Sicché, le tradizionali metodologie di campagna elettorale – caratterizzate da inserzioni sulla stampa cartacea, materiale propagandistico cartaceo, comunicazioni pubblicitarie televisive e radiofoniche, confronti pubblici, messaggistica telefonica e comizi – sono state affiancate, se non addirittura sostituite, da contenuti diffusi su Twitter, Facebook, Instagram, TikTok, ecc. In tale contesto, la costante accessibilità digitale ha ridotto la distanza fisica tra elettori ed eletti, consentendo ai primi un monitoraggio continuo e ubiquitario dell'operato dei secondi.

Tuttavia, questa digitalizzazione ha anche generato dinamiche che trascendono i tradizionali schemi della democrazia rappresentativa, manifestandosi in forme di mobilitazione spontanea in diversi paesi. Ne è testimonianza l'on-

* Ricercatrice di Diritto pubblico comparato, Università degli Studi di Urbino Carlo Bo.

data di proteste guidate dalla Generazione Z che, dal Nepal al Marocco, dal Madagascar all'Indonesia, dall'Algeria al Perù, ha scosso numerosi governi negli ultimi mesi. Questi movimenti, privi di leader formali e strutture partitiche tradizionali, si caratterizzano per un'organizzazione orizzontale interamente mediata da piattaforme digitali quali TikTok, Instagram, Discord e Telegram. In Madagascar, lo slogan "Miala Rajoelina!" è diventato virale sui social, alimentando manifestazioni che hanno provocato 22 vittime, costringendo il presidente Andry Rajoelina allo scioglimento del governo. In Marocco, il movimento della Gen Z 212 si è rivolto direttamente a re Mohammed VI chiedendo lo scioglimento dell'esecutivo per la mancata tutela dei diritti costituzionali. In Nepal, il tentativo governativo di vietare i social network ha innescato proteste di massa culminate nell'incendio del palazzo del Parlamento e nelle dimissioni del governo. Questi giovani manifestanti, armati di smartphone e slogan tratti dalla cultura pop, dimostrano come le piattaforme digitali non costituiscano più soltanto strumenti di propaganda elettorale *top-down*, ma abbiano abilitato forme di partecipazione politica *bottom-up* capaci di far vacillare la stabilità dei governi. Le strategie di protesta si condividono in tempo reale attraverso canali digitali, in una logica di coordinamento decentralizzato che sfugge ai meccanismi di controllo tradizionali.

Nel frattempo, l'accesso asimmetrico alle informazioni e il controllo sui dati stanno plasmando un nuovo regime tecnocratico¹, dove algoritmi e intelligenza artificiale non si limitano a mediare, ma determinano processi sociali ed economici e, parimenti, politici, instaurando una forma di dominio invisibile che sfugge ai tradizionali meccanismi di controllo democratico².

1. N. Bobbio, *Il futuro della democrazia*, Giappichelli, 1995; M. Volpi, *Tecnocrazia e crisi della democrazia*, in M. Volpi (a cura di), *Governi tecnici e tecnici al governo*, Giappichelli, 2017, pp. 1-20; M. Calamo Specchia, *Processi politici, nuove tecnologie e democrazia digitale. Una nuova frontiera del costituzionalismo*, in Id. (a cura di) *Processi politici e nuove tecnologie*, Giappichelli, 2024, p. 1 ss.

2. Sull'utilizzo dei social media e degli algoritmi quale fattore di condizionamento sul versante prettamente politico, cfr. B. Zarouli, T. Dobber, G. De Pauw, C. De Vreese, *Using a Personality-Profiling Algorithm Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media*, in *Communication Research*, vol. 49, n. 8, 2022, pp. 1066 ss. Lungimiranti sono le osservazioni di B. Chul Han, *Infocrazia. Le nostre vite manipolate dalla rete*, Einaudi, 2024. L'autore, nel proporre una differenza tra regime disciplinare e regime dell'informazione, osserva come nel primo caso dominare è l'isolamento, mentre nel secondo è la comunicazione, nell'ambito della quale «la sorveglianza avviene attraverso i dati», cit. p. 4. Peraltro, è doveroso ricordare i più lampanti esempi in cui si è riscontrata un'indebita influenza dei social media e delle piattaforme su votazioni deliberative ed elettive come quelle del referendum Brexit, dello scandalo *Cambridge Analytica* e, da ultimo, le presidenziali americane e rumene del 2024. Nel caso americano, il candidato repubblicano Donald J. Trump ha fatto grande uso della piattaforma X (ex Twitter) di Elon Musk, poi entrato a far parte della

Com'è noto, la legittimazione di un sistema democratico si fonda essenzialmente sulla presenza di appuntamenti elettorali liberi, competitivi e periodici, volti all'attribuzione di cariche all'interno degli apparati rappresentativi dello Stato. Tale impostazione configura un modello di democrazia rappresentativa, in cui i cittadini, pur mantenendo formalmente la titolarità della sovranità, esercitano il loro potere principalmente attraverso il meccanismo elettorale e, solo in circostanze specifiche previste dall'ordinamento, mediante strumenti di democrazia diretta. In questo quadro istituzionale, ai cittadini viene garantita la facoltà decisionale, che presuppone un adeguato livello di libertà affinché sia possibile una scelta consapevole tra diverse alternative politiche.

In questa prospettiva, perciò, risulta cruciale la questione della trasparenza del mercato politico, elemento che influenza in modo determinante sia la forma di governo sia la struttura costituzionale degli Stati. Difatti, le implicazioni di una *sregolata*³ vicinanza tra partiti e piattaforme digitali hanno prodotto

squadra governativa. Sulle elezioni americane si v. G. Caravale, *Dall'assalto a Capitol Hill alla riconquista della Casa Bianca: la definizione dei caratteri della nuova presidenza passa anche dalla Corte suprema*, in *Nomos – Le attualità nel diritto*, n. 3, 2024; F. Clementi, *Tra dilemmi e opportunità: la vittoria di Donald Trump e le sfide per la democrazia negli Stati Uniti*, in *federalismi.it*, n. 28, 2024; mentre sul caso *Cambridge Analytica* P. Norris, R. Inglehart, *Cultural Backlash, Trump, Brexit and Authoritarian Populism*, Cambridge, 2019; E. Assante, *Cosa ci può insegnare il caso Cambridge analytica*, in *federalismi.it*, Editoriale, 25 aprile 2018. Sull'impatto delle piattaforme sulle elezioni italiane G. Ziccardi, *L'uso dei social network in politica tra alterazione degli equilibri democratici, disinformazione, propaganda e dittatura dell'algoritmo: alcune considerazioni informatico-giuridiche*, in *Ragion pratica*, n. 1/2020, p. 51 ss. Non ci si può esimere dal far riferimento all'intervento della Corte costituzionale rumena. Dopo il primo turno del 24 novembre, si è tenuto il ballottaggio tra i due candidati più votati (Calin Georgescu ed Elena Lasconi). Tuttavia, nonostante la precedente conferma con decisione n. 31/2024 del 2 dicembre, la Corte ha poi deciso di annullare il procedimento elettorale sulla base di nuovi elementi emersi. La Corte ha rinviato a data da destinarsi il secondo turno delle elezioni e ha stabilito l'applicabilità dell'articolo 83 paragrafo 2 della Costituzione, che prevede che il Presidente in carica continui a esercitare il mandato fino al giuramento del nuovo eletto. La decisione della Corte costituzionale mette in luce la crisi democratica che sta attraversando la Romania, influenzata sia dal contesto geopolitico attuale che dalle dinamiche politico-istituzionali interne. Sebbene le motivazioni della Corte sulla necessità di garantire correttezza e veridicità nel processo elettorale siano condivisibili, suscita perplessità il fatto che l'intervento è avvenuto all'ultimo momento, favorendo oggettivamente il Primo ministro in carica Ion-Marcel Ciolacu, presidente del partito socialdemocratico, che era arrivato terzo nelle votazioni del 24 novembre. Sul punto si v. A.L. Solea, *Why Romania's election was annulled – and what happens next?*, in *The Conversation*, 16 dicembre 2024; S. Sassi, A. Sterpa, *La Corte costituzionale della Romania difende la democrazia liberale dalla disinformazione. Prime note sulla sentenza n. 32 del 6 dicembre 2024*, in *federalismi.it*, n. 4, 2025. Si v. anche F. Rosa, *L'annullamento delle elezioni presidenziali rumene e la difficile difesa della democrazia*, in *federalismi.it*.

3. Cfr. F. Lanchester, *Il problema del partito politico: regolare gli sregolati*, in *Quaderni costituzionali*, n. 3, 1988, p. 437 ss.

effetti sotto svariati profili, quali, principalmente, quelli inerenti alla garanzia di una formazione corretta e consapevole delle volontà elettorali da parte dei cittadini. Se da un lato, l'avvicinamento elettore-eletti si riflette su un continuo incremento dell'astensionismo⁴, dall'altro la propaganda elettorale⁵ tramite social network – e le conseguenti implicazioni derivanti dal funzionamento algoritmico – rischia di provocare storture che inevitabilmente si ripercuotono sul funzionamento del circuito democratico.

Sebbene le strategie di controllo e condizionamento dell'opinione pubblica non costituiscano un fenomeno inedito – con chiaro riferimento ai regimi totalitari della prima metà del XX secolo –, le criticità contemporanee risiedono nell'applicazione massiva di tecniche digitali per l'analisi dettagliata dei profili degli elettori, combinate con pratiche di marketing commerciale particolarmente invasive.

La tendenza all'assimilazione tra corpo elettorale e bacino di consumatori⁶, fondata sull'esigenza di prevedere le loro inclinazioni e di persuaderli circa la validità delle informazioni e dei concetti veicolati attraverso le piattaforme digitali, oltre a generare una rappresentazione distorta della realtà e a provocare uno smarrimento nella sfera politica – poiché l'elettore viene privato della possibilità di autodeterminarsi in modo consapevole –, comporta una deforma-

4. Sul punto si veda L. Ferrajoli, *Poteri selvaggi*, Laterza, 2011; sul rapporto tra astensionismo e digitale A. Mazzola, *L'effettività del suffragio tra astensionismo e nuove modalità di espressione alla luce dei principi costituzionali*, in *Consulta online*, n. 3, 2023, pp. 991-1008; A. Calafato, *Il voto digitale: tra assenteismo parlamentare e astensionismo popolare: problemi e prospettive*, in *Diritti fondamentali*, n. 1, 2024, pp. 227-253.

5. F. Lanchester, *Propaganda elettorale*, in *Enc. dir.*, vol. XXXVII, p. 126 ss.; Id., *La propaganda elettorale (e referendaria) in Italia tra continuità sregolata e difficile rinnovamento*, in *Quad. cost.*, 1996, fasc. 3, p. 383 ss. Si reputa necessario operare una distinzione concettuale, muovendo dal presupposto che, nel tempo presente, la propaganda non è più circoscritta agli appuntamenti elettorali, ma è divenuta una pratica quotidiana. Si può parlare di campagna e propaganda permanenti. Tale distinzione concettuale attiene al concetto di comunicazione, dal momento che può essere sia politica sia elettorale. La prima trova fondamento nella libertà di manifestazione del pensiero, nel principio democratico pluralista e nel modello (proprio del costituzionalismo americano) del "libero mercato delle idee", poggiandosi su una cornice regolatoria minimale. La comunicazione elettorale, invece, si configura come una sottocategoria caratterizzata da una perimetrazione normativa più stringente, manifestandosi con particolare intensità in prossimità delle consultazioni elettorali. La propaganda, lungi dall'essere circoscritta a momenti elettorali, configura ormai una pratica quotidiana di ri-negoziazione permanente del consenso, ove i confini tra informazione, persuasione e partecipazione divengono sempre più fluidi. Si v. altresì R. Borrello, *Par condicio e radiotelevisione*, vol. 1: *Introduzione alla tematica, analisi dei principali ordinamenti europei*, Giappichelli, 2007.

6. G. Gori, *Social media ed elezioni. I limiti del diritto e il rischio di una modulated democracy*, in *Informatica e diritto*, n. 1-2/2017; M. Cuniberti, *Il contrasto alla disinformazione in rete tra logiche del mercato e (vecchie e nuove) velleità di controllo*, in *MediaLaws*, 2017.

zione sostanziale della partecipazione alla vita politica e, aspetto di maggiore gravità, una compromissione dell'autenticità del suffragio⁷.

Stante ciò, l'Unione europea ha cercato di arginare tale fenomeno attraverso strumenti normativi di diversa natura, da ultimo il Regolamento (UE) 2024/900 relativo alla trasparenza e al targeting della pubblicità politica (Ttpa).

Il presente contributo si propone di offrire una prima risposta a una serie di quesiti:

1. Quali strumenti normativi sono necessari per tutelare l'integrità del processo democratico nell'era digitale?
2. In che misura il Regolamento UE 2024/900 riuscirà effettivamente a contrastare la disinformazione e le interferenze esterne nei processi elettorali europei?
3. In che misura il nuovo quadro normativo europeo influenzerà la qualità del dibattito democratico e la partecipazione elettorale consapevole dei cittadini?
4. Quale equilibrio ottimale è possibile raggiungere tra libertà di espressione politica e necessità di trasparenza nella comunicazione digitale nell'ambito del nuovo regolamento europeo?

Alla luce di ciò, nelle considerazioni che seguiranno verrà ricostruito l'impianto normativo europeo (par. 2), anche alla luce del regolamento 900/2024 di recente approvazione (par. 3). Infine, dopo aver posto in evidenza alcuni timori paventati dai parlamenti nazionali di Francia, Italia e Germania (par. 4) si cercherà di offrire alcune riflessioni di sintesi (par. 5).

2. La comunicazione politica online nel contesto euro-unionale

Sia negli Stati Uniti che in Europa, la regolamentazione dell'ecosistema digitale⁸ è stata caratterizzata, in una prima fase storica, da un approccio di tipo prevalentemente liberistico. La preoccupazione principale consisteva nel pro-

7. Vi è chi ha posto l'attenzione sul «l'ingresso del potere economico nel cyberspazio, che quelle libertà ha finito per cannibalizzare e soprattutto plasmare secondo i propri fini». M. Betzu, *I baroni del digitale*, Editoriale scientifica, 2022, p. 22.

8. Sul c.d. costituzionalismo digitale la letteratura è divenuta sconfinata. A ogni modo *ex multis* si v. O. Pollicino, *Judicial protection of fundamental rights on the Internet. A road towards digital constitutionalism?*, Bloomsbury, 2021; G. De Gregorio, *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, in Cambridge University Press, 2022; M. Betzu, *I baroni del digitale*, cit. Si v. altresì M.C. Girardi, *Libertà e limiti della comunicazione nello spazio pubblico digitale*, in *federalismi.it*, n. 17, 2024.

muovere la libera circolazione e lo sviluppo dei servizi della società dell'informazione, rimuovendo gli ostacoli alla crescita del commercio elettronico.

È possibile riscontrare questo approccio regolatorio tanto nel contesto statunitense⁹ quanto in quello europeo. Purtuttavia, nell'alveo euro-unionale, l'e-

9. Il riferimento va ovviamente alla Sezione 230 del Titolo 47 dello *U.S. Code* (*Communication Decency Act*), che ha stabilito l'esonero totale di responsabilità per i fornitori di servizi Internet relativamente ai contenuti pubblicati da terzi. L'atto è stato approvato nel 1996 dal Congresso americano. Sul punto si v. M. Bassini, *Internet e libertà di espressione. Prospettive Costituzionali e Sovranazionali*, 2019; R. Petruso, *Le responsabilità degli intermediari della rete telematica. I modelli statunitense ed europeo a confronto*, Giappichelli, 2019, XVIII; M.R. Allegri, *Oltre la par condicio, Comunicazione politico-elettorale nei social media, fra diritto e autodisciplina*, FrancoAngeli, 2020; Id. *Il futuro digitale dell'Unione europea: nuove categorie di intermediari digitali, nuove forme di responsabilità*, in *Rivista italiana di informatica e diritto*, n. 3, 2021, pp. 7-23; Si v. anche G. Pitruzzella, *La libertà di informazione nell'era di Internet*, in *MediaLaws*, 2018; G.E. Vigevani, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritto Costituzionale*, n. 1, 2023, pp. 41-54; L. Fabiano, *Content moderation e free speech clause: il controverso rapporto fra libertà e responsabilità delle piattaforme digitali nella più recente giurisprudenza della Corte suprema federale USA*, in *federalismi.it*, n. 27, 2024. Per una ricostruzione nel quadro statunitense si v. C.R. Carlson, *Let the Sunshine In: Crafting Constitutional Transparency Regulations for Content Moderation*, in *Seattle U. L. Rev.*, vol. 48, n. 3, 2025. Vi è però da osservare che in tempi più recenti il problema della moderazione dei contenuti "*Lawful but Awful*" online – cioè, contenuti offensivi o moralmente "ripugnanti" – si è riproposto in maniera poderosa, benché trovi comunque protezione nel Primo Emendamento. A tal proposito, alcuni autori hanno identificato tre possibili approcci per governare questo tipo di contenuti: a) il controllo delle piattaforme, ove le piattaforme private decidono le proprie regole, ma questo concentra troppo potere in poche aziende; b) il controllo governativo (come avviene nel caso di Texas e Florida), ove lo stato impone regole alle piattaforme, ma questo crea problemi costituzionali e pratici enormi; c) il controllo degli utenti (cd. approccio "*middleware*"), ove gli utenti possono scegliere servizi di moderazione personalizzati, mantenendo la possibilità di comunicare tra loro. Si è ritenuto che proprio il terzo approccio sia il migliore perché riduce il potere delle piattaforme senza creare censura governativa, dando agli utenti la libertà di scegliere che tipo di contenuti vedere online. Sul punto si v. D. Keller, *Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users*, in *The University of Chicago Law Review Online*, 2022. Si v. anche J.M. Balkin, *Free Speech versus the First Amendment*, in *UCLA Law Review*, vol. 70, n. 5, 2023, pp. 1206-1273. Quest'ultimo analizza come l'era digitale abbia trasformato la regolamentazione e la protezione della libertà di espressione, evidenziando il crescente divario tra i valori che giustificano la libertà di parola e ciò che il *First Amendment* effettivamente protegge. In particolar modo, proprio l'era digitale, avendo introdotto un modello pluralista di regolamentazione del discorso – ove non solo i governi territoriali, ma anche le aziende private che possiedono infrastrutture digitali giocano un ruolo centrale – ha reso il Primo emendamento meno rilevante rispetto alla governance privata del discorso. Balkin sostiene che lo stesso, pur essendo fondamentale, non è però sufficiente per proteggere i valori della libertà di parola nell'era digitale. Egli sottolinea che la governance privata e la società algoritmica abbiano trasformato il discorso in un fenomeno misurabile e manipolabile, creando un divario crescente tra i valori della libertà di parola e ciò che il *First Amendment* tutela. Per affrontare queste sfide, egli propone di concentrarsi sulla riforma del capitalismo informativo e sulla rico-

voluzione della regolamentazione della comunicazione politica online ha rivelato un paradosso fondamentale che ha caratterizzato per oltre due decenni l'approccio normativo. Si è trattato del convincimento relativo alla presunta neutralità tecnologica delle piattaforme che, però, non solo si è rivelata empiricamente falsa, ma addirittura strutturalmente impossibile se applicata al contesto della comunicazione politica.

A lungo il quadro normativo di riferimento si è basato sulla Direttiva 2000/31/CE (la cosiddetta direttiva sul commercio elettronico), che ha rappresentato la pietra angolare della regolamentazione in materia¹⁰. Proprio tale Direttiva ha codificato il principio di "irresponsabilità degli intermediari digitali", presupposto teorico oggi chiaramente insostenibile. Questo approccio, mutuato dall'ordinamento statunitense – dalla *Section 230 del Communications Decency Act* – postulava una separazione netta tra infrastruttura tecnologica e contenuti veicolati, configurando le piattaforme come meri canali o spazi neutrali per la trasmissione di contenuti.

Invero, a differenza dei mercati finanziari, dove gli attori tendono a un equilibrio razionale, le piattaforme digitali presentano meccanismi di amplificazione che hanno favorito la diffusione di contenuti emotivamente coinvolgenti piuttosto che verificati. Ciò in considerazione del fatto che gli algoritmi privilegiano l'*engagement*, non la veridicità, rendendo l'autocorrezione non solo improbabile, ma strutturalmente impossibile. In questo modo si viene a creare una sorta di "amplificatore di polarizzazione", intrinseco al modello di business delle piattaforme¹¹.

struzione delle istituzioni che promuovono la conoscenza e la democrazia. la soluzione risiede nel riformare il capitalismo informativo e le strutture industriali digitali. Si v. altresì M. Monti, *The "unity of opposites" in the regulation of social media platforms: content moderation between the EU "Digital Services Act" and the US First Amendment theories*, in *EUI Law Working Paper*, 07, 2024, pp. 1-18; Id., *Regolazione del discorso pubblico online e processi costituzionali di integrazione: una comparazione UE Usa*, Editoriale Scientifica, 2025.

10. Proprio la direttiva 2000/31/CE si basava sull'originaria impostazione di irresponsabilità degli *Internet Service Provider*, che configurava una presunta neutralità delle piattaforme digitali. Tale impostazione ha determinato una sostanziale deresponsabilizzazione rispetto ai contenuti veicolati, con particolare criticità nel contesto politico. Com'è stato opportunamente osservato: «*Constitutional democracies cannot leave private actors to acquire areas of power outside constitutional limits*». O. Pollicino, G. Di Gregorio, *Constitutional Law in the Algorithmic Society*, in H.-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor, G. De Gregorio (eds.), *Constitutional challenges in the Algorithmic Society*, Cambridge University Press, 2021, cit., p. 24.

11. J. Burrell, Z. Tufekci, *Seeing with Algorithms: How Data Science and Machine Learning Shape and Limit Human Understanding*, in *Communication Studies*, vol. 70, n. 3, 2019, pp. 270-287. Questo meccanismo genera una contraddizione strutturale: mentre l'obiettivo economico consiste nel massimizzare il tempo di permanenza e l'interazione degli utenti, l'interesse

La presunta neutralità tecnologica ha generato quella che può essere definita una “zona grigia” normativa, in cui le decisioni algoritmi che – che determinano quali contenuti politici debbano raggiungere specifici cittadini – sfuggono tanto al controllo democratico quanto ai meccanismi di responsabilità¹². Questa lacuna regolativa è diventata particolarmente evidente durante le elezioni del 2016 e del 2018, quando attori esterni all’Unione europea hanno potuto manipolare i processi elettorali nazionali sfruttando proprio questa assenza di governance. Tale “zona grigia” evidenzia la necessità di ripensare le categorie classiche del costituzionalismo contemporaneo lungo tre direttrici principali. In *primis*, il principio di territorialità viene messo in discussione da piattaforme che operano su scala globale, rendendo necessario un ripensamento delle modalità di esercizio dell’autorità pubblica oltre i confini nazionali. In secondo luogo, la distinzione tradizionale tra sfera pubblica e privata risulta inadeguata quando soggetti privati esercitano funzioni di controllo sui flussi informativi che incidono sui processi democratici. In terzo luogo, il bilanciamento tra diritti fondamentali richiede nuovi strumenti interpretativi che tengano conto delle specificità dell’ambiente digitale, dove la velocità e la pervasività dei flussi comunicativi alterano i tradizionali equilibri tra libertà di espressione e protezione dei processi democratici.

L’oramai celebre caso *Cambridge Analytica*¹³ rappresenta l’esempio lampante di questo fallimento sistemico: la possibilità di utilizzare dati personali per micro-targeting politico, senza trasparenza né controllo, ha dimostrato come la neutralità tecnologica sia, in realtà, una forma di governance privatizzata che trasferisce alle *corporations* decisioni fondamentali. Proprio questo caso ha messo in luce una problematica di rilevante spessore, ossia l’emersione

pubblico richiede un’informazione politica accurata, equilibrata e che conduce a scelte democratiche consapevoli. Il microtargeting politico rappresenta l’applicazione più problematica di questo modello economico alla sfera politica, trasformando i dati personali in materia prima per la manipolazione elettorale e le opinioni politiche in prodotti di consumo. Proprio il caso *Cambridge Analytica* risulta essere emblematico, dimostrando come la presunta neutralità tecnologica sia, in realtà, una forma di governance privatizzata che trasferisce alle *corporations* decisioni fondamentali per il funzionamento democratico, compromettendo il diritto all’informazione corretta e l’autonomia decisionale degli elettori. La sfida consiste nell’evitare che l’agorà democratica si trasformi in un mercato dove le idee politiche vengono commercializzate come prodotti di consumo attraverso sofisticate tecniche di persuasione digitale.

12. La neutralità supposta dal modello economico classico ne risulta completamente sovvertita quando viene inserita e applicata a un ecosistema digitale, giacché in quest’ultimo esistono – e agiscono – forti incentivi economici alla proliferazione della disinformazione, rendendo l’autoregolazione un meccanismo intrinsecamente fallace.

13. Oltre ai saggi citati nella nota 7, si v. P. Day, *Cambridge Analytica and Voter Privacy*, in *Georgetown Law Technology Review*, vol. 4, n. 2, 2020; P. Wagner, *Data Privacy – The Ethical, Sociological, and Philosophical Effects of Cambridge Analytica*, in *SSRN Electronic Journal*, 2021.

di un potere privato che incide direttamente sui processi democratici senza essere sottoposto ai tradizionali meccanismi di controllo. Le piattaforme digitali si sono così trasformate in attori con capacità di *gatekeeping* informativo che richiede forme di limitazione dell'esercizio del potere privato analoghe a quelle tradizionalmente riservate ai soggetti pubblici.

Questo fenomeno impone una riflessione, poiché la tradizionale distinzione tra sfera pubblica e privata risulta inadeguata quando soggetti formalmente privati esercitano funzioni che incidono sull'essenza stessa dei processi democratici.

La contraddizione sistemica ha così spinto le istituzioni europee verso un approccio inizialmente basato sull'autoregolazione, che si è rivelato altrettanto inadeguato. Sicché, solo negli ultimi anni si è registrato un cambio di passo nell'alveo dell'Unione europea¹⁴. Difatti, si è inaugurata una fase di rapida espansione normativa¹⁵, che ha finito per generare una complessa stratificazione legislativa di difficile catalogazione e consultazione per quanto attiene al settore digitale. Non solo, questo incremento quantitativo ha comportato una crescente difficoltà di coordinamento tra le diverse disposizioni, creando potenziali problemi di reciproca interferenza¹⁶.

Se si volesse tentare di offrire una prima sistematizzazione, tre sarebbero le fasi che sarebbe possibile enucleare:

1. autoregolazione tramite atti di *soft law* (2018);
2. approccio interventista (2019-2020);
3. attuale regolamentazione vincolante.

14. Le motivazioni che hanno spinto Stati Uniti e Unione europea ad assumere un approccio diverso, che rispondono a ragioni di tipo storico, economico e politico, sono richiamate in L. Fabiano, *Intelligenza artificiale e decisione algoritmica nelle esperienze europea e statunitense: differenze culturali e geopolitiche e divergenze nelle soluzioni giuridiche*, in L. Cassetti, F. Fabrizzi, A. Morrone, F. Savastano, A. Sterpa (a cura di), *Studi in memoria di Beniamino Caravita*, Editoriale Scientifica, 2024, p. 1395 ss., spec. 1400-1401. Sul cambio di passo dell'Unione europea si v. C. Pinelli, *L'evoluzione della normativa dell'Unione europea*, in C. Pinelli, U. Ruffolo (a cura di), *I diritti nelle piattaforme*, Giappichelli, 2023, p. 13 ss. Sulla proposta di regolamento europeo si v. D. Messina, *La proposta di regolamento europeo in materia di Intelligenza Artificiale: verso una "discutibile" tutela individuale di tipo consumer-centric nella società dominata dal "pensiero artificiale"*, in *MediaLaws*, 2022, pp. 204-221.

15. È stato però osservato che il passaggio da un approccio ordo-liberale a un approccio particolarmente dettagliato abbia causato una eccessiva ipertrofia normativa. Sul punto si v. A. Lucarelli, *Nuovi mezzi di comunicazione, gli assetti imprenditoriali e le soggettività politiche*, in *Rivista AIC*, n. 3, 2025.

16. Sul punto si v. E.A. Rossi, *Problemi di coordinamento normativo nel contrasto alla disinformazione politica e commerciale*, in *federalismi.it*, che osserva come tale situazione finisce con incrementare il rischio di sovrapposizioni normative contraddizioni applicative e difformità interpretative tra i vari atti legislativi del settore.

La prima fase¹⁷ si è basata sull'applicazione acritica della teoria dell'autocorrezione dei mercati al contesto della disinformazione online, che però rappresenta un fraintendimento delle dinamiche comunicative digitali. Nel 2018, la Commissione Europea ha deciso di importare un modello economico statunitense in un contesto valoriale europeo profondamente diverso. Questa trasposizione meccanica ha generato un primo codice di condotta contro la disinformazione¹⁸. Tale strumento ha invitato le varie piattaforme di social media ad assicurare livelli superiori di trasparenza riguardo alla provenienza dei contenuti e ad adottare misure efficaci per contrastare la diffusione di notizie false¹⁹.

17. Sul problema della regolamentazione o dell'autoregolazione vi sono state molteplici posizioni. Sul punto si v. le osservazioni di G. Pitruzzella, *La libertà di informazione nell'era di internet*, cit. Secondo l'autore, che analizza il controllo delle informazioni online, esistono due principali approcci: da un lato l'autoregolazione, dove le piattaforme come Google e Facebook gestiscono autonomamente il filtraggio dei contenuti (fake news, hate speech, contenuti terroristici); dall'altro l'eteroregolazione, una regolazione pubblica che renderebbe le piattaforme responsabili dei contenuti ospitati, sia equiparandole agli editori tradizionali sia imponendo obblighi di rimozione con relative sanzioni. L'autore evidenzia che, nonostante sembrano approcci contrapposti, entrambi finiscono per delegare a compagnie private il controllo effettivo sulle informazioni che circolano online. Si v. anche G. De Minico, *Internet regole e anarchia*, Editoriale Scientifica, 2012; L. Scaffardi, *Internet fra auto-limitazione e controllo pubblico*, in *Rivista AIC*, n. 4/2023, p. 355 ss.

18. Inizialmente, questo tipo di approccio è sembrato essere il più adatto, anche in considerazione delle competenze storicamente limitate dell'UE in materia di processi democratici nazionali. Così, attraverso la comunicazione della Commissione europea del 2018 prima, e il Codice di autoregolamentazione sulla disinformazione poi, si è sottolineata la necessità di regolamentare la pubblicità politica, sollecitando le piattaforme a implementare meccanismi di trasparenza, pur mantenendo un carattere di adesione volontaria. Come è stato osservato da autorevole dottrina, in un sistema basato sul libero mercato risulta «eccezionalmente difficile regolamentare la manipolazione in quanto tale», R. Sunstein, *Fifty Shades of Manipulation*, in *Marketing Behav.*, 2016. Si v. altresì O. Pollicino, *I Codici di Condotta tra self-regulation e hard law: esiste davvero una terza via per la regolazione Digitale? Il caso della Strategia europea contro la disinformazione online*, in *Rivista Trimestrale di Diritto Pubblico*, 4, 2022, p. 2 ss.; M. Monti, *Il Code of Practice on Disinformation dell'UE: tentativi in fieri di contrasto alle fake news*, *MediaLaws*, n. 1, 2019; Id., *Lo strengthened Code of Practice on Disinformation: un'altra pietra della nuova fortezza digitale europea?*, *MediaLaws*, n. 2, 2022.

19. P. Bonini, *L'autoregolamentazione dei principali Social Network. Una prima ricognizione delle regole sui contenuti politici*, in *federalismi.it*, n. 11, 2020; M. Mori, *La disinformazione online, la crisi del rapporto pubblico esperti e il rischio della privatizzazione della censura nelle azioni dell'Unione europea (Code of practice on disinformation)*, in *federalismi.it*, n. 11, 2020. È stato osservato come l'Unione Europea e le piattaforme online abbiano definito e implementato i concetti di "pubblicità politica" e "trasparenza" nel contesto delle iniziative politiche e normative per regolamentare le piattaforme digitali. Nel confrontare le politiche dell'UE, come il Codice di Condotta sulla Disinformazione e il Regolamento sulla Trasparenza e il Targeting della Pubblicità Politica, con le pratiche di piattaforme tecnologiche come Google, Meta, Microsoft, TikTok, Twitter/X, Telegram e Mastodon, alcuni autori hanno osservato come la trasparenza,

La svolta si è registrata nel biennio 2019-2020, quando la capillare diffusione di campagne di disinformazione e di manipolazione dei processi elettorali, portate avanti da attori esterni all'Unione europea, ha indotto le istituzioni euro-unionali a ripensare – finalmente – gli strumenti regolatori e le strategie di contrasto²⁰. Nel frattempo, la pandemia da Covid-19²¹ ha svolto il ruolo di catalizzatore, intensificando l'uso di piattaforme digitali e la diffusione di narrazioni fuorvianti. Questa proliferazione di attività disinformative ha segnato l'avvio della seconda fase nell'evoluzione della disciplina dell'agorà digitale europea, caratterizzata da un approccio più interventista da parte delle istituzioni. Cosicché l'Unione ha risposto con diversi strumenti: particolarmente significativa è stata l'implementazione del sistema di allerta precoce (*Rapid Alert System* – Ras), istituito in conformità al Piano d'azione contro la disinformazione, finalizzato a ottimizzare lo scambio informativo (attraverso tecniche di micro-targeting) tra l'Unione e gli Stati membri, nonché a contrastare efficacemente i fenomeni disinformativi. Inoltre, vi è anche il Regolamento 2015/2020²², che si proponeva di imporre ai provider di servizi internet un trattamento neutrale del traffico digitale, indipendentemente dalla fonte, dai contenuti, dalle applicazioni e dalle apparecchiature utilizzate.

L'evoluzione normativa successiva ha progressivamente introdotto forme di controllo e moderazione dei contenuti, mediante un complesso bilanciamento tra esigenze di libertà espressiva e necessità di arginare comportamenti illeciti. Così, gli interventi più recenti si sono concentrati, invece, su iniziative

così come implementata, sia insufficiente per affrontare le sfide democratiche poste dalla pubblicità politica online. Sarebbe necessario un approccio più ampio che includa la responsabilità delle piattaforme e degli attori politici, oltre a una maggiore regolamentazione delle pratiche di moderazione dei contenuti. Sul punto si v. T. Meyer, A. Vetulani-Cegiel, *Transparency as an empty signifier? Assessing transparency in EU and platform initiatives on online political advertising and actors*, in *Policy&Internet*, 4 settembre 2024.

20. È stato osservato come, nonostante le profonde differenze strutturali e funzionali sussistenti tra sistemi presidenziali e parlamentari, si stia verificando una convergenza significativa nelle dinamiche comportamentali delle formazioni politiche. In particolare, si è osservato un processo di "americanizzazione" per cui i partiti europei stanno progressivamente adottando strategie comunicative e tecniche digitali originariamente sviluppate e consolidate nel contesto del presidenzialismo statunitense. Cfr. C.J. Bennett, *Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America?*, in *International Data Privacy Law*, vol. 6, n. 4, 2016.

21. A. Vedaschi, C. Graziani, *Post-Pandemic Constitutionalism: Covid-19 as a Game-Changer for "Common Principles"?*, in *University of Pennsylvania Journal of International Law*, vol. 44, n. 4, 2023.

22. A tal proposito M. Orofino, *La declinazione della net-neutrality nel Regolamento europeo 2015/2020. Un primo passo per garantire un'Internet aperta?*, in *federalismi.it*, n. 2, 2016.

legislative mirate: la proposta di Regolamento sulla trasparenza e targeting della pubblicità politica (2021) e l'*European Media Freedom Act* (Emfa)²³. Questi strumenti hanno avuto come obiettivo l'armonizzazione, affinché fossero garantiti tanto l'equilibrio quanto la correttezza nel dibattito pubblico digitale, introducendo forme di responsabilizzazione per i distributori di contenuti politici e meccanismi di etichettatura dei contenuti.

L'obiettivo complessivo si configura come un poderoso processo di regolamentazione che mira a contemperare la libertà di espressione che, pur essendo un diritto fondamentale, non può essere assoluta e illimitata. Pertanto, si rende necessario un equo bilanciamento con l'esigenza di controllare i potenziali rischi di manipolazione informativa, in un contesto caratterizzato dall'estrema velocità e pervasività dei flussi comunicativi online.

Un messaggio pubblicitario in ambito politico dovrebbe garantire trasparenza nei contenuti, pluralismo, protezione dei dati personali e obiettività dell'informazione. È solo attraverso un'informazione politica puntuale e imparziale che i cittadini possono esercitare consapevolmente i propri diritti, compiendo scelte elettorali fondate su una comprensione non manipolata della realtà. L'obiettivo dovrebbe essere arduo, giacché si dovrebbero contemperare due esigenze apparentemente contrapposte: preservare la libertà di espressione e, al contempo, tutelare il diritto dei cittadini a un'informazione politica autentica, completa e rispettosa della loro privacy.

23. Tra i primi commenti sull'Emfa si v. Y.M. Citino, *Verso l'European Media Freedom Act*, in *MediaLaws*, n. 2, 2022; G.E. Vigevari, *Il Regolamento europeo sulla libertà dei media: un passo verso un servizio pubblico indipendente*, in *Quaderni costituzionali*, n. 3, 2024. Secondo quest'ultimo, l'intervento si colloca in un contesto preoccupante, dove la libertà e il pluralismo dei media sono sotto pressione in diversi Stati membri, con una crescente dipendenza dal potere politico e un indebolimento delle fonti di finanziamento. Il regolamento stabilisce standard minimi comuni che includono la protezione delle fonti giornalistiche, limiti alle ingerenze come intercettazioni e sequestri, e tutele per l'indipendenza editoriale, ponendo obblighi positivi agli Stati per garantire un effettivo pluralismo informativo. Si v. altresì E. Brogi, D. Borges, R. Carlini, I. Nenadic, K. Bleyersimon, J. Kermer, U. Reviglio, M. Trevisan, S. Verza, *The European Media Freedom Act: media freedom, freedom of expression and pluralism*, Technical Report, European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, 2023; P. Cesarini, G. De Gregorio, O. Pollicino, *The Media Privilege in the European Media Freedom Act*, in *MediaLaws*, 4 settembre 2023; T. Scipp, R.Ó. Fathaigh, M. Van Drunen, *Defining the "media" in Europe: pitfalls of the proposed European Media Freedom Act*, in *Media Law*, vol. 15, n. 1, 2023, pp. 39-51; I. Nenadić, E. Brogi, *Why news media need article 17 of the European Media Freedom Act*, in *Centre for Media Pluralism and Freedom*, 16 novembre 2023.

3. La comunicazione politica alla prova del nuovo regolamento europeo 2024/900

L'Unione Europea ha cercato di contrastare la disinformazione politica rafforzando le fonti esistenti e, al tempo stesso, introducendone di nuove orientate a proteggere l'integrità dei processi elettorali²⁴. Perciò, a seguito di un ampio processo consultivo che ha coinvolto i principali portatori d'interesse, la Commissione europea ha predisposto due distinte proposte regolamentari: il *Digital Services Act* e il *Digital Markets Act*²⁵. Entrambe le proposte, poi

24. Appare doveroso sottolineare come la Commissione europea abbia manifestato un chiaro interesse per la regolamentazione dell'ambiente digitale e per la tutela dei processi democratici già nel dicembre 2020, quando ha varato un ampio programma d'azione per la democrazia. Tra gli obiettivi principali figurava la promozione di consultazioni elettorali libere ed eque, con particolare attenzione alle problematiche emergenti legate alla pubblicità politica e all'uso di tecnologie innovative per la personalizzazione dei messaggi pubblicitari basata sui dati personali degli utenti. In tale contesto, la Commissione ha avviato l'iter di approvazione di una serie di provvedimenti volti a rafforzare l'integrità dei processi elettorali e a consolidare la partecipazione democratica. Tra questi si annoverano una comunicazione ufficiale, due proposte normative sui diritti elettorali attivi e passivi dei cittadini UE residenti in Stati membri diversi da quello d'origine, una proposta per la modernizzazione delle regole sul finanziamento dei partiti politici e delle fondazioni europee, nonché una proposta normativa sulla trasparenza e la personalizzazione della pubblicità politica. Sicché, la Commissione ha presentato un "pacchetto digitale", orientato a riaffermare la sovranità digitale dell'Unione Europea e a garantire una governance antropocentrica e personalista dell'innovazione, proteggendo i diritti fondamentali e promuovendo un mercato digitale equo e sicuro. I punti principali sono analizzati da M.R. Allegri, *Il futuro digitale dell'Unione europea: nuove categorie di intermediari digitali, nuove forme di responsabilità*, in *Rivista italiana di informatica e diritto*, n. 2, 2021, pp. 7-23. Si v. altresì S. Rudohradská, D. Tresčáková, *Proposals for the digital markets act and digital services act: broader considerations in context of online platforms*, in *EU and Comparative Law Issues and Challenges Series (Eclic)*, n. 5/2021, p. 495 ss.; L.D. Dąbrowski, M. Suska (eds.), *The European Union Digital Single Market. Europe's Digital Transformation*, Routledge, 2023; F. Pizzetti (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024.

25. Il *Digital Markets Act* (2022) ha introdotto il contrasto ai monopoli delle grandi piattaforme tecnologiche, la limitazione del potere dei "gate keeper" digitali e la promozione della concorrenza e innovazione nel mercato digitale. Il *Digital Services Act* (2022), invece, è orientato a tutelare gli utenti online, a introdurre obblighi di trasparenza per le piattaforme, a contrastare la disinformazione, a introdurre meccanismi di rimozione di contenuti illegali, proteggere i minori e tracciare la pubblicità online. Cfr. G.M. Ruotolo, *Le proposte di disciplina di digital services e digital markets della Commissione del 15 dicembre 2020*, in *DPCE on line*, n. 4, 2020; M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *La Rivista "Gruppo di Pisa"*, n. 2, 2021; Id., *I poteri privati nella società digitale: oligopoli e antitrust*, in *Diritto pubblico*, n. 3, 2021, pp. 739-760; G. Wagner, M. Eifert, A. Metzger, H. Schweitzer, *Taming the giants: The Dma/Dsa package*, in *Common Market Law Review*, vol. 58, 4, 2021, pp. 987-1028; O. Grandinetti, *Le piattaforme digitali come "poteri privati" e la censura online*, in *Rivista italiana di informatica e diritto*, n. 1, 2022; A. Iannotti Della Valle, *Il digital markets act e il ruolo dell'U-*

approvate, testimoniano il riconoscimento da parte dell'Unione dell'esigenza di avere un quadro normativo uniforme che potesse superare l'eterogeneità legislativa²⁶.

Con queste iniziative si è così inaugurata la terza fase della disciplina dell'agorà digitale, cui è possibile ricondurre anche il recente Regolamento (UE) 2024/900 sulla trasparenza della pubblicità politica²⁷, su cui si vuole concentrare l'attenzione.

Tutti questi provvedimenti rappresentano un punto di svolta nella governance digitale europea, spostando l'approccio da una logica di autoregolazione

nione europea verso un costituzionalismo digitale, in *Giurisprudenza costituzionale*, n. 3, 2022. Con specifico riguardo al Dsa si v. M. Broadbent, *The Digital Services Act, the Digital Markets Act, and the New Competition Tool*, *Center for strategic and international studies*, 20 novembre 2021; G. De Gregorio, *The Digital Services Act: A Paradigmatic Example of European Digital Constitutionalism*, in *Diritti Comparati*, 17 maggio 2021; M. Husovec, I. Roche Laguna, *Digital Services Act: A Short Primer*, in M. Husovec, I. Roche Laguna (eds.), *Principles of the Digital Services Act*, Oxford University Press, 2023.

26. Al pari di altre normative adottate nel periodo più recente, il Dsa ha modificato il tipo di regolamentazione applicata dall'Unione. Si interroga sulla loro aspirazione/natura federalista M. Monti, *Towards a Federal-Type Regulation of Online Public Discourse by the EU?*, in *European Public Law*, n. 4, vol. 30, 2024, pp. 441 ss. Pur avendo obiettivi ambiziosi, come la lotta ai contenuti illegali e alla disinformazione, la protezione dei diritti fondamentali e l'armonizzazione delle leggi tra gli Stati membri, il Dsa solleva numerose critiche per il suo impatto negativo sulla libertà di espressione e altri diritti fondamentali. È possibile ravvisare la presenza di definizioni vaghe e ampie, la mancanza di armonizzazione, obblighi eccessivi per le piattaforme, il ruolo eccessivo delle autorità politiche e le restrizioni strutturali alla libertà di espressione. Ne parla approfonditamente M. Atzori, *The Digital Services Act and the Freedom of Expression in the European Union: A Political Perspective*, 2024, disponibile su: papers.ssrn.com/sol3/papers.cfm?abstract_id=4887181; si v. altresì G. Schlag, *European Union's Regulating of Social Media: A Discourse Analysis of the Digital Services Act*, in *Politics and Governance*, vol. 11, n. 3, pp. 168-177. Cercando di offrire un paragone con il contesto normativo britannico, ove è di recente stato adottato l'*Online Safety Act* britannico (OSA), alcuni autori hanno osservato che anche questo strumento, similmente al Dsa europeo, pur essendo stato concepito per contrastare la disinformazione, in realtà presenta criticità analoghe nel proteggere adeguatamente il diritto alla partecipazione politica. Entrambi gli strumenti, perciò, sarebbero fallimentari nel garantire i due componenti fondamentali dell'articolo 25 del Patto Internazionale sui Diritti Civili e Politici (il diritto di voto e il diritto di candidarsi) quando minacciati dalla disinformazione online. Questo perché presentano definizioni vaghe e generiche e risultano essere eccessivamente deferenti verso le piattaforme online. Sul punto si v. E. Abrusci, *The UK Online Safety Act, the EU Digital Services Act and online disinformation: is the right to political participation adequately protected?*, in *Journal of Media Law*, vol. 16, n. 2, 2024, pp. 440-467.

27. Regolamento (UE) 2024/900 del Parlamento europeo e del Consiglio, del 13 marzo 2024, «relativo alla trasparenza e al targeting della pubblicità politica». Sul punto si v. L. Ullrich, *The European Regulation on the Transparency and Targeting of Political Advertising*, in B. Steinrötter, C. Heinze, M. Deng (eds.), *EU Platform Law. A Handbook*, Nomos, vol. I, 2025, pp. 657 ss.

a un intervento normativo strutturato e vincolante, con l'obiettivo di proteggere diritti individuali e collettivi all'interno dello spazio digitale.

In questo modo, l'evoluzione regolatoria mira a creare un ecosistema digitale più trasparente, responsabile e democratico, bilanciando la libertà di espressione con la necessità di un controllo rispetto a rischi paventati o concreti²⁸.

Nello specifico, il Regolamento (UE) 2024/900 è stato approvato, a larghissima maggioranza²⁹, oramai quasi due anni fa dal Parlamento europeo, dopo un lungo iter legislativo iniziato nel 2021³⁰.

Il provvedimento è scaturito da una profonda riflessione critica concernente i rischi derivanti dalla potenziale alterazione dei flussi informativi e dalle possibili interferenze di attori internazionali nei processi elettorali³¹. Questo perché si parte dal presupposto che esiste un innato conflitto tra il ricorso all'intelli-

28. Nell'analizzare la regolamentazione della pubblicità politica ingannevole sulle piattaforme online nell'Unione Europea (UE), alcuni autori hanno introdotto il concetto di "mercantilismo regolatorio". Questo approccio evidenzia come la percezione di vulnerabilità dell'UE di fronte a minacce ibride, come la disinformazione, abbia portato a un cambiamento nelle politiche di governance digitale, passando da un modello di autoregolamentazione a uno di co-regolamentazione con sanzioni. Difatti, è stato evidenziato come la regolamentazione della pubblicità politica online rifletta un cambiamento verso un approccio più rigoroso e centralizzato, motivato dalla percezione di vulnerabilità dell'UE. Questo approccio mira a garantire la sicurezza, la democrazia e la sovranità digitale, ma il successo dipenderà dalla capacità di far rispettare le norme alle grandi piattaforme tecnologiche. A tal proposito si v. B. Farrand, *Regulating misleading political advertising on online platforms: an example of regulatory mercantilism in digital policy*, in *Policy Studies*, vol. 45, n. 5, 2024, pp. 730-749.

29. All'esito dello scrutinio i voti a favore sono stati 470, mentre quelli contrari 50 e 105 gli astenuti. Il primo commento in dottrina che tiene conto della novità è di E. Caterina, *Verso il nuovo regolamento UE in materia di pubblicità politica: mercato delle idee o della propaganda?*, in *Quaderni costituzionali*, n. 1, 2024. Si v. anche C. Massa, *Proposta di regolamento sulla pubblicità politica nell'UE: più trasparenza e meno targeting*, in *Quaderni AISDUE*, n. 1, 2022, p. 363 ss.; M. Monti, *Towards a Federal-Type Regulation of Online Public Discourse by the EU?*, cit.; P. Villaschi, *The regulation of political targeting in the Italian and European Union legal framework*, in *Riv. dir. internaz. priv. e proc.*, n. 1, 2024, p. 136 ss.

30. Nel frattempo, la Commissione europea ha adottato il "pacchetto per la difesa della democrazia", che ha come elemento principale una proposta di direttiva che riguarda la trasparenza nelle attività di lobbying svolte per conto di paesi stranieri. Tanto il Regolamento quanto le iniziative incluse nel "pacchetto" fanno parte del più ampio Piano d'azione per la democrazia europea che la Commissione ha lanciato nel 2020. Quest'ultimo, partendo dal presupposto che la libertà e il pluralismo dei media sono essenziali per le democrazie e sono sanciti dalla Carta dei diritti fondamentali, prevede misure per promuovere elezioni libere ed eque, per rafforzare la libertà dei mezzi di informazione e per lottare contro la disinformazione. Il contenuto della comunicazione COM(2020) 790 final è consultabile on line all'indirizzo: eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52020DC0790.

31. Le iniziative, seppur allora in fase del tutto embrionale, che hanno poi portato al regolamento relativo alla pubblicità politica possono essere fatte risalire al 2018. Tuttavia, solo nel 2020 la Commissione europea ha presentato una bozza di regolamento sulla trasparenza e il tar-

genza artificiale per la profilazione e la necessità di un ambiente informativo che dovrebbe essere ispirato a un maggior grado di democraticità.

Il regolamento, perciò, rappresenta il precipitato normativo di una nuova concezione della sovranità digitale, intesa come la capacità delle istituzioni di esercitare un controllo effettivo sui processi di formazione dell'opinione pubblica. Questa concezione si fonda sul principio che la democrazia non può essere considerata un semplice aggregato di preferenze individuali, ma richiede la costruzione di uno spazio pubblico caratterizzato da trasparenza, pluralismo e accessibilità informativa.

Sebbene diversi siano gli obiettivi da perseguire, tra i principali è possibile ravvisare la volontà di implementare un meccanismo di trasparenza comunicativa che consenta ai cittadini di identificare agevolmente i contenuti di pubblicità politica digitale.

Siccome l'intelligenza artificiale, nel corso delle campagne politiche, consente di "colpire" gli utenti con estrema precisione in base ai loro dati, la stessa finisce per creare delle "bolle" informative che limitano opinioni diverse e possono amplificare la disinformazione³².

Alla luce di ciò, il nuovo Regolamento è orientato ad affiancare e integrare la legislazione in materia, completando il Dsa già in vigore da quattro anni³³. Tra i suoi obiettivi vi è anche quello di porre un freno all'incontenibile mes-

geting della pubblicità politica. Il regolamento è stato redatto nell'ambito del Piano d'azione per la democrazia della Commissione europea. Sul punto si cfr. il sito ufficiale.

32. In questo contesto, il Regolamento sulla protezione dei dati personali offre un ulteriore strumento per regolamentare e limitare le pratiche di microtargeting politico. Perciò, in ottemperanza al principio di trasparenza – cui il regolamento si ispira –, nel momento in cui i dati personali sono utilizzati per il targeting pubblicitario, i responsabili devono seguire regole specifiche: devono fornire informazioni aggiuntive accanto agli annunci politici, spiegando la logica e i parametri principali delle tecniche utilizzate, incluso se è stata impiegata l'IA per il targeting o la distribuzione dell'annuncio. Ad es. il Gdpr richiede già la trasparenza sulle modalità di trattamento dei dati personali e impone un divieto condizionato all'uso di dati sensibili e di tecnologie di profilazione. Cfr. Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati). Sul tema, *ex multis* L. Califano, C. Colapietro (a cura di), *Innovazione tecnologica e valore della persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Editoriale Scientifica, 2017; L. Scaffardi (a cura di), *I "profili" del diritto Regole, rischi e opportunità nell'era digitale*, Giappichelli. Si v. altresì il recentissimo volume di L. Califano, F. Galli, V. Fiorillo, *La protezione dei dati personali: natura, garanzie e bilanciamento di un diritto fondamentale*, Giappichelli, 2024.

33. Peraltro, è opportuno osservare come proprio nel Dsa è stato esplicitato che il dovere di diligenza sia quello di attenuare i rischi sistemici aventi effetti reali o prevedibili in relazione ai processi elettorali rientra nella logica dell'attuazione (*ex art. 44*).

saggistica personalizzata, che finora ha distorto il dibattito politico e violato i diritti fondamentali dei cittadini europei.

Secondo il Parlamento europeo e la Commissione, il Regolamento stabilirà un quadro normativo comune che migliorerà la pubblicità politica sponsorizzata, l'equità e la parità di trattamento nel processo elettorale, nonché il diritto fondamentale a un'informazione obiettiva, trasparente e pluralistica.

Come evidenziato nei considerando 9-14, l'Unione ha deciso di affrontare la frammentazione del mercato interno nel settore attraverso un intervento normativo volto all'armonizzazione. Questo perché, tale frammentazione ha generato un contesto di incertezza per gli operatori che preparano, collocano e diffondono messaggi pubblicitari di natura politica, con conseguente ostacolo alla libera circolazione dei servizi, distorsioni della concorrenza (particolarmente evidenti nell'interazione tra servizi online e offline) e incremento degli oneri di conformità. Si vuole così perseguire l'obiettivo di stabilire un livello uniforme ed elevato di trasparenza in tutto il territorio unionale, prevenendo divergenze normative che potrebbero compromettere il funzionamento del mercato interno. A tal fine, l'articolo 114 Tfe costituisce la base giuridica per l'introduzione di norme armonizzate relative alla fornitura di servizi di pubblicità politica, con particolare riferimento agli obblighi di chiarezza e di dovere di diligenza per sponsor e prestatori di servizi.

L'approccio adottato è quello dell'armonizzazione completa, che preclude agli Stati membri la facoltà di mantenere o introdurre disposizioni nazionali divergenti, siano esse più o meno rigorose. Tale approccio mira a incrementare la certezza del diritto e a ridurre la frammentazione degli obblighi gravanti sugli operatori del settore.

È significativo notare come il regolamento delimiti precisamente il proprio ambito di applicazione alle norme sulla trasparenza, agli obblighi di diligenza, all'uso di tecniche di targeting e alla vigilanza, senza interferire con il contenuto dei messaggi né con altre normative dell'Unione o nazionali che disciplinano aspetti diversi della pubblicità politica. Rimangono pertanto impregiudicate le disposizioni che regolano la conduzione e il finanziamento delle campagne politiche, i periodi di silenzio elettorale e, in particolare, i diritti fondamentali alla libertà di opinione e di espressione.

In questa prospettiva, la normativa è orientata a proteggere le persone fisiche per quanto riguarda il trattamento dei dati personali, introducendo norme specifiche sull'uso di tecniche di targeting e diffusione nel contesto della pubblicità politica. Tutte queste regole si applicheranno ai responsabili del trattamento, compresi i fornitori di servizi che fanno ricorso a tali tecniche.

In linea generale, l'architettura normativa mira a favorire un processo di consapevolizzazione dei cittadini, ponendoli nelle condizioni di acquisire

una piena cognizione dei meccanismi di veicolazione dei messaggi politico-elettorali. In particolare, l'obiettivo è duplice: da un lato rendere immediatamente identificabile la matrice generativa del messaggio pubblicitario, dall'altro consentire una valutazione critica circa la natura *targeted* del contenuto ricevuto.

Tale impianto si configura quale strumento di tutela democratica, volto a preservare l'autonomia decisionale degli elettori mediante la promozione di una sfera pubblica fondata sulla chiarezza informativa e sulla consapevolezza dei meccanismi di persuasione mediatica. L'intento ultimo è quello di porre i cittadini nelle condizioni di effettuare una scelta elettorale realmente informata e non influenzata da meccanismi occulti di condizionamento, che rischiano di snaturare la partecipazione politica e, aspetto particolarmente critico, di compromettere l'autonomia delle scelte democratiche. In tal modo, si vuole preservare l'integrità del processo democratico, consentendo all'elettorato di orientarsi consapevolmente nelle proprie decisioni, affinché il baluardo della libertà del voto³⁴, elemento imprescindibile dell'architettura democratica, non ne risulti snaturato o svuotato di significato.

La normativa si inserisce in un più ampio contesto di riflessione critica sulle nuove forme di comunicazione politica nell'ecosistema digitale, ove i confini tra informazione, persuasione e manipolazione divengono sempre più labili e complessi da decifrare³⁵.

Ad ogni modo, si tratta di un regolamento complesso, composto da 114 Considerando, 30 articoli e 5 Capi, che si pone ardui obiettivi, tra cui: il rafforzamento della trasparenza e dell'integrità dei processi digitali affinché le scelte possano essere caratterizzate da democraticità e consapevolezza; l'armonizzazione del mercato interno della pubblicità politica e la tutela dei diritti fondamentali, in *primis* vita privata e protezione dati personali. Lo stesso concerne la regolamentazione, la promozione e la disseminazione degli annunci politici, abbracciando un ampio spettro di comunicazione pubblicitaria politica – inclusa quella cosiddetta “*issue-based*” – che sia prodotta, veicolata o indirizzata

34. Sul diritto di voto, seppur oramai risalente, si v. M. Rubechi, *Il diritto di voto. Profili costituzionali e prospettive evolutive*, Giappichelli, 2016.

35. Nel frattempo, nel giugno 2022, la Commissione europea ha anche adottato un codice di condotta sulla disinformazione online. Tale codice si basa su circa quaranta impegni, principalmente orientati a contrastare il fenomeno della disinformazione, attraverso: l'incremento della trasparenza e l'implementazione di meccanismi di etichettatura della pubblicità politica, nonché il potenziamento del ruolo dei verificatori di informazioni (*fact checker*). I 43 soggetti firmatari si sono impegnati a rispettare tali principi e a sviluppare un sistema di risposta tempestiva in contesti di crisi o durante le consultazioni elettorali. È opportuno sottolineare che l'operatività di tale sistema sarà condizionata dall'implementazione delle disposizioni del Dsa, che conferirà carattere vincolante mediante l'introduzione di un apparato sanzionatorio.

all'interno dell'ecosistema europeo, prescindendo dalla localizzazione geografica del soggetto erogante.

Una scelta particolarmente significativa riguarda l'approccio extraterritoriale del Regolamento, che estende la propria applicazione non solo ai soggetti privati europei, ma anche a soggetti extraterritoriali che operano nel mercato interno europeo. Questa opzione rappresenta un'evoluzione del principio di territorialità, introducendo una logica di "effetto sul mercato interno" che giustifica l'applicazione della normativa europea indipendentemente dalla localizzazione geografica del soggetto erogante. L'approccio, già sperimentato con il Gdpr, solleva questioni complesse, ma appare necessario per contrastare fenomeni e, al tempo stesso, garantire l'effettività per quanto riguarda la tutela dei diritti. Ciò è particolarmente rilevante considerando che molte delle interferenze nei processi elettorali europei provengono da attori esterni all'Unione, come evidenziato dalle elezioni del 2016 e del 2018.

Questa dimensione extraterritoriale rivela inoltre una concezione della sovranità digitale che trascende i confini nazionali, riconoscendo che la protezione della democrazia europea richiede un approccio sistemico capace di contrastare le interferenze esterne. Tale estensione della giurisdizione si ritiene possa rappresentare un'innovazione poderosa, che potrebbe inaugurare nuovi scenari nella governance globale.

Come si legge nella relazione di accompagnamento, lo scopo del regolamento è contribuire al buon funzionamento del mercato interno della pubblicità politica con norme tra loro armonizzate. Pertanto, tali norme si applicheranno tanto ai prestatori di servizi, quanto agli editori, prevedendo specifici obblighi.

Di estremo interesse appare il perimetro definitorio che, come si legge nel primo considerando, può assumere forme svariate: contenuti a pagamento, risultati di ricerca sponsorizzati, messaggi mirati a pagamento, promozione in classifiche, promozione di qualcosa o qualcuno integrato in contenuti, come un inserimento di prodotti, influencer e altri testimonial. In questo modo, è stata introdotta una definizione armonizzata di pubblicità politica, finora inesistente, lasciando campo libero a ingerenze straniere e al suo finanziamento esclusivamente da parte di organizzazioni europee (*ex art. 5*)³⁶.

Inoltre, ne viene fornita una nuova definizione, che comprende «la preparazione, collocazione, promozione, pubblicazione, consegna o diffusio-

36. Le attività correlate possono comprendere ad esempio la diffusione di messaggi di pubblicità politica su richiesta di uno sponsor o la pubblicazione di contenuti a titolo oneroso o che prevedano altre forme di remunerazione.

ne di un messaggio» che sia in grado di influenzare l'esito di un'elezione o referendum³⁷, il processo legislativo o il comportamento di voto.

Non rientrano nell'ambito di applicazione del testo le opinioni politiche (art. 1), né le comunicazioni provenienti da autorità governative e istituzionali che assolvono a funzioni di carattere informativo e procedurale strettamente connesse ai processi democratici (art. 3). Si tratta, in particolare, di specifiche comunicazioni, quali gli avvisi concernenti le modalità di partecipazione elettorale, gli annunci di candidature e, parimenti le presentazioni dei candidati che, disciplinate da normative specifiche, garantiscono condizioni di trasparenza e parità³⁸.

Appare dunque evidente come il Regolamento miri a trasformare la pubblicità politica da potenziale strumento di manipolazione a mezzo di informazione trasparente e consapevole.

Benché l'iniziale desiderata dell'UE si concentrasse sulla disinformazione online e sulla pubblicità mirata, il regolamento non traslascia la televisione e la stampa³⁹.

La definizione di pubblicità politica non è vincolata alla modalità di distribuzione a pagamento. In altre parole, un contenuto può essere considerato tale anche se non erogato a titolo oneroso. Ciò comporta che qualsiasi comunicazione che potenzialmente possa influenzare l'orientamento di voto degli elettori rientra nella categoria della pubblicità politica, indipendentemente dal fatto che sia stata diffusa gratuitamente o a pagamento.

Un elemento degno di considerazione riguarda il fatto che il Ttp è orientato ad attuare due livelli di trasparenza. Uno riguarda i fornitori di servizi, che dovranno registrare le informazioni rilevanti, in particolare quelle relative

37. Come si legge nel considerando n. 31 per votazioni elettive si intendono «le elezioni del Parlamento europeo, tutte le elezioni o referendum organizzati a livello nazionale, regionale e locale negli Stati membri, e le elezioni per stabilire la leadership dei partiti politici». Da questa ampia definizione sono invece escluse altri tipi di elezioni, come le votazioni professionali o organizzate privatamente.

38. Si tratta di messaggi provenienti da fonti ufficiali degli Stati membri o dell'Unione strettamente riguardanti l'organizzazione e le modalità di partecipazione a elezioni o referendum, compreso l'annuncio delle candidature o dei quesiti referendari, oppure la promozione della partecipazione a elezioni o referendum; ii) la comunicazione pubblica finalizzata a fornire informazioni ufficiali al pubblico da parte, a favore o per conto di un'autorità pubblica di uno Stato membro o dell'Unione, incluso da parte, a favore o per conto di membri del governo di uno Stato membro, a condizione che non possa e che non sia intesa a influenzare l'esito di un'elezione o di un referendum, un comportamento di voto o un processo legislativo o regolamentare; e iii) la presentazione di candidati in determinati spazi pubblici o nei mezzi di comunicazione, esplicitamente prevista dalla legge, effettuata a titolo gratuito e in condizioni di parità di trattamento dei candidati.

39. Regolamento, considerando nn. 17 e 32.

all'origine del finanziamento, all'interno di uno specifico registro (art. 8) trasmesso all'editore pubblicitario (art. 10) per un'etichettatura chiara, precisa e univoca, visibile agli utenti al momento della diffusione dei messaggi (art. 11). L'altro livello spetta alla Commissione che creerà, due anni dopo l'entrata in vigore del testo, un registro europeo, aggiornato con i messaggi pubblicitari, le informazioni associate relative alle tecniche di targeting e amplificazione, dal loro utilizzo quando comportano il trattamento di dati personali per un periodo massimo di sette anni (art. 13).

La *ratio legis* risiede nel perseguimento di un modello di trasparenza comunicativa, implementato attraverso articolati meccanismi di *disclosure* obbligatori che sono rivolti agli editori. Si tratta di obblighi specifici in materia di etichettatura. Ciascun messaggio pubblicitario di natura politica deve essere accompagnato da un complesso di informazioni chiaramente visibili e "privi di ambiguità"⁴⁰. È anche prevista l'adozione, entro il 10 luglio 2025, di atti di esecuzione da parte della Commissione europea per la definizione standardizzata del formato e del modello delle etichette, con l'obiettivo di garantirne l'adequatezza rispetto ai diversi media, considerando le specificità dei mezzi di comunicazione, le evoluzioni tecnologiche e di mercato, nonché le evidenze scientifiche e le migliori pratiche in materia.

Nello specifico, ogni comunicazione pubblicitaria deve essere corredata da:

1. un'avvertenza che espliciti la natura a pagamento del contenuto;
2. una "*transparency notice*" che documenti gli elementi costitutivi della campagna, quali l'identità dello sponsor, l'entità dell'investimento economico e i criteri di targeting dei destinatari.

Inoltre, il provvedimento introduce significative limitazioni alle tecniche di profilazione e amplificazione, stabilendo vincoli stringenti come il divieto di profilazione per soggetti minorenni e l'esclusione di pratiche di micro-

40. Tra questi vi rientrano: la dichiarazione esplicita della natura pubblicitaria del messaggio, l'identificazione del soggetto sponsor e dell'eventuale entità che esercita il controllo ultimo, l'indicazione del contesto elettorale, referendario o legislativo di riferimento, una dichiarazione relativa all'eventuale impiego di tecniche di targeting, nonché un avviso o l'indicazione della sua reperibilità. La disposizione normativa non si limita a prescrivere la mera presenza di tali informazioni, ma ne garantisce la completezza e l'accuratezza, ponendo in capo agli editori la responsabilità della loro verifica. Inoltre, sono stabiliti requisiti formali per le etichette informative, che devono essere adeguate al mezzo di comunicazione utilizzato, mantenere la loro visibilità anche in caso di ulteriore diffusione del messaggio pubblicitario e facilitare l'identificazione immediata della natura politica del contenuto da parte dei destinatari. Si v. quanto disposto dagli artt. 11 e 12 del Regolamento.

targeting che non siano previamente autorizzate mediante consenso esplicito dell'utente⁴¹.

In una prospettiva di implementazione normativa – con decorrenza dal 10 aprile 2024 ma applicabilità differita al 10 ottobre 2025, a eccezione di talune disposizioni immediatamente precettive⁴² – numerosi principi sono stati già recepiti nelle Linee guida elaborate dalla Commissione europea nel marzo 2024, in preparazione delle elezioni europee dello scorso giugno⁴³.

Gli editori e le piattaforme assumono una responsabilità fondamentale nel garantire la pubblicazione delle informazioni obbligatorie per ogni annuncio pubblicitario. L'attività di controllo sul rispetto dei molteplici obblighi è stata assegnata a diverse autorità: le autorità per la protezione dei dati personali vigilano sugli aspetti relativi al targeting, anche attraverso un sistema di reclami individuali, mentre i coordinatori dei servizi digitali (in Italia rappresentati dall'Agcom, mentre per le grandi piattaforme e motori di ricerca la competenza spetta direttamente alla Commissione) controllano gli altri aspetti normativi relativi alla pubblicità online. È anche previsto che ogni Stato membro designi un'autorità indipendente responsabile per la vigilanza, eventualmente diversa da quelle già specificate dal Regolamento stesso.

È stato previsto anche un apparato sanzionatorio *una tantum* in caso di violazioni ripetute, ai sensi del Dsa, che possono ammontare fino al 6% del fatturato o del reddito annuo di un fornitore di pubblicità (art. 25), mentre altre sanzioni potrebbero essere imposte ai sensi del Gdpr.

Una notazione riguarda l'art. 27 che non esclude la possibilità di modificare il regolamento entro due anni da ciascuna elezione del Parlamento europeo. Qualora il sistema dovesse rivelarsi insufficiente, la Commissione può presentare al Parlamento europeo e al Consiglio una relazione sulla valutazione e su una possibile richiesta di riesame.

Sebbene la disciplina introdotta costituisca un valido punto di partenza per regolamentare un fenomeno in rapida evoluzione come la pubblicità politica digitale, fornendo un quadro di limiti e regole necessari, tuttavia, vi potrebbero essere margini significativi di miglioramento.

Posto che le principali linee di intervento (rafforzamento della trasparenza delle campagne elettorali, limitazione microtargeting politico, obblighi di etichettatura contenuti politici sponsorizzati e divieto profilazione basata su dati

41. Peraltro, una volta che l'utente nega il proprio consenso, non può più essere richiesto.

42. Fanno eccezione gli articoli 3, relativo alle definizioni, e 5, che disciplina le prestazioni di servizi di pubblicità politica nell'Unione.

43. V. Isaia, *I complessi contorni della nozione di pubblicità politica alla prova delle elezioni europee 2024*, in *MediaLaws*, 19 giugno 2024.

sensibili) possono essere raggiunte, rimangono talune criticità che non possono e non devono essere trascurate e che riguardano, per esempio, le difficoltà tecniche nell'uniformazione degli standard, le resistenze di alcune piattaforme digitali, nonché le complessità implementazione algoritmi di controllo.

Un limite di non poco conto riguarda l'art. 25, che lascia ai singoli Stati membri il compito di stabilire le sanzioni in caso di violazione. Come già anticipato, il regolamento fissa solo un tetto massimo del 6% delle entrate o del bilancio annuale dello sponsor o del fornitore di servizi, oppure del loro fatturato globale annuo. Questa formulazione potrebbe comportare una disparità di trattamento, laddove gli Stati membri più vulnerabili alle influenze esterne – come evidenziato dall'Osservatorio europeo dei media digitali (Edmo)⁴⁴ – potrebbero non essere in grado di imporre sanzioni sufficientemente dissuasive nei confronti di chi viola il Regolamento. A ciò si aggiunge un ulteriore rischio di propagazione a macchia d'olio, giacché non solo i soggetti esterni intenzionati a compromettere il corretto svolgimento dei processi elettorali riuscirebbero a farlo, ma addirittura non si riuscirebbe a evitare la diffusione di messaggi pubblicitari politici in tutto l'ambiente digitale dell'Unione.

4. L'armonizzazione europea della pubblicità politica digitale: i timori di alcuni Parlamenti nazionali

Nel corso dell'iter di approvazione non sono mancate osservazioni da parte di alcuni Parlamenti nazionali, che vanno oltre a considerazioni in merito alla ripartizione di competenze tra Unione europea e Stati membri. È stato infatti osservato come il Regolamento, basandosi sugli articoli 114 e 16 del Tfeue, esula dalle competenze del mercato interno legate alla concorrenza. Di conseguenza, il principio di sussidiarietà deve essere rispettato – dalla Commissione – nell'ambito delle competenze condivise.

Al di là della mera ripartizione di competenze, numerose questioni sono state sottoposte all'attenzione del legislatore europeo. In particolare, il *Bundesrat* tedesco, pur riconoscendo e condividendo la necessità di creare norme armonizzate in materia di piattaforme online, non ha mancato di evidenziare come l'eccessiva ipertrofia normativa degli ultimi anni⁴⁵ non solo possa causare ai

44. Report on Disinformation narratives during the 2023 elections in Europe, March 2024.

45. Il Consiglio federale ha richiamato tutte le proposte legislative presentate dalla Commissione negli ultimi anni, tra cui il Regolamento generale sulla protezione dei dati (Gdpr), il Regolamento ePrivacy, la Direttiva sui servizi di media audiovisivi (Avmsd), la Legge sui servizi digitali e sui mercati digitali e la governance dei dati, e il "Data Act".

destinatari un'incapacità di cogliere la visione d'insieme, ma anche che questo possa avere un impatto negativo sull'economia.

Considerando il ruolo che la pubblicità politica assume in ogni ordinamento, vale a dire essere un elemento costitutivo di ogni democrazia, il Consiglio federale ha osservato che gli obblighi derivanti dal (allora proposta) di Regolamento non dovrebbero né ostacolarla, né sottoporla a una generalizzata presunzione di illegittimità. Anche le strategie di profilazione dell'utenza derivanti dall'adattamento dei contenuti informativi a specifici gruppi demografici non costituiscono di per sé una distorsione della comunicazione politica, bensì rappresentano un elemento funzionale al processo deliberativo democratico⁴⁶.

Anche il Senato francese ha offerto considerazioni attraverso l'*Avis politique* del 17 febbraio 2022⁴⁷. Con tale documento sono state espresse diverse preoccupazioni sulla portata delle competenze dell'UE, sottolineando che, mentre l'Unione può stabilire linee guida per le elezioni europee, non ha però autorità sull'organizzazione delle elezioni nazionali e locali negli Stati membri. Inoltre, il testo ha richiamato l'attenzione sul diritto elettorale francese, che già limita rigorosamente la pubblicità politica durante i periodi di campagna elettorale, al fine di proteggere gli elettori da pressioni indebite.

A ogni modo, il Senato francese ha espresso preoccupazioni tanto per l'ambiguità di alcune definizioni e per il ruolo delle "entità interessate" che potrebbero richiedere informazioni sulle pubblicità politiche, quanto al finanziamento dei partiti politici europei: in particolar modo, rispetto a quest'ultimo punto, non pochi i timori sulla possibilità dei partiti politici europei di ricevere contributi (fino al 10% del totale) da partiti membri con sede in paesi appartenenti al Consiglio d'Europa, che potrebbe rappresentare un rischio di ingerenza straniera a detrimento della loro indipendenza.

Nel caso italiano, come osservato nella risoluzione della XIV commissione del Senato⁴⁸, il regolamento non avrebbe soddisfatto pienamente il principio di proporzionalità. Questo perché la base giuridica dell'articolo 114 del Tfue, concernente il ravvicinamento delle legislazioni nazionali in materia di merca-

46. Il testo del parere tedesco disponibile su: [www.bundesrat.de/SharedDocs/drucksachen/2021/0801-0900/826-21\(B\).pdf?__blob=publicationFile&v=1](http://www.bundesrat.de/SharedDocs/drucksachen/2021/0801-0900/826-21(B).pdf?__blob=publicationFile&v=1).

47. L'Avis è divenuto poi una risoluzione, ai sensi dell'articolo 73 quinquies, commi 2 e 3, del Regolamento del Senato. Il testo della risoluzione è disponibile su: www.senat.fr/leg/tas21-122.pdf.

48. Si v. Risoluzione della 14ª Commissione permanente (Politiche dell'Unione europea) approvata nella seduta del 13 aprile 2022 sulla proposta di regolamento del Parlamento europeo e del Consiglio relativo alla trasparenza e al targeting della pubblicità politica (COM(2021) 731), disponibile su: www.senato.it/service/PDF/PDFServer/DF/370159.pdf.

to interno, sarebbe risultata inadeguata per legittimare l'intervento dell'Unione europea nelle campagne elettorali nazionali e locali, nonché per imporre obblighi ai candidati elettorali nazionali e ai partiti politici in qualità di sponsor. Come si evince nel testo della risoluzione, la "frammentazione" normativa che si intende superare non costituirebbe tanto una problematica inerente al mercato interno, quanto piuttosto una manifestazione delle peculiarità dei sistemi istituzionali nazionali, delle tradizioni politiche e delle procedure operative caratterizzanti la vita democratica di ciascuno Stato membro. Pertanto, il Regolamento avrebbe dovuto essere circoscritto ai requisiti minimi volti a garantire l'efficiente funzionamento esclusivamente del mercato transfrontaliero dei prodotti e servizi di pubblicità politica, al fine di non interferire con lo svolgimento delle campagne elettorali nazionali e locali, né con le strategie di comunicazione politica dei candidati e delle formazioni partitiche. Inoltre, sarebbe stato preferibile che l'atto giuridico avesse assunto la veste di direttiva anziché regolamento, così da assicurare agli Stati membri un margine di discrezionalità più ampio in un ambito normativo sensibile e potenzialmente influente sulla vita democratica nazionale o locale⁴⁹.

Sia consentito offrire alcune riflessioni, prima di avviarci alle conclusioni. Ovviamente, questo Regolamento comporterà importanti implicazioni per tutti gli Stati membri, a partire dagli ampi aspetti definitori. Invero, la portata delle ripercussioni emerge lampante se si considerano le attuali modalità di approccio alla pubblicità politica, potendo operare una distinzione tra:

1. Stati membri che dispongono di una definizione consolidata di pubblicità politica e di un apparato normativo nazionale implementato da numerosi decenni (come in Germania);
2. Stati deficitari di specifiche disposizioni (come in Svezia);
3. Stati sprovvisti di regolamentazione specifica per la pubblicità diffusa;
4. Stati che disciplinano la trasparenza attraverso piattaforme digitali (come in Francia).

49. Come si legge nella relazione di accompagnamento, il regolamento è «lo strumento giuridico più appropriato per definire il quadro di trasparenza della pubblicità politica nell'Unione. La scelta [...] trova giustificazione nella necessità che sia data un'applicazione uniforme alle nuove norme, come la definizione di pubblicità politica e gli obblighi di trasparenza cui sono tenuti i prestatori di servizi di pubblicità politica quando preparano o diffondono messaggi. Godendo di applicabilità diretta a norma dell'articolo 288 Tfu, il regolamento ridurrà la frammentazione giuridica e garantirà maggiore certezza del diritto introducendo elevati standard di trasparenza per la pubblicità politica a beneficio dei pertinenti prestatori di servizi, e prevenendo l'insorgere di divergenze che ostacolano la libera prestazione dei servizi nel mercato interno».

Un aspetto critico che emerge è relativo ai rischi sui diritti fondamentali derivante dal potenziale conflitto tra diverse libertà costituzionalmente protette. Da un lato, la libertà di espressione e di impresa dei soggetti privati; dall'altro, il diritto dei cittadini a un'informazione plurale e non manipolata.

Tuttavia, come evidenziato dalle diverse osservazioni dei Parlamenti nazionali durante l'iter di approvazione, nel contesto eurounionale non esiste una visione omogenea su quale sia la migliore tutela dei diritti fondamentali nell'ecosistema digitale. Questa eterogeneità si manifesta in approcci nazionali diversificati (alcuni Stati privilegiano la libertà di espressione, altri la protezione dei processi democratici), tradizioni costituzionali differenti e sensibilità culturali specifiche. Il Regolamento tenta di mediare tra queste differenze attraverso un approccio procedurale che privilegia la trasparenza rispetto alla censura. Tuttavia, la tensione tra le diverse concezioni rimane irrisolta e potrebbe emergere nella fase di implementazione, richiedendo un costante bilanciamento.

Anche la dimensione procedurale del regolamento solleva importanti interrogativi sulla natura stessa del controllo democratico nell'era digitale. L'approccio adottato dal legislatore europeo privilegia infatti una strategia di proceduralizzazione dei comportamenti che, pur preservando formalmente la libertà di espressione, rischia di incidere sul profilo sostanziale delle questioni in gioco. Questa scelta metodologica, se da un lato evita i rischi di censura diretta, dall'altro può generare forme di controllo indiretto e comportare un'eccessiva proceduralizzazione. Proprio quest'ultima, che si manifesta attraverso una serie di strumenti, finisce per trasformare la valutazione politica in verifica di conformità amministrativa: gli obblighi di etichettatura, i registri di trasparenza, le dichiarazioni di conformità e i meccanismi di *accountability* algoritmico configurano un sistema di controllo che opera principalmente sul piano formale-procedurale. Questo approccio, anche se garantisce una maggiore trasparenza, rischia di spostare il focus dalla qualità sostanziale del dibattito democratico alla mera conformità tecnica delle procedure. La questione assume particolare rilevanza se considerata alla luce del rapporto tra costituzionalismo e governance digitale. Il costituzionalismo contemporaneo, tradizionalmente orientato alla limitazione del potere pubblico in senso verticale, deve confrontarsi con l'emergere di poteri privati digitali che esercitano forme di controllo orizzontale sui processi democratici. Le piattaforme digitali, attraverso i loro algoritmi e le loro politiche di moderazione, esercitano un potere che incide profondamente sulla formazione dell'opinione pubblica e sui processi democratici.

A ogni modo, nonostante alcuni dubbi e talune problematicità, si vuole offrire una prospettiva di ottimismo rispetto ad alcuni elementi: è stimato

che dall'entrata a regime del regolamento si possa determinare una riduzione dei contenuti disinformativi stimata intorno al 30%, una maggiore consapevolezza elettorale e un parziale contenimento delle interferenze esterne. Tale impatto positivo si estenderà presumibilmente anche alla qualità complessiva del dibattito democratico, rafforzando l'integrità dei processi elettorali e contribuendo all'armonizzazione delle discipline comunicative all'interno dello spazio pubblico europeo, con possibili effetti benefici sulla legittimazione delle istituzioni euro-unionali e su una partecipazione maggiormente consapevole dei cittadini.

5. La democrazia nell'era algoritmica: tra sfide e prospettive

L'analisi sin qui condotta ha cercato di porre in luce quanto la rivoluzione digitale abbia profondamente trasformato il paradigma della comunicazione politica, generando sfide inedite per la tenuta del sistema democratico. Questo perché, come è stato sostenuto, il fenomeno digitale non può essere circoscritto, ma finisce col ricomprendere una molteplicità di sfaccettature⁵⁰.

Le campagne elettorali sui social media possono compromettere il diritto all'informazione corretta, presupposto fondamentale per l'esercizio del diritto di voto. Parimenti, l'incidenza di tali strumenti non si esaurisce nell'ambito dei diritti individuali, ma determina anche una progressiva erosione dei principi fondamentali su cui si basa la democrazia rappresentativa. Questo aspetto è particolarmente evidente se si considera la trasformazione della funzione di intermediazione dei partiti nell'era della comunicazione digitale⁵¹.

L'Unione europea sembrerebbe porre un'attenzione diversa sulla tematica. Consapevole delle svariate e importanti implicazioni derivanti dall'assenza di una regolamentazione, l'UE ha promosso un percorso in cui l'interventismo istituzionale ha seguito una crescita esponenziale nel corso degli ultimi anni. Un percorso che vede nel Regolamento UE 2024/900 l'ultimo importante tassello nella governance della comunicazione politica digitale, pur rivelando lacune strutturali che potrebbero comprometterne l'efficacia complessiva. Nonostante ciò, il suo principale merito risiede nell'aver finalmente introdotto principi di trasparenza vincolanti e armonizzati per i Paesi membri, colmando un vuoto normativo che per anni ha esposto i processi democratici a forme

50. C. Pinelli, *Disinformazione, comunità virtuali e democrazia: un inquadramento costituzionale*, in *Diritto pubblico*, n. 1, 2022, p. 173 ss.

51. A. Papa, *Disintermediazione dell'informazione e della comunicazione politica nelle democrazie digitali*, in *federalismi.it*, n. 15, 2025.

s sofisticate di manipolazione algoritmica. Alcune delle misure previste rappresentano una conquista di rilievo, giacché capaci di restituire ai cittadini quel *quantum* di consapevolezza riguardo ai meccanismi di persuasione cui sono sottoposti.

Particolarmente significativo, in questa prospettiva, è il cambio di paradigma operato dall'Unione. Per anni, la regolamentazione della comunicazione politica digitale è stata affidata prevalentemente a strumenti di *soft law*: codici di condotta, linee guida e meccanismi di autoregolamentazione. Questo approccio, pur rispettoso della libertà imprenditoriale e della natura transnazionale di internet, si è rivelato ampiamente insufficiente di fronte alla rapidità e pervasività dei fenomeni manipolativi. L'Unione ha così abbandonato la logica del *nudging* per abbracciare uno strumento capace di imporsi in modo uniforme su tutto il territorio europeo. Tuttavia, proprio la scelta dello strumento normativo del regolamento solleva interrogativi non trascurabili circa la sua effettiva capacità di incidere sulla realtà. La diretta applicabilità, se da un lato garantisce uniformità, dall'altro si scontra con la natura intrinsecamente dinamica e sfuggente dell'ecosistema digitale. Un regolamento, per sua natura, presenta una rigidità strutturale che mal si concilia con il contesto tecnologico in continua e rapidissima evoluzione. Ciò che oggi costituisce una tecnica di *microtargeting* manipolativo domani potrebbe essere sostituita da strumenti non contemplati dal testo normativo. A ciò si deve aggiungere anche che la vincolatività formale non garantisce automaticamente l'effettività sostanziale, soprattutto quando i destinatari delle norme sono colossi tecnologici dotati di risorse immense e operanti su scala globale. Il rischio concreto è che le sanzioni previste, per quanto significative sulla carta, si rivelino insufficienti rispetto ai profitti derivanti da pratiche illecite.

Tali criticità emergono con particolare evidenza dall'analisi degli orientamenti elaborati dalla Commissione europea in attuazione del Regolamento, entrato in vigore il 10 ottobre 2024. L'articolo 8, paragrafo 2, del regolamento invita esplicitamente la Commissione a elaborare orientamenti comuni per contribuire all'efficace attuazione dello stesso e, in particolare, per sostenere gli sponsor o i prestatori di servizi pubblicitari nel dichiarare e nell'identificare i messaggi di pubblicità politica, nonché per aiutare i prestatori di servizi a facilitare e gestire adeguatamente tali dichiarazioni.

Gli orientamenti pubblicati dovrebbero fungere da guida pratica per assistere le autorità competenti nelle loro attività di esecuzione e i diversi soggetti contemplati dal regolamento nel garantire il rispetto degli obblighi imposti.

Tuttavia, un'attenta disamina rivela che gli orientamenti non riescono a chiarire i problemi interpretativi relativi all'apparato definitorio e ai criteri di individuazione dei messaggi. La Commissione si è limitata a riprodurre il con-

tenuto normativo dell'articolato, senza fornire quella necessaria attività ermeneutica che avrebbe potuto guidare gli operatori nell'applicazione pratica delle disposizioni.

Per quanto concerne i criteri di identificazione di tali messaggi, gli orientamenti offrono solo la formulazione di alcune domande orientative che gli operatori dovrebbero porsi nel valutare i potenziali *political advertisement*, senza tuttavia fornire parametri interpretativi chiari o casistiche concrete. Questa lacuna applicativa appare particolarmente problematica se si considera la centralità che l'individuazione della pubblicità politica riveste nell'intero impianto del regolamento: senza criteri definiti e condivisi, il rischio è quello di generare interpretazioni disomogenee tra gli Stati membri, vanificando proprio quell'obiettivo di armonizzazione che costituiva la *ratio* della scelta dello strumento regolamentare. La debolezza degli orientamenti evidenzia, ancora una volta, la difficoltà del diritto europeo di fornire risposte concrete a fenomeni complessi e in continua trasformazione.

Di fronte a tale impasse interpretativa, sovviene una provocazione: se gli strumenti giuridici tradizionali si rivelano inadeguati nell'individuare con precisione i contenuti di pubblicità politica, potrebbe essere utile ricorrere proprio a quegli strumenti tecnologici che si intendono regolare. Detto in altri termini, si potrebbero allenare algoritmi e sistemi di intelligenza artificiale specificamente progettati per identificare, classificare e monitorare tali messaggi per ovviare al problema della vaghezza definitoria che affligge il quadro normativo europeo. Questa prospettiva, per quanto paradossale, trova una sua logica nell'idea che la complessità e la velocità dell'ecosistema digitale richiedano strumenti di controllo altrettanto sofisticati e dinamici. Tuttavia, questa soluzione tecnologica solleva alcuni interrogativi: chi controllerebbe questi algoritmi? Secondo quali criteri sarebbero addestrati? Il rischio concreto è quello di sostituire l'opacità algoritmica della propaganda politica con l'opacità algoritmica del controllo pubblico, generando un circolo vizioso in cui la tecnologia viene utilizzata per regolare se stessa, sottraendosi completamente al controllo democratico. La provocazione, pertanto, non va intesa come una soluzione definitiva, ma come una sollecitazione a ripensare radicalmente gli strumenti di governance della comunicazione politica digitale, superando l'illusione che il diritto tradizionale possa, da solo, governare fenomeni la cui natura tecnologica sfugge alle categorie giuridiche consolidate.

A ogni modo, il Regolamento 900/2024 è declinato secondo un approccio che privilegia la trasparenza procedurale rispetto alla censura contenutistica, mantenendo così l'equilibrio tra protezione democratica e libertà di espressione.

Le disposizioni relative all'etichettatura obbligatoria dei contenuti pubbli-

citari, alla *disclosure* delle tecniche di targeting e alla limitazione del microtargeting per i minori riflettono una strategia normativa volta a ripristinare le condizioni basilari affinché il dibattito possa essere democratico e informato, senza compromettere la libertà di espressione. Dall'analisi, lo stesso sembra riflettere una concezione diversa del rapporto tecnologia-democrazia. L'obbligo di trasparenza algoritmica per le tecniche di targeting, la creazione di registri pubblici delle pubblicità politiche e l'introduzione di meccanismi di *accountability* per le piattaforme digitali configurano un modello di governance che riconosce la necessità di rendere visibili e controllabili i processi algoritmici che influenzano la formazione dell'opinione pubblica.

Nonostante questi progressi normativi, l'esperienza delle elezioni europee del 2024 ha rivelato alcune criticità. Pur rappresentando un primo banco di prova dell'approccio alla trasparenza politica digitale, quest'ultimo appuntamento elettorale ha confermato la necessità di un maggiore coordinamento tra autorità nazionali e istituzioni europee. La frammentazione delle competenze di vigilanza tra diverse autorità genera sovrapposizioni e vuoti di controllo che gli attori malintenzionati possono facilmente sfruttare.

Emerge perciò lampante la necessità di interventi immediati, quali per esempio: l'istituzione di un'Autorità europea dedicata, la creazione di un sistema di monitoraggio in tempo reale durante le campagne elettorali e l'armonizzazione delle sanzioni minime.

Il modello europeo di "democrazia digitale" deve fondarsi su tre pilastri: trasparenza algoritmica, responsabilizzazione delle piattaforme e maggiore consapevolezza informativa dei cittadini⁵².

La sfida consiste nel bilanciare la libertà di espressione politica con l'esigenza di preservare l'integrità del processo democratico, evitando che proprio l'agorà democratica possa trasformarsi in un mercato dove le idee politiche vengono commercializzate come prodotti di consumo attraverso sofisticate tecniche di persuasione digitale.

In questa prospettiva, emerge con particolare evidenza un paradosso strutturale delle democrazie contemporanee: la persistente gerontocrazia nei corpi rappresentativi si scontra con la crescente digitalizzazione della comunicazione politica, ambito nel quale i giovani dimostrano una maggiore dimestichezza. I parlamenti e i governi delle democrazie occidentali risultano infatti dominati da classi dirigenti sopra una certa soglia di età, mentre la Generazione Z rimane sistematicamente sottorappresentata nelle istituzioni, nonostante costitui-

52. R. Borrello, *Le forme di raccordo tra le autorità indipendenti nazionali e la realizzazione di un modello comune europeo di disciplina della comunicazione politica: recenti sviluppi*, in *DPCE online*, 47, 2021.

sca una parte significativa della popolazione e sia la principale destinataria di decisioni politiche i cui effetti si dispiegano nel lungo periodo – si pensi alle politiche climatiche, previdenziali, occupazionali, educative e sanitarie. Questa asimmetria generazionale produce un doppio cortocircuito democratico: da un lato, coloro che padroneggiano gli strumenti digitali della comunicazione politica contemporanea non hanno accesso alle sedi decisionali formali; dall'altro, chi detiene il potere spesso non comprende appieno le dinamiche della sfera pubblica digitale che governano.

Non sorprende, dunque, che proprio i giovani abbiano fatto ricorso alle piattaforme digitali per aggirare i canali tradizionali di rappresentanza, dando vita a quelle forme di mobilitazione spontanea e transnazionale che, come evidenziato in apertura, hanno scosso numerosi governi dal Madagascar al Nepal, dal Marocco all'Indonesia. Le proteste della Generazione Z testimoniano l'emersione di una nuova modalità di partecipazione politica che prescinde dalle strutture partitiche tradizionali e si organizza attraverso reti digitali orizzontali, rivelando al contempo i limiti della democrazia rappresentativa nell'intercettare le istanze di una generazione che si percepisce esclusa dai processi decisionali. Questa dinamica pone interrogativi fondamentali sulla sostenibilità di un modello democratico in cui la rappresentanza formale e la partecipazione informale seguono logiche sempre più divergenti, alimentando un deficit di legittimazione che né il Regolamento 2024/900 né gli altri strumenti normativi analizzati affrontano adeguatamente.

Alla luce di ciò, il Ttp costituisce un primo passo, ma non del tutto sufficiente. La sua efficacia dipenderà dalla volontà politica di procedere rapidamente alle modifiche strutturali e dalla capacità delle istituzioni europee di adattare continuamente gli strumenti normativi all'evoluzione tecnologica⁵³. Solo attraverso questo approccio dinamico e integrato sarà possibile garantire che la rivoluzione digitale costituisca un'opportunità di arricchimento del dibattito democratico e non una minaccia alla sua stessa sopravvivenza.

Peraltro, proprio quest'atto evidenzia come la disciplina della comunicazione politica online rappresenti un banco di prova per l'evoluzione del costituzionalismo europeo nell'era digitale. La transizione dalla neutralità tecnologica alla governance attiva dei contenuti politici segna il passaggio a una situazione che richiede nuovi strumenti, tanto teorici quanto pratici.

In prospettiva, tre questioni paiono centrali per lo sviluppo di una governance costituzionale dell'ecosistema digitale:

53. F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino (a cura di), *La regolazione europea della società digitale*, cit.

1. la definizione di standard comuni per la limitazione del potere privato che preservino il pluralismo democratico;
2. lo sviluppo di meccanismi che garantiscano l'effettività delle tutele;
3. la creazione di un sistema di bilanciamento dei diritti fondamentali che tenga conto delle specificità tecnologiche senza compromettere i principi costituzionali.

Il Regolamento 2024/900, pur rappresentando un significativo passo avanti, costituisce solo l'inizio di un processo di adattamento del diritto alle sfide della digitalizzazione. La sua effettiva implementazione fornirà elementi importanti per valutare l'adequatezza dell'approccio europeo e orientare future evoluzioni normative verso una governance digitale pienamente compatibile con i principi del costituzionalismo democratico.

FIGHTING FOREIGN INTERFERENCE.
UN'ANALISI COMPARATA DELLE LEGISLAZIONI
CONTRO LE INTERFERENZE STRANIERE
NEI PROCESSI ELETTORALI IN AUSTRALIA, CANADA,
IRLANDA, REGNO UNITO E NUOVA ZELANDA

*Emanuele Gabriele**

SOMMARIO: 1. Vulnerabilità democratica ed elezioni nel disordine informativo – 2. Una ricognizione sulla *ratio* e la natura delle fonti rilevanti nei sistemi del modello Westminster – 3. Le principali misure di prevenzione e protezione dell'integrità elettorale negli ordinamenti del modello – 4. Per una sintesi sulle traiettorie normative emergenti.

1. Vulnerabilità democratica ed elezioni nel disordine informativo

Le interferenze straniere nei processi elettorali si presentano, negli ultimi anni in misura sempre maggiore, come uno dei fattori di massima vulnerabilità per l'integrità e la tenuta delle democrazie contemporanee.

Certamente, la questione non è di recente emersione. In principio era l'ingerenza francese nelle presidenziali americane del 1796, ma la pratica si è arricchita di numerosi casi nel XX secolo, con l'interferenza statunitense in Italia nel 1948 e in Cile nel 1970, e quella russa in Ucraina nel 2004, solo per citarne alcune tra le più indagate¹. E neppure il fenomeno gode di rara applicazione, dato che le stime riportano come l'11,3% delle elezioni tra il 1946 e il 2000 abbia subito qualche forma di ingerenza esterna, spesso tramite pressioni economiche, propaganda o sostegno occulto a partiti e candidati².

Ciò che è cambiato oggi non è, dunque, la pratica, ma gli strumenti per attuarla: le reti digitali, i social media e l'intelligenza artificiale permettono

* Dottore di ricerca in Diritto pubblico, comparato e internazionale, Sapienza Università di Roma.

1. Cfr. D.H. Levin, *Should We Worry about Partisan Electoral Interventions? The Nature, History, and Known Effects of Foreign Interference in Elections*, in D.B. Hollis, J.D. Ohlin (eds.), *Defending Democracies*, Oxford University Press, 2021, pp. 19-39.

2. V. Mohan, A. Wall, *Foreign Electoral Interference*, in *Georgetown Journal of International Affairs*, n. 20, 2019, pp. 110-119, 111.

un'ingerenza capillare e invisibile, capace di manipolare le percezioni più che le urne. Dall'influenza russa sulle elezioni presidenziali americane del 2016³ fino alle campagne di disinformazione che hanno accompagnato le consultazioni in diversi Paesi europei nel 2024⁴, il problema non è più solo politico o tecnologico, diventa pure un problema di diritto internazionale e costituzionale, poiché assume i caratteri dello sgretolamento dell'integrità elettorale, ossia di quella combinazione di libertà, equità, trasparenza e sicurezza dei processi di voto che garantisce la legittimità democratica del risultato⁵. La crescente articolazione delle tecniche di influenza sollecita, dunque, l'elaborazione di strumenti normativi coordinati e la ricostruzione dei tratti comuni ai diversi sistemi di protezione dell'integrità elettorale.

Sul tema, in via preliminare, è fondamentale distinguere tra influenza e interferenza, ossia tra la promozione di valori o interessi attraverso strumenti diplomatici o culturali – che rientra nella normale dialettica delle relazioni internazionali – e l'azione intenzionale di un soggetto esterno – statale o non statale – col fine di alterare un processo elettorale, manipolare l'opinione pubblica o condizionare l'esito del voto⁶.

È in questo passaggio che l'influenza si trasforma in violazione di sovranità. L'interferenza elettorale, infatti, non soltanto lede il principio di non ingerenza negli affari interni (come stabilito dall'art. 2 della Carta delle Nazioni Unite)⁷, ma può pure costituire una violazione del diritto dei popoli all'autodeterminazione (sancito, tra l'altro, dall'art. 1 della Convenzione internazionale sui diritti civili e politici e dall'art. 1 della Convenzione internazionale sui diritti economici, sociali e culturali) quando limita o intacca la libera formazione dell'opinione elettorale⁸.

3. Cfr. J. Blake, *Russian Interference in U.S. Elections: How to Protect Critical Election Infrastructure from Foreign Participation*, in *Public Contract Law Journal*, vol. 49, n. 4, 2020, pp. 709-734; A. Burak Daricili, *Analysis of Manipulation of the Russian Federation in the 2016 Presidential Elections of the United States of America within the Scope of Intelligence Techniques*, in *Güvenlik Bilimleri Dergisi*, vol. 8, n. 1, 2019, pp. 133-150.

4. Cfr. E. Saeva, I. Tasheva, *The 2024 EU elections and cybersecurity: A retrospective and lessons learned*, in *European View*, vol. 23, n. 2, 2024, pp. 226-234; A.L. Pup, *Intelligence Agencies and the Ballot Box: Countering Foreign Interference in Romania's 2024 Presidential Elections*, in *International Journal of Intelligence and CounterIntelligence*, 2025, pp. 1-36.

5. Sul concetto di integrità elettorale si v. anzitutto P. Norris, *Why Electoral Integrity Matters*, Cambridge University Press, 2014.

6. Cfr. P. Baines, N. Jones, *Influence and Interference in Foreign Elections*, in *The RUSI Journal*, vol. 163, n. 1, 2018, pp. 12-19.

7. D.B. Hollis, J. Neutze, *Defending Democracies via Cybern norms*, in D.B. Hollis, J.D. Ohlin (a cura di), *Defending Democracies*, cit., pp. 315-348, 315.

8. P. Rothkirch, *Foreign Election Interference: A Violation of the Right to Self-Determination*, in *Voelkerrechtsblog*, disponibile su: voelkerrechtsblog.org/.

Il problema, tuttavia, è che il diritto internazionale non dispone ancora di un'architettura normativa efficace, capace di sanzionare in modo univoco queste condotte⁹, e pure le posizioni degli Stati restano divergenti, con alcuni che riconoscono come illecito solo l'attacco diretto alle infrastrutture elettorali, e altri che estendono la violazione anche alle campagne di disinformazione e alle manipolazioni online.

Oggi, infatti, le interferenze avvengono prevalentemente nel dominio digitale, con l'intelligenza artificiale che amplifica in modo esponenziale la capacità di creare e diffondere contenuti falsi, manipolati o mirati¹⁰. La disinformazione algoritmica sfrutta la libertà di espressione e il pluralismo dell'informazione per destabilizzare i valori democratici che questi principi dovrebbero invece difendere. In questo contesto, «votare nel disordine informativo»¹¹ significa esercitare le proprie prerogative democratiche in un contesto in cui il cittadino, sommerso da informazioni contraddittorie, perde la capacità di distinguere il vero dal verosimile, il dibattito politico dal condizionamento.

È in questa congerie particolarmente convulsa e costantemente mutante che occorre prestare attenzione alle prime, sistematiche risposte legislative degli ordinamenti nazionali, che negli ultimi anni hanno approvato strumenti normativi con lo scopo di arginare il fenomeno, talvolta ponendosi in linea di continuità con quelli approvati a livello sovranazionale (si pensi al *Digital Services Act* eurounitario).

In particolare, un esame comparato delle legislazioni in materia di trasparenza politica, sicurezza informativa e protezione dei dati elettorali dei sistemi anglosassoni di matrice Westminster (e, nello specifico, di quelli australiano, neozelandese, britannico, canadese e irlandese) permette di cogliere prospettive diverse – derivanti dalla sensibilità propria di ciascun ordinamento alla materia – ma complementari, per comprendere quali strumenti giuridici siano stati introdotti per proteggere l'integrità elettorale e quali criticità rimangano ancora aperte.

9. D.B. Hollis, J. Neutze, *Defending Democracies via Cybernorms*, cit., pp. 317-323.

10. O. Pollicino, P. Dunn, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *federalismi.it*, n. 12, 2024, pp. IV-XXII.

11. M.G. Sessa, *Le interferenze elettorali: votare nel disordine informativo*, in C. Fumagalli, V. Ottonelli (a cura di), *Votare o no. La pratica democratica del voto, tra diritto individuale e scelta collettiva*, Feltrinelli, 2022, pp. 121-134.

2. Una ricognizione sulla *ratio* e la natura delle fonti rilevanti nei sistemi del modello Westminster

L'introduzione di legislazioni volte a preservare della fiducia pubblica nei processi democratici e a salvaguardia dell'integrità elettorale dalla minaccia delle interferenze straniere ha preso avvio, negli ordinamenti del sistema Westminster, a partire dall'ingerenza russa nelle presidenziali statunitensi del 2016, che hanno rappresentato da questo punto di vista l'espressione plastica di un cambio di rotta legislativo sul tema. Ciascuno degli ordinamenti in esame ha reagito più o meno speditamente alla questione, moltiplicandosi, nell'ultimo decennio, le risposte legislative a una problematica tanto permeante quanto mutevole.

Passando in rassegna in ordine cronologico queste reazioni, occorre partire dal caso australiano. L'esperienza di Canberra costituisce, infatti, il primo e più articolato tentativo di disciplinare organicamente il fenomeno. La strategia legislativa nasce nel biennio 2017-2018, in un contesto segnato dalle rivelazioni nelle elezioni statunitensi e dalle crescenti preoccupazioni per l'influenza politica cinese, oltre che dall'urgenza di garantire lo svolgimento democratico delle elezioni suppletive del 2018¹². La riforma, promossa dal Governo liberale di Malcolm Turnbull, si è concretizzata in una complessa strategia multilivello, ossia in un pacchetto di «*interlocking and overlapping legislation*»¹³ che poggia su tre strumenti principali: il *National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018*, il *Foreign Influence Transparency Scheme Act 2018* (Fits) e l'*Electoral Legislation Amendment (Electoral Funding and Disclosure Reform) Act 2018*. Questa struttura normativa è coordinata dall'*Home Affairs Department* nell'ambito della *Counter-Foreign Interference Strategy* (Cfis), una strategia nazionale permanente lanciata nel 2018. Il primo strumento legislativo introduce un nuovo articolo 92 nel *Criminal Code* australiano, definendo la *foreign interference* come l'insieme di quelle attività coercitive, corruttive, ingannevoli o clandestine condotte da o per conto di un attore straniero, intese a compromettere la sicurezza nazionale o i processi democratici. Dall'interferenza straniera viene distinta la *foreign influence*, che non presuppone un intento manipolativo, e che è regolata dal *Foreign Influence Transparency Scheme Act 2018*, il quale riconosce la liceità delle azioni di in-

12. D. Mack, *An Era of Foreign Political Interference: Impulsive, Overcompensation of Australia, and a Comparison of Legislative Schemes with the United States*, in *Emory International Law Review*, vol. 34, 2020, pp. 367-397, 368-370.

13. S. Sorial, S. Morris, P. Grete, *Foreign Interference and the Incremental Chilling of Free Speech*, in *Federal Law Review*, vol. 52, n. 1, 2024, pp. 103-127, 104.

fluenza, ma istituisce un registro pubblico per chiunque intrattenga rapporti di lobbying o comunicazione politica con soggetti stranieri. Infine, l'*Electoral Legislation Amendment Act 2018* regola, restringendo ulteriormente la normativa già in vigore, le donazioni straniere nei confronti dei soggetti partitici australiani.

La strategia canadese si sviluppa tra il 2017 e il 2019, quando il Governo Trudeau avvia un piano di riforme in risposta alle preoccupazioni derivanti dalle interferenze russe nelle elezioni statunitensi del 2016, dai tentativi di influenza cinese sulle comunità sino-canadesi e sui partiti politici nazionali e dalle preoccupazioni per l'uso manipolativo delle piattaforme digitali nelle campagne elettorali nazionali e provinciali. Nel 2017, il *Minister of Democratic Institutions* riceveva mandato di proteggere il processo elettorale da minacce informatiche, dando avvio a quel modello integrato di *governance*, oggi evoluto in un vero "ecosistema di protezione democratica", fondato su cooperazione interagenzia, trasparenza e responsabilità pubblica. Gli strumenti principali adottati da Ottawa per far fronte alla questione delle interferenze straniere sono stati, in una prima fase, il *National Security and Intelligence Committee of Parliamentarians Act 2017*, che istituiva una commissione bicamerale sulla materia, e l'*Elections Modernization Act 2018*, che modificava il *Canada Elections Act* introducendo nuove regole su trasparenza, spesa politica e pubblicità online, vietava le donazioni estere e impostava la tracciabilità delle comunicazioni digitali. Negli anni seguenti sono stati istituiti ulteriori comitati, che hanno affiancato la commissione bicamerale nata nel 2017, con lo scopo di studiare la materia e proporre ulteriori raccomandazioni. In particolare, nel 2019 è stata costituita la *National Security and Intelligence Review Agency*, mentre nel 2023 è stata istituita la *Foreign Interference Commission*. È proprio a seguito dei rapporti pubblicati nel 2024 da questi tre organi che è stata approvata un'ulteriore e più strutturata legislazione, il *Countering Foreign Interference Act 2024*¹⁴. Secondo la definizione che di interferenza straniera dà l'articolo 20.4(1) della nuova legislazione, rientrano nella condotta illecita quei comportamenti occulti o ingannevoli, realizzati su istruzione di, o in collaborazione con un'entità straniera, con l'intento di influenzare un processo politico o governativo, la gestione di un'istituzione educativa, l'adempimento di un dovere connesso a tali processi o gestioni, ovvero l'esercizio di un diritto democratico nel territorio del Canada. Tra i *political or governmental processes*, ai sensi dell'articolo 20.4(4), rientrano anche lo svolgimento di elezioni e referendum. Il provvedimento crea, inoltre, un *Foreign Influence Transparency Registry*, obbligando in-

14. Sul punto si v. A. Fiorentino, *Il Canada si avvicina alle elezioni federali all'ombra dello scandalo sulle interferenze straniere*, in *Nomos. Le attualità nel diritto*, n. 2, 2024, pp. 1-15, 9.

dividui e organizzazioni che agiscano per conto di governi o entità estere a registrarsi presso un *Foreign Influence Transparency Commissioner* indipendente.

Nel caso irlandese, la risposta normativa ha preso la forma di una legge organica, l'*Electoral Reform Act 2022*, che ha riformato l'intero sistema elettorale, introducendo misure innovative di regolazione delle campagne online e di contrasto alla disinformazione¹⁵. La legge, nata dopo le controversie legate al referendum sull'aborto del 2018 e alle interferenze digitali connesse al voto sulla Brexit, ha riconosciuto la necessità di un controllo istituzionale indipendente sul flusso informativo elettorale¹⁶. Il provvedimento istituisce la *Electoral Commission*, organismo autonomo con poteri di vigilanza, monitoraggio e ricerca, e disciplina la pubblicità politica online, imponendo l'obbligo di etichettatura, la tracciabilità dei finanziatori e un *repository* pubblico delle inserzioni, vietando al contempo la vendita di pubblicità politica online ad attori stranieri (articolo 125 della legge).

La strategia britannica in materia di interferenze straniere si colloca all'interno di una più ampia riconfigurazione del *National Security Law*. Dopo un ventennio dominato dal paradigma del *counter-terrorism*, il Regno Unito ha avviato un «*pivot back to state threats*»¹⁷: un ritorno alla centralità delle minacce poste da Stati ostili e dai loro *proxy*, in particolare Russia, Cina e Iran. Il punto di svolta nell'implementazione di una legislazione in materia è rappresentato dal lavoro promosso dalla *Law Commission Review on the Protection of Official Data* (2017-2020), che aveva evidenziato la vetustà del regime degli *Official Secrets Acts* (approvati tra il 1911 e il 1939). A ciò si aggiungono due fattori determinanti: le raccomandazioni del c.d. *Russia Report* del 2020¹⁸, stilato dall'*Intelligence and Security Committee of Parliament*, che denunciava la vulnerabilità del sistema elettorale britannico a interferenze digitali e finanziarie esterne, e la crescente attività di disinformazione e influenza occulta nel contesto post-Brexit¹⁹. Su questa base, il Governo Johnson prima e quello Sunak poi, hanno promosso il *National Security Bill*, presentato nel 2022 e di-

15. Cfr. H. Vasani, T.S. James, H.A. Garnett, *Ireland has moved to improve electoral integrity – other countries should follow suit*, in *LSE Blog*, disponibile su: blogs.lse.ac.uk/.

16. D. Shattock, *Knowledge, deception, and freedom of expression: a critical examination of Ireland's approach to disinformation under the Electoral Reform Act 2022*, in *International Review of Law, Computers & Technology*, 2024, pp. 1-30, 2.

17. P.F. Scott, *"State threats", security, and democracy: the National Security Act 2023*, in *Legal Studies*, vol. 44, 2024, pp. 260-276, 261.

18. Intelligence and Security Committee of Parliament, *Russia. Presented to Parliament pursuant to section 3 of the Justice and Security Act 2013*, HC 632, disponibile su: isc.independent.gov.uk/.

19. P.F. Scott, *"State threats", security, and democracy: the National Security Act 2023*, cit., pp. 269-271.

venuto legge nel luglio 2023. Il provvedimento abroga in larga parte gli *Official Secrets Acts* e introduce, nelle sezioni 13, 14 e 16 i reati di *foreign interference*, intesi come condotte occulte o ingannevoli poste in essere per conto di una potenza straniera con lo scopo di influenzare un processo politico, elettorale o decisionale nel Regno Unito. Nel linguaggio della disposizione, la norma presuppone la sussistenza di una *foreign power condition* (clausola di collegamento con una potenza straniera) e di un *prejudicial purpose* (scopo pregiudizievole per la sicurezza o gli interessi del Regno Unito). A completare il quadro è il *Foreign Influence Registration Scheme* (Firs), operativo dal luglio 2025, che impone l'obbligo di registrazione per le attività di lobbying e comunicazione politica condotte per conto di governi stranieri.

La strategia neozelandese in materia di interferenze straniere è relativamente recente, e nasce da una progressiva consapevolezza, maturata abbandonando la precedente autopercezione di perifericità, dell'importanza della materia²⁰. A partire dal *National Security Strategy 2023-2028*, il Governo ha incluso la *foreign interference* tra le principali minacce alla sicurezza nazionale. Il risultato di questa evoluzione è stato il *Crimes (Countering Foreign Interference) Amendment Act 2025*, approvato nel novembre 2025 e destinato a modificare il *Crimes Act 1961*. Questo provvedimento rappresenta il primo tentativo organico di criminalizzare esplicitamente le condotte di interferenza straniera, analogamente al modello australiano. In particolare, il nuovo articolo 78AAA del *Crimes Act 1961*, introduce una nozione intenzionale e teleologica di interferenza straniera, che presuppone un legame con un potere straniero (diretto o indiretto) e la finalità di compromettere interessi fondamentali del Paese, tra cui «*the fair and impartial conduct of any election, poll, or referendum under the law of New Zealand*».

3. Le principali misure di prevenzione e protezione dell'integrità elettorale negli ordinamenti del modello

L'emersione del fenomeno delle interferenze straniere nei processi democratici ha determinato, negli ordinamenti del Commonwealth, quando non una trasformazione strutturale del diritto della sicurezza e del diritto elettorale, almeno una rimodulazione del quadro normativo in materia. La legislazione di cui si è dato brevemente conto dimostra come, negli ultimi anni, si sia tentato di combinare misure di prevenzione, trasparenza e repressione in risposta a

20. Cfr. G. Orr, D. Geddis, *Islands in the Storm? Responses to Foreign Electoral Interference in Australia and New Zealand*, in *Election Law Journal*, 2020, pp. 1-16, 12.

una minaccia – la manipolazione occulta dell’informazione, del finanziamento politico e del consenso elettorale – ritenuta non più eccezionale ma sistemica.

In questo senso, le strategie preventive rappresentano il primo livello di risposta legislativa, e assumono configurazioni differenti nei vari ordinamenti. Nel caso australiano, il *Foreign Influence Transparency Scheme Act 2018* (Fits) istituisce un sistema di registrazione obbligatoria per tutte le persone o organizzazioni che intrattengono rapporti di tipo politico, comunicativo o decisionale per conto di un’entità estera (art. 14). L’obiettivo è rendere visibile ciò che, in assenza di tale schema, potrebbe avvenire occultamente e in modo illecito: lobbying, campagne mediatiche o attività di *advocacy* dirette a influenzare le scelte elettorali²¹. Il *Fits Register* impone l’obbligo di dichiarazione dettagliata delle attività e delle risorse impiegate, prevedendo sanzioni penali per l’omissione (art. 57).

Un meccanismo analogo è introdotto nel Regno Unito con il *Foreign Activities and Foreign Influence Registration Scheme* (Firs), disciplinato dal *National Security Act 2023* agli articoli 65-83 ed entrato in operatività, come si è detto, nel luglio 2025. Il Firs si articola in due livelli (“*tiers*”) principali. Il livello primario (o *political tier*) riguarda le *foreign influence arrangements*, ossia accordi in cui un potere straniero incarica un soggetto di svolgere attività di influenza politica nel Regno Unito. Tali accordi devono essere registrati entro 28 giorni, pena sanzioni penali. Le attività di “influenza politica” sono definite in senso ampio, includendo la distribuzione di denaro o servizi, la comunicazione pubblica (salvo che la provenienza straniera sia chiaramente indicata) e l’interazione con funzionari pubblici, parlamentari o membri delle forze armate²². Lo schema non vieta tali attività, ma le subordina a un obbligo di trasparenza. Il livello avanzato (*enhanced tier*) consente al Segretario di Stato di designare specifiche potenze straniere come soggetti a obbligo di registrazione rafforzata, qualora ciò sia ritenuto necessario per la sicurezza nazionale. Chiunque concluda un accordo con tali soggetti deve registrarlo entro dieci giorni. Anche in questo caso, la mancata registrazione costituisce reato, sia per chi promuove l’attività sia per chi la realizza²³. Subito dopo l’entrata in vigore del Firs, il

21. Cfr. M.E. Dowling, *Foreign interference and Australian electoral security in the digital era*, in *Australian Journal of International Affairs*, vol. 76, n. 1, 2022, pp. 40-56, 44; sul punto anche D. Mack, *An Era of Foreign Political Interference: Impulsive, Overcompensation of Australia, and a Comparison of Legislative Schemes with the United States*, cit., pp. 372-374.

22. Cfr. P.F. Scott, “*State threats*”, *security, and democracy: the National Security Act 2023*, cit., p. 268.

23. Per un’analisi della disposizione si rimanda a R. Ward, *State Threat Activity. Principal Offences*, in R. Ward, D. Blundel (a cura di), *National Security Law, Procedure and Practice*, II. ed., Oxford University Press 2024, pp. 649-684, 681-682.

Minister of State for Security Dan Jarvis, tramite *regulation*, ha definito Russia e Iran come le prime potenze straniere per cui è richiesto un obbligo di registrazione rafforzata²⁴. In questo modo, il *Foreign Influence Registration Scheme* rappresenta un tentativo di bilanciare la tutela della sicurezza nazionale con i principi di trasparenza e libertà politica²⁵.

In Canada, la prevenzione è costruita intorno al *Countering Foreign Interference Act 2024*, che nella sua Part 4 contiene il *Foreign Influence Transparency and Accountability Act*, a sua volta istituyente un *Foreign Influence Transparency Registry* ispirato al modello australiano. Gli artt. 5-8 della legge prevedono che chiunque svolga attività politiche o di comunicazione per conto di un governo straniero debba dichiararlo, consentendo un controllo *ex ante* delle interazioni tra interessi esteri e processi decisionali. Comunque, alla fine del 2025, a causa della *dissolution* del Parlamento di Ottawa intercorsa in primavera, il Registro doveva ancora entrare in vigore²⁶.

L'approccio irlandese si discosta in modo significativo. Qui la prevenzione si fonda non su obblighi di registrazione, ma su un modello istituzionale di vigilanza permanente, rappresentato dalla *Electoral Commission*. La Commissione svolge funzioni di monitoraggio, informazione civica e supervisione sulle campagne digitali, pubblicando rapporti periodici sullo stato dell'integrità elettorale²⁷. Si tratta di un paradigma "fiduciario" della prevenzione, basato su trasparenza e responsabilizzazione sociale piuttosto che su misure penali.

Certamente, il nucleo più incisivo delle normative analizzate riguarda la criminalizzazione diretta dell'interferenza straniera, concepita come condotta ostile contro la sovranità statale e il corretto funzionamento delle istituzioni democratiche. Il riferimento originario è il *Criminal Code* australiano, emendato dal *National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018*, che introduce nella *Division 92* i reati di *foreign interference* ed *espionage*. La *ratio* è eminentemente securitaria: l'interferenza è trattata come minaccia alla sicurezza nazionale, non come violazione dell'equità elettorale²⁸.

24. UK Government, *UK launches Foreign Influence Registration Scheme*, disponibile su: www.gov.uk/.

25. Cfr. P.F. Scott, "State threats," *security, and democracy: the National Security Act 2023*, cit., pp. 267-270.

26. Canada Government, *Question Period Note: Implementation of the Foreign Influence Transparency Registry*, disponibile su: search.open.canada.ca/.

27. D. Shattock, *Knowledge, deception, and freedom of expression: a critical examination of Ireland's approach to disinformation under the Electoral Reform Act 2022*, cit., p. 4.

28. Cfr. D. Mack, *An Era of Foreign Political Interference: Impulsive, Overcompensation of Australia, and a Comparison of Legislative Schemes with the United States*, cit., p. 376; S. Sorial, S. Morris, P. Greste, *Foreign Interference and the Incremental Chilling of Free Speech*, cit., p. 116.

Un impianto simile è stato adottato in Nuova Zelanda con il *Crimes (Countering Foreign Interference) Amendment Act 2025* che, come si è detto, modifica il *Crimes Act 1961*. Le nuove sezioni 78AAA e 78AAB puniscono, rispettivamente, l'interferenza straniera e la collaborazione consapevole con una potenza estera, con il primo che include espressamente i processi elettorali. L'ampiezza della formulazione è stata oggetto di osservazioni critiche da parte della *New Zealand Law Society*, che ha segnalato il rischio di un effetto dissuasivo sulla libertà di associazione²⁹.

Nel Regno Unito, il *National Security Act 2023* (art. 16) codifica il reato di *foreign interference in elections*. L'*offence* richiede la presenza della *foreign power condition* (art. 31), ossia un legame sostanziale con un governo estero, e di un intento pregiudizievole (*prejudicial purpose*). Si tratta di una norma che riprende la logica australiana, ma inserita in un quadro più ampio di "*state threats*" comprendente spionaggio, sabotaggio e tradimento³⁰.

In Canada, il *Countering Foreign Interference Act 2024* ha introdotto un reato analogo emendando il *Security of Information Act*, definendo l'interferenza straniera come, si è visto, quelle condotte occulte o ingannevoli, realizzate su istruzione di, o in collaborazione con un'entità straniera, con l'intento di influenzare un processo politico o governativo, la gestione di un'istituzione educativa, l'adempimento di un dovere connesso a tali processi o gestioni, ovvero l'esercizio di un diritto democratico nel territorio del Canada.

La regolazione del finanziamento politico rappresenta un ulteriore strumento di contrasto alle interferenze, mirato a interrompere i canali economici di influenza straniera. In Australia, l'*Electoral Legislation Amendment (Electoral Funding and Disclosure Reform) Act 2018* modifica il *Commonwealth Electoral Act 1918* introducendo limiti rigorosi alle donazioni estere. Gli articoli 287AA e 287AB vietano i contributi politici provenienti da donatori non elettori o da soggetti stranieri, sia diretti sia indiretti, e impongono ai partiti di dichiarare ogni pagamento superiore a 1.000 Aud proveniente da fonte estera.

In Canada, il *Canada Elections Act 2000* (art. 349) vietava già qualsiasi donazione o spesa elettorale da parte di cittadini o entità straniere, mentre l'*Elections Modernization Act 2018* ha esteso il divieto anche alle attività digitali, incluse le sponsorizzazioni pubblicitarie online durante il periodo elettorale (art. 7).

In Irlanda, infine, l'art. 125 dell'*Electoral Reform Act 2022* vieta la pubblicità politica finanziata da soggetti non residenti nell'Unione europea, chiu-

29. New Zealand Law Society, *Submission to the Crimes (Countering Foreign Interference) Amendment Bill*, disponibile su: www.lawsociety.org.nz/.

30. P.F. Scott, "State threats", *security, and democracy: the National Security Act 2023*, cit.

dendo così una delle principali vie di accesso dell'influenza estera al dibattito pubblico.

Le nuove legislazioni Westminster hanno anche affrontato l'ambito della regolazione delle campagne digitali e della disinformazione elettorale. Ancora l'Irlanda è il primo ordinamento europeo ad aver introdotto una definizione normativa di *disinformation* (art. 166 *Electoral Reform Act 2022*), distinguendola dalla *misinformation* e attribuendo alla Commissione Elettorale il potere di emettere *take-down notices* (che rimuovono il contenuto disinformativo da una piattaforma), *correction notices* (che rettificano le informazioni parzialmente manipolate) e *labelling orders* (che evidenziano come l'informazione sia attualmente sotto indagine da parte della Commissione)³¹. In Canada, l'*Elections Act*, così come modificato dall'*Elections Modernization Act 2018*, impone (all'art. 325) a tutte le piattaforme digitali di mantenere un *registry of digital advertisements* accessibile al pubblico per due anni, contenente le informazioni sugli sponsor delle inserzioni.

Le normative esaminate mostrano una convergenza verso un modello di protezione integrata della democrazia, in cui prevenzione, trasparenza e repressione convivono in un equilibrio non sempre stabile. L'Australia, la Nuova Zelanda e il Regno Unito adottano un approccio securitario e penale, mentre Canada e Irlanda privilegiano la trasparenza e la fiducia pubblica come strumenti di resistenza democratica. In tutti i casi, tuttavia, la *foreign interference* è definita come condotta intenzionale e occulta volta a distorcere la sovranità decisionale dello Stato, e la sua repressione si estende oggi ben oltre la sfera penale, fino a comprendere il finanziamento, la comunicazione e la protezione dei dati.

4. Per una sintesi sulle traiettorie normative emergenti

L'analisi comparata delle esperienze di Australia, Canada, Irlanda, Nuova Zelanda e Regno Unito mette in luce un dato comune e una divergenza strutturale. Da un lato, tutti questi ordinamenti condividono la convinzione che l'interferenza straniera rappresenti non solo una minaccia alla sicurezza ma pure una sfida alla legittimità democratica. Dall'altro, ciascun sistema ha elaborato una propria "filosofia normativa" del rischio, riflessa nella struttura e nelle finalità della legislazione adottata.

Nel caso australiano, il tratto distintivo pare essere l'adozione, a paradigma interpretativo della legislazione in materia, della centralità della sicurezza na-

31. D. Shattock, *Knowledge, deception, and freedom of expression: a critical examination of Ireland's approach to disinformation under the Electoral Reform Act 2022*, cit., p. 6.

zionale, per cui la risposta di Canberra al problema delle ingerenze straniere in ambito elettorale è stata di natura eminentemente securitaria³². Il *National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018* ha inserito nel *Criminal Code* una definizione ampia di *foreign interference*, mentre il *Foreign Influence Transparency Scheme Act 2018* (Fits) ha istituito un registro obbligatorio per chiunque svolga attività di lobbying o *advocacy* per entità estere. Questa combinazione di prevenzione amministrativa e repressione penale, sostenuta da un forte coinvolgimento dell'intelligence, fa dell'Australia un modello che potrebbe definirsi di *state-centric democracy*, dove la protezione del processo elettorale è concepita come funzione della sicurezza nazionale, di cui non a caso viene data una definizione normativa ampia e onnicomprensiva³³.

Il Canada, al contrario, si distingue per un approccio cooperativo e istituzionalmente diffuso, che è stato definito dalla dottrina "ecosistemico"³⁴. L'idea centrale è quella che guarda all'adozione di una strategia di resilienza che integri prevenzione, educazione civica e trasparenza, equilibrando gli aspetti di tutela dei diritti e di fiducia nei confronti delle istituzioni. La difesa della sovranità elettorale, in altre parole, è perseguita non solo mediante la criminalizzazione delle interferenze, ma anche con la promozione della trasparenza digitale (come dimostra il registro delle pubblicità politiche online). In questo senso, il modello canadese rappresenta una forma di *governance* basata sul consenso e sulla consapevolezza, che tenta di prevenire la politicizzazione della sicurezza, sulla scia delle raccomandazioni che le diverse commissioni istituite sul tema hanno avanzato nel corso degli ultimi anni³⁵.

L'Irlanda, invece, costituisce un caso paradigmatico di regolazione della trasparenza. Con l'*Electoral Reform Act 2022*, Dublino ha affrontato il tema dell'integrità elettorale combinando innovazione tecnologica e garanzie costituzionali. La creazione dell'*Electoral Commission* e l'introduzione dei poteri di *take-down notices*, *correction notices* e *labelling orders* delineano un modello centrato sull'intervento informativo e non repressivo nella gestione della disinformazione.

32. Cfr. D. Mack, *An Era of Foreign Political Interference: Impulsive, Overcompensation of Australia, and a Comparison of Legislative Schemes with the United States*, cit., pp. 375-378.

33. Cfr. S. Sorial, S. Morris, P. Grete, *Foreign Interference and the Incremental Chilling of Free Speech*, cit., p. 116.

34. Y. Dawood, *Combating Foreign Election Interference: Canada's Electoral Ecosystem Approach to Disinformation and Cyber Threats*, in *Election Law Journal*, vol. 20, n. 1, 2021, pp. 10-31.

35. Cfr. ivi, p. 18; sul punto anche W. Wark, *How Allies Do It: Five Eyes Foreign Influence Transparency Registries*, in *Centre for International Governance Innovation. Policy Brief No. 184*, disponibile su: www.cigionline.org/, p. 3; Id., *Canadian democracy and foreign interference*, in *The Round Table*, vol. 114, n. 2, 2025, pp. 155-158, 157.

La Nuova Zelanda segue un percorso differente, caratterizzato da una forte enfasi sulla protezione della sovranità, assumendo un orientamento penale che riflette un modello difensivo e *intelligence-driven* simile, in molti aspetti, a quello australiano.

Infine, il Regno Unito si distingue per aver adottato la risposta più istituzionalmente complessa e securitaria. Con il *National Security Act 2023* e il *Foreign Influence Registration Scheme* (Firs), Londra ha ridefinito il concetto di *state threat*, includendo la *foreign interference* come reato specifico e introducendo un regime obbligatorio di registrazione per attività politiche condotte per conto di potenze straniere. Questo assetto consacra un modello di democrazia difensiva, nel quale la protezione delle elezioni è subordinata alla tutela della sicurezza nazionale e affidata all'interazione tra *Home Office*, Mi5 ed *Electoral Commission*. Certamente, comunque, il sistema britannico tradisce una persistente gerarchia tra Stato e società, in cui la libertà comunicativa resta subordinata alla logica del controllo preventivo.

IL CONTRASTO ALLA DISINFORMAZIONE NEI PROCEDIMENTI ELETTORALI: SPUNTI COMPARATIVISTICI TRA ITALIA E SPAGNA

*Lorenzo De Carlo**

SOMMARIO: 1. Introduzione – 2. Il modello spagnolo: la *Junta Electoral Central* come baricentro della tutela dell'integrità elettorale – 3. Il modello italiano tra frammentazione istituzionale e limitata reattività – 4. Una valutazione dell'efficacia dei due modelli in prospettiva comparata.

1. Introduzione

L'integrità e la sicurezza dei processi democratici nell'era digitale sono minacciate da un fenomeno pervasivo e insidioso: la disinformazione all'interno dei procedimenti elettorali, che rappresenta una minaccia diretta al corretto esercizio della sovranità popolare, all'equità del confronto politico e alla legittimità stessa degli esiti delle votazioni. Sia l'Italia sia la Spagna, due democrazie mature con una storia costituzionale caratterizzata da un processo di transizione dall'autoritarismo e dalla presenza di sistemi mediatici assai articolati, percepiscono questa esigenza come assolutamente urgente. Tuttavia, le risposte normative e istituzionali che i due Paesi hanno sviluppato divergono in modo significativo, offrendo un fertile terreno per un'analisi comparata. Partendo da un quadro normativo tendenzialmente scarno in entrambi i paesi, parzialmente colmato dall'entrata in vigore del regolamento europeo *Digital Services Act* (Dsa), l'analisi si concentrerà sul ruolo cruciale svolto dalle istituzioni pubbliche sul tema. In particolare, nel modello spagnolo si evidenzierà la centralità della *Junta Electoral Central* (Jec), un organo indipendente con poteri di intervento rapido e in grado di consolidare orientamenti giurisprudenziali. A questo si contrappone il modello italiano, frammentato e scarsamente reattivo,

* Dottorando di ricerca in Diritto, Scuola superiore di Studi Universitari e di Perfezionamento "Sant'Anna" di Pisa.

dove le competenze sono parcellizzate tra diverse autorità e la tutela in fase pre-elettorale appare insufficiente. Attraverso un esame dettagliato delle procedure e dei risultati, facendo riferimento anche a casi concreti di applicazione delle normative, si tenterà di valutare il grado di efficacia dei due sistemi, domandandosi se la presenza di un'autorità elettorale indipendente e specializzata, come la Jec, costituisca un fattore determinante per garantire una risposta più tempestiva, coerente ed efficace alla disinformazione e di conseguenza preservare l'integrità del procedimento elettorale.

Prima di entrare nel dettaglio sulle specifiche misure adottate dagli ordinamenti in esame, è necessario inquadrare il tema da una prospettiva più ampia, delineando il quadro normativo vigente a livello nazionale, ma anche eurounitario, profilo fondamentale in quanto mira a ridefinire il panorama della regolamentazione della comunicazione politica (e non solo) degli Stati membri. Né in Italia né in Spagna esistono previsioni di legge che definiscano e sanzionino specificamente il reato di disinformazione, come invece accaduto, con difficoltà applicative e interventi giurisprudenziali rilevanti, in Germania (seppur senza specifiche previsioni in materia elettorale) e Francia¹. Nel caso francese, in particolare, il legislatore ha introdotto l'art. L163-1 del Codice elettorale, volto a imporre alle piattaforme digitali di grandi dimensioni obblighi di informazione sull'utilizzo dei dati personali, sull'identità dei finanziatori di contenuti di propaganda politica *online* durante i periodi di campagna elettorale (i tre mesi antecedenti al primo giorno del mese in cui si tengono le elezioni generali). In parallelo, l'art. L163-2 del medesimo Codice mira a contrastare specificamente la disinformazione elettorale *online*, introducendo la possibilità per il giudice dei *référé*s, su richiesta di qualsivoglia soggetto interessato o del pubblico ministero, di imporre alle piattaforme di mettere in atto tutti i provvedimenti necessari per cessare la diffusione di contenuti recanti «*allégations ou imputations inexactes ou trompeuses d'un fait de nature à altérer la sincérité du scrutin*», con possibilità di riconoscere risarcimenti e pronuncia entro 48 ore. La disciplina è stata successivamente oggetto dello scrutinio del *Conseil Constitutionnel*, che ha provveduto a restringerne il perimetro di applicazione e a chiarirne i presupposti².

I tentativi di introdurre norme *ad hoc*, come le proposte di legge “*anti-fake news*” in Italia, si sono spesso scontrati con comprensibili preoccupazioni ri-

1. Si veda M. Foglia, *Le esitazioni del legislatore francese nel contrasto ai discorsi d'odio in rete*, in Aa.Vv., *Liber Amicorum per Pasquale Costanzo*, 2020, pp. 1-11, disponibile su: www.giurcost.org.

2. *Funditus*, P. Passaglia, *La disinformazione, i contenuti illegali e i limiti alla libertà di espressione online: un'inevitabile evaporazione delle garanzie costituzionali?*, in *federalismi.it*, numero speciale, 30 maggio 2025, spec. pp. 37-47

guardanti la libertà di espressione e la difficoltà di delineare un confine netto tra disinformazione volontaria e semplice scorretta informazione o legittima critica politica. Di conseguenza, il contrasto al fenomeno si è basato sull'applicazione di norme preesistenti: in Italia, infatti, si fa riferimento al Codice penale (es., reati di diffamazione o procurato allarme), alla legge sul diritto d'autore e, in misura minore, alla legge n. 28/2000 in materia di *par condicio*, che disciplina l'accesso ai media in condizioni di parità tra gli attori politici, ma non affronta direttamente il contrasto alla diffusione di contenuti falsi. In Spagna, invece, si applicano le disposizioni del Codice penale (es. calunnia, ingiurie), benché siano scarsamente funzionali in una materia che richiede risposte immediate quale quella in esame, e la *Ley Orgánica del Régimen Electoral General* (Loreg), che garantisce l'imparzialità e l'indipendenza dell'amministrazione elettorale³.

È necessario poi ricordare la recente entrata in vigore del *Digital Services Act* (Regolamento UE 2022/2065), che rappresenta la risposta più ambiziosa messa in atto dall'Unione Europea per creare uno spazio digitale più sicuro. Pienamente applicabile per tutte le piattaforme a partire da febbraio 2024, il Dsa introduce obblighi stringenti per i servizi digitali, con particolare attenzione ai cosiddetti *Very Large Online Platforms* (Vlop) e *Very Large Online Search Engines* (Vlose), e sottopone le piattaforme a una disciplina comune in materia di contrasto alla disinformazione anche a fini elettorali. In parallelo, l'*Artificial Intelligence Act* (Regolamento UE 2024/1689), nel mirare a regolamentare l'utilizzo e lo sviluppo dell'intelligenza artificiale nell'Unione, è indirettamente andato a disciplinare una materia che rappresenta la nuova frontiera della disinformazione elettorale, date le sue sconfinata potenzialità, si pensi alla produzione di contenuti *deepfake*. Non è un caso, infatti, se nella classificazione introdotta dall'*AI Act* i sistemi volti a «influenzare l'esito di elezioni o referendum o il comportamento di voto delle persone fisiche nell'esercizio del loro voto alle elezioni o ai referendum» rientrano nella fascia di rischio più alta, con eccezione di quelli che non prevedono un'esposizione diretta dei singoli utenti agli *output* (es. strumenti per organizzare le campagne elettorali da un punto di vista logistico e amministrativo). Tuttavia, l'implementazione e l'*enforcement* di tali regolamenti in materia elettorale specifici dipendono in larga misura dalla funzionalità e dalle competenze delle autorità nazionali. Ed è proprio sotto questo profilo che le differenze tra Italia e Spagna diventano decisive.

3. Per una prospettiva più ampia ed esaustiva sul tema, si veda P. Santolaya Machetti, *Procedimiento y garantías electorales*, Editorial Civitas, 2013.

2. Il modello spagnolo: la *Junta Electoral Central* come baricentro della tutela dell'integrità elettorale

La Spagna ha sviluppato un sistema di organizzazione e giustizia elettorale che si distingue per la sua centralizzazione, tempestività e specializzazione. Il perno di questo sistema è la *Junta Electoral Central* (Jec): un organo amministrativo *sui generis*⁴ imparziale e permanente, istituito dalla Loreg e composto da 13 membri (otto magistrati del *Tribunal Supremo* e cinque professori universitari di diritto, scienze politiche o sociologia in attività nominati dal Congresso dei Deputati⁵). Tale organo si caratterizza per la sua indipendenza dal governo in carica⁶, elemento di cruciale rilevanza per garantirne l'imparzialità e favorirne la credibilità⁷.

Sebbene la Loreg sia stata approvata prima dell'era digitale, la Jec ha interpretato estensivamente le sue competenze di «garantire la trasparenza e l'obiettività del processo elettorale» (art. 8) ricomprendendovi anche le campagne elettorali svolte *online*. I poteri della *Junta* durante la campagna elettorale, che inizia con la convocazione delle votazioni, sono molto ampi: oltre a supervisionare la regolarità del procedimento nel suo complesso, essa ha la competenza di intervenire su qualsiasi elemento che possa mettere a repentaglio l'equità della competizione⁸. Il processo è snello e rapido, adatto alla natura *time-sensitive* delle campagne elettorali: esso prende generalmente avvio da una denuncia presentata alla Jec riguardante una violazione della Loreg posta in essere da esponenti politici o dai media. A seguito di ciò, l'organo si riunisce e si pronuncia a strettissima distanza (24-72 ore). Le pronunce della Jec si sostanziano in ordinanze (*acuerdos*) che ingiungono ai responsabili (partiti, media, piattaforme) di cessare la diffusione dei contenuti in violazione della Loreg o, se del caso, rimuoverli da siti *web* o *account* di

4. Sul punto E. Arnaldo Alcubilla, *La interpretación de la normativa por la Junta Electoral Central* (1977-1997), in *Revista de las Cortes Generales*, n. 41, 1997, p. 173.

5. Sull'evoluzione della composizione cfr. M.V. García Soriano, *Jueces y magistrados en el procedimiento electoral*, Centro de Estudios Políticos y Constitucionales, 2000, spec. p. 23 ss.

6. Si faccia riferimento a M. Satrustegui Gil-Delgado, *Las garantías del derecho electoral*, in *Revista de las Cortes Generales*, n. 20, 1990, p. 98.

7. La disinformazione *online* relativa alle attività della *Junta* è però diffusa, si veda L. Rodríguez-Fernández, S. González-Fernández, S. Arce-García, *Desinformación sobre procesos electorales en España: el caso de la plataforma Elecciones Transparentes*, in *Revista Mediterránea de Comunicación*, n. 2, 2025, spec. p. 9.

8. La nozione di campagna elettorale è pertanto da intendersi in senso restrittivo, si vedano gli *acuerdos* Jec N. 469/2015 e 56/2016, cfr. I. Lauzirika Ansola, *La desinformación en campañas en línea: contexto y consecuencias del Reglamento (ue) sobre transparencia y segmentación en la publicidad política*, in *Legal*, n. 2, 2024, pp. 9-10.

social network, di pubblicare la decisione che smentisce la notizia o di pagare sanzioni pecuniarie. Tali competenze sono integrate dall'art. 68 Loreg, che attribuisce a candidati o dirigenti delle formazioni politiche che concorrano alle elezioni il diritto di richiedere a un giudice (mentre la Jec non è competente in materia, v. in ultimo l'*acuerdo* n. 150/2023) di disporre la rettifica in caso di diffusione su qualsiasi mezzo di comunicazione⁹ di fatti inesatti o idonei ad arrecare loro pregiudizio, con decisione resa in forma orale entro quattro giorni.

Pur non ordinando mai direttamente la rimozione di contenuti in quanto falsi o fuorvianti, dal momento che una simile valutazione sulla veridicità delle informazioni esula dalle competenze della Jec, la giurisprudenza da essa sviluppata appare comunque rilevante ai fini di inquadrare lo strumentario potenzialmente a disposizione del legislatore spagnolo per contrastare il fenomeno della disinformazione. Ad esempio, la *Junta* ha interpretato in modo rigoroso il divieto di propaganda nella "*jornada de reflexión*" (il giorno prima del voto), applicandolo anche ai *social media* e ordinando la rimozione di *post* o messaggi su piattaforme come X o WhatsApp inviati da candidati o partiti in quel giorno; l'art. 53 Loreg, poi, ha esteso espressamente ai media digitali il divieto di propaganda nel periodo precedente alla campagna elettorale. È proprio sotto questo profilo che si può individuare una delle principali differenze a livello di effettività rispetto all'ordinamento italiano: ai sensi dell'art. 9, comma 1, della legge n. 212 del 1956 nel giorno precedente e in quelli stabiliti per le elezioni sono vietate «le riunioni di propaganda elettorale diretta o indiretta». Tale previsione è stata estesa anche alle emittenti radiotelesive private dal d.l. n. 807/1984, ma non contempla i *social media* e Internet in generale, lasciando pertanto un colossale vuoto normativo in materia¹⁰, al punto che l'Agcom stessa nelle linee guida pubblicate in occasione delle elezioni politiche del 2018 si limitava a definire soltanto «auspicabile» che in tale periodo «fosse evitata, da parte dei soggetti politici, ogni forma di propaganda, per evitare di influenzare con pressioni indebite l'elettorato ancora indeciso». Analogo controllo, ai sensi dell'art. 69 Loreg, viene effettuato poi sulla diffusione di sondaggi, con attento scrutinio sulla correttezza dello svolgimento delle operazioni di rilevazione e la fedeltà dei risultati pubblicati ai dati raccolti, con obbligo per i mezzi

9. Tuttavia, la dottrina ha evidenziato i limiti di questo istituto nel momento in cui l'attività di disinformazione è condotta *online* da *bot* o da utenti localizzati in Paesi esteri, essendo la disciplina concepita per mezzi di stampa o radiotelesivi, v. O. Sánchez Muñoz, *La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales*, CEPC – Universidad de Valladolid, 2020, p. 254.

10. Sul punto *amplius* D. Bellacicco, *L'applicabilità della disciplina del silenzio elettorale ai social media*, in *Cammino Diritto*, n. 10, 2021, spec. pp. 6 ss.

di informazione che li abbiano divulgati in violazione della legge di pubblicare entro tre giorni le rettifiche richieste dalla Jec nei medesimi spazi o pagine in cui era presente l'inchiesta originaria¹¹.

La Jec vigila poi attentamente sull'attuazione dell'art. 50.2 della Loreg, che vieta alle istituzioni pubbliche di fare propaganda, sanzionando *leader* politici per l'uso di *account* istituzionali sui social media per promuovere la propria candidatura¹². Inoltre, sebbene la *Junta* sia cauta riguardo alla veridicità delle informazioni oggetto di contestazione, dal momento che esula dalle sue competenze pronunciarsi sul merito delle questioni, essa senz'altro interviene quando la disinformazione viola principi specifici in materia elettorale. Ad esempio, l'organo si è pronunciato sulla copertura informativa delle varie formazioni politiche in condizioni di parità¹³ e sull'utilizzo di marchi e loghi di partito in sede di campagna elettorale¹⁴, nonché, appunto, sulla diffusione di informazioni fuorvianti da parte dei mezzi di comunicazione. È il caso degli *acuerdos* n. 123/2017 e n. 133/2017, che in occasione delle elezioni della *Generalitat* hanno censurato la violazione del principio di neutralità informativa ex art. 66 Loreg da parte dell'emittente televisiva *Corporació Catalana de Mitjans Audiovisuals*, che nelle sue trasmissioni aveva utilizzato i termini «esiliati» e «governo in esilio» con riferimento ai politici coinvolti nella nota vicenda della dichiarazione d'indipendenza della Catalogna, favorendo implicitamente determinati candidati e fuorviando pertanto gli ascoltatori. Tuttavia, la Jec non ha richiesto una ritrattazione esplicita né ha imposto delle sanzioni, ammonendo l'emittente di non reiterare simili violazioni.

La Jec non opera in solitudine: il sistema spagnolo prevede un'architettura integrata dalla *Comisión Nacional de los Mercados y la Competencia* (Cnmc), l'ente di regolamentazione dei media audiovisivi, che collabora con la *Junta* nel vigilare sulla correttezza dell'informazione durante la campagna elettorale, soprattutto per quanto riguarda i media tradizionali, oltre a rappresentare a partire dal 2024 l'autorità responsabile della supervisione delle reti sociali e delle piattaforme digitali prevista dal *Digital Services Act* e dal *Digital Markets Act*. In parallelo, la Jec è coadiuvata a livello territoriale dalle *Juntas* delle Comunità

11. Per la giurisprudenza più recente si vedano in particolare gli *Acuerdos Jec* 464, 465 e 539/2023, cfr. L. Corredoira y Alfonso, *Garantías de la libertad de expresión en redes sociales y medios. Doctrina de la Junta Electoral Central durante los comicios españoles celebrados en 2023*, in *Rivista de Derecho Politico*, n. 5, 2024, p. 228.

12. *Amplius*, L.A. Gálvez Muñoz, *Las campañas institucionales. Normas, Doctrina de la Junta Electoral Central y Jurisprudencia*, Aranzadi, 2010, pp. 23-71.

13. Si vedano gli *acuerdos Jec* n. 162, 273 e 503 del 2023, sui quali *amplius* L. Corredoira y Alfonso, *Garantías de la libertad de expresión en redes sociales y medios*, cit., pp. 228-230.

14. Ivi, pp. 230-231, sugli *acuerdos Jec* N. 261 e 287 del 2023.

Autonome, provinciali e di zona, sebbene con modificazioni a livello di numero di membri (tre magistrati e due professori nelle giunte provinciali e di zona) e di modalità e organi abilitati per la nomina¹⁵.

3. Il modello italiano tra frammentazione istituzionale e limitata reattività

L'approccio italiano al contrasto della disinformazione elettorale è radicalmente diverso da quello spagnolo, essendo caratterizzato da una frammentazione delle competenze e da una minore capacità di intervento rapido e preventivo. Non esistendo un'autorità equivalente alla Jec con poteri così ampi e specifici in materia di contrasto alla disinformazione elettorale, le competenze sono suddivise tra una molteplicità di soggetti.

Anzitutto, l'Autorità per le Garanzie nelle Comunicazioni (Agcom) si occupa della regolazione dei media tradizionali e, sempre più, dei servizi digitali. Essa ha poteri di intervento sulle emittenti televisive e radiofoniche per violazione della *par condicio*; tuttavia, il suo raggio d'azione sui contenuti *online* è più limitato e in materia elettorale si basa su un approccio improntato all'adesione puramente volontaria delle piattaforme secondo codici di autoregolamentazione¹⁶. Uno dei principali limiti di tale organo è rappresentato dalla sua proiezione strettamente nazionale, difficilmente in grado di contrastare efficacemente la diffusione di informazioni scorrette su piattaforme che hanno una natura transnazionale. In parallelo, il Garante per la protezione dei dati personali interviene non sulla veridicità dei contenuti, bensì quando la disinformazione coinvolge il trattamento illecito di dati personali a fini di propaganda politica: tale autorità indipendente può vietare l'utilizzo di elenchi di contatti acquisiti illecitamente (es. da *database* pubblici) per l'invio di messaggi elettorali o sanzionare il *micro-targeting* basato su dati sensibili (opinioni politiche, salute) in violazione del Gdpr. Tuttavia, sebbene di grande rilevanza, il suo intervento è settoriale e non copre la parte più consistente di disinformazione "virale", che non impiega dati personali specifici. Inoltre, i reati di diffamazione, procurato allarme o istigazione all'odio sono di competenza della magistratura ordinaria, con possibilità di ottenere risarcimenti; tuttavia, i tempi della

15. Funditus, D. Camoni, *La Junta Electoral Central tra amministrazione, funzione normativa e control judicial della competizione elettorale spagnola*, in Osservatorio costituzionale AIC, n. 2, 2020, p. 69.

16. Sul punto P. Bonini, *L'autoregolamentazione dei principali Social Network. Una prima ricognizione delle regole sui contenuti politici*, in *federalismi.it*, n. 11, 2020, pp. 265-281.

giustizia civile e penale sono incompatibili con la brevità di una campagna elettorale, dal momento che eventuali sentenze di condanna, giungendo a elezioni concluse, sono comunque inefficaci per tutelare l'integrità delle procedure di voto. Anche la Polizia Postale ha tentato di porre un freno alla diffusione di *fake news*, introducendo sul proprio sito Internet un "pulsante rosso" in grado di permettere agli utenti di segnalare istantaneamente la presenza in Rete di contenuti potenzialmente forieri di disinformazione a esperti appartenenti alle forze dell'ordine, con pubblicazione di smentite ufficiali sui siti e sui canali *social* istituzionali; tuttavia, i limiti di tale strumento appaiono evidenti, non fosse altro che per la genericità dei criteri di valutazione e l'eccessiva indeterminazione delle informazioni oggetto di controllo, nonché per l'insufficiente base normativa dell'intervento¹⁷. Un ruolo particolare è infine giocato dall'Agenzia per la Cybersicurezza Nazionale (Acn), che ha un ruolo di vigilanza sulla sicurezza delle infrastrutture critiche e su eventuali campagne di influenza sulle elezioni provenienti da Paesi terzi, ma non ha poteri diretti sulla gestione dei contenuti.

Questa frammentazione crea un "vuoto di responsabilità" e una mancanza di coordinamento: non è chiaro a chi si possa rivolgere un candidato o una forza politica nel momento in cui voglia denunciare la diffusione di *fake news* confidando in un intervento celere ed efficace, il risultato è spesso l'inazione o un intervento tardivo. L'Italia soffre quindi di una insufficiente estensione della tutela in fase preelettorale, poiché i suoi strumenti più forti (la *par condicio*) si attivano solo a campagna elettorale ufficialmente convocata, mentre quelli deboli (i protocolli) mancano di coerenza.

In assenza di un quadro normativo definito, le autorità italiane hanno tentato di reagire con i limitati strumenti a loro disposizione¹⁸: in primo luogo con le delibere dell'Agcom (coadiuvata a livello regionale dai Corecom), volte a contrastare le pratiche commerciali sleali *online* che sfruttano la disinformazione e a garantire trasparenza in tema di pubblicità politica. Inoltre, il Governo e le altre istituzioni pubbliche hanno promosso protocolli di intesa con le principali piattaforme *social* (Meta, Google, TikTok, X). Esempi sono la delibera 423/17/CONS contenente linee guida per la parità di accesso alle piattaforme *online* per la campagna elettorale per le elezioni politiche del 2018, volte a favorire la trasparenza dei messaggi pubblicitari elettorali e l'eliminazione

17. M. Mezzanotte, *Fake news nelle campagne elettorali digitali. Vecchi rimedi o nuove regole?*, in *federalismi.it*, n. 24, 2018, pp. 17-18.

18. Evidenzia i limiti di tale, pur meritorio, tentativo M. Monti, *Le categorie del costituzionalismo e le fonti alla prova della propaganda politica online: utente-consumatore o utente-cittadino?*, in *DPCE online*, numero speciale, 2021, p. 1294.

dei contenuti illeciti o diffamatori nei confronti di persone o soggetti politici con raccomandazione soprattutto a Facebook e Google di impiegare strumenti di *fact-checking*¹⁹, e il «Codice di autodisciplina per la trasparenza dell'informazione durante le campagne elettorali» promosso in vista delle elezioni europee 2019, nonché vari accordi successivi²⁰. Questi protocolli si basano sulla volontarietà e prevedono che le piattaforme si impegnino a creare canali di comunicazione rapida per la segnalazione di contenuti palesemente falsi (es. *deepfake* di candidati, notizie infondate sulla chiusura dei seggi), promuovere l'informazione ufficiale (es. *prompt* del Ministero dell'Interno su come si vota), rafforzare i sistemi di *fact-checking*, garantire trasparenza sugli annunci di natura politica, contrastare *account* falsi e *bot* e segnalare all'Agenzia per la Cybersicurezza Nazionale attività sospette che possano essere ricondotte a tentativi di influenza straniera. Tuttavia, tali documenti soffrono di limiti intrinseci: non sono vincolanti, la loro attuazione dipende dalla buona volontà delle piattaforme, dal momento che mancano meccanismi di *enforcement* efficaci ed effettivamente sanzionatori, a differenza della Jec.

Altrettanto insufficienti appaiono gli altri strumenti potenzialmente idonei a contrastare la disinformazione mediante la diffusione di contenuti maggiormente attendibili da parte delle istituzioni pubbliche: i limiti della *governance* della Rai sono ben noti e difficilmente superabili fintantoché il condizionamento politico degli organi interni rimarrà invariato, con conseguente perdita di fiducia nell'imparzialità e nell'obiettività del servizio pubblico²¹. Come nel modello spagnolo, risultano poi ampiamente obsolete le previsioni dell'art. 8 della legge n. 47 del 1948 in tema di obbligo di rettifica in capo ai direttori o ai responsabili di giornali che abbiano pubblicato immagini o abbiano attribuito atti o pensieri o affermazioni di soggetti che le abbiano ritenute lesive della loro dignità o contrarie a verità. In parallelo, come è stato suggerito, potrebbe risultare utile a fronteggiare le dinamiche disinformative, soprattutto per le consultazioni referendarie, la predisposizione e diffusione a cura delle istituzioni di brevi documenti esplicativi del contenuto dei quesiti, in modo tale da evitare che i cittadini, sovente privi degli strumenti tecnico-giuridici che ne consentono un'adeguata comprensione, si rivolgano prima-

19. Sul punto si vedano M. Bellezza, F. Frigerio, Italy: *First Attempt to (Self)Regulate the Online Political Propaganda*, in *MediaLaws*, 6 febbraio 2018; F. Sciacchitano, A. Panza, *Fake news e disinformazione online: misure internazionali*, in *MediaLaws*, n. 1, 2020, p. 126.

20. G. Stegher, *Da cittadini elettori a cittadini consumatori: osservazioni sull'importanza di regolare le campagne elettorali sui social media*, in *Nomos – Le attualità nel diritto*, 1, 2023, pp. 11-12.

21. G.E. Vigevari, *I media di servizio pubblico nell'età della rete. Verso un nuovo fondamento costituzionale, tra autonomia e pluralismo*, Giappichelli, 2017, p. 31 ss.

riamente a Internet, dove l'attendibilità delle informazioni reperite non è affatto garantita²².

Non avendo avuto buon esito i vari tentativi di introdurre una disciplina organica in materia di contrasto alla disinformazione²³, la legge sulla *par condicio* (legge n. 28/2000) risulta dunque lo strumento principale per garantire l'equità della competizione nei media tradizionali. Tuttavia, è uno strumento concepito per un'epoca predigitale, nonostante la delibera 60/23/Cons dell'Agcom (e le sue precedenti) abbia esteso alcuni principi (come il divieto di propaganda nel giorno di silenzio) alle piattaforme, ma con poteri di *enforcement* molto blandi: i suoi limiti nell'affrontare la disinformazione sono pertanto evidenti. In primo luogo, essa si applica principalmente ai media tradizionali e non alle piattaforme *online*, nonostante in dottrina sia stata suggerita una sua applicazione estensiva ai *social network* e ai motori di ricerca riconducibili a società con sede in Italia²⁴; inoltre, essa è focalizzata sull'equità di accesso (spazi, tempi) più che sulla veridicità dei contenuti, cosicché i suoi interventi sono spesso percepiti come troppo lenti e poco incisivi rispetto alla viralità e alla velocità di diffusione delle *fake news*.

Bisogna infine ricordare che la recentissima legge n. 132 del 2025, volta ad adeguare l'ordinamento italiano alle previsioni dell'*AI Act*, introduce all'art. 26 la nuova fattispecie di reato di illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale (art. 612-*quater* c.p.), con reclusione da uno a cinque anni per chiunque arrechi «un danno ingiusto a una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità». Inoltre, dopo aver introdotto all'art. 61 c.p., n. 11-*undecies*), una circostanza aggravante comune consistente nella commissione del fatto mediante sistemi di intelligenza artificiale tali da costituire un mezzo insidioso, da ostacolare la pubblica o privata difesa o da acuire le conseguenze del reato, all'art. 294 c.p. si prevede al secondo comma la reclusione da due e sei anni se l'inganno volto

22. I. Spadaro, *Contrasto alle fake news e tutela della democrazia*, in *Ianus*, n. 20, 2019, pp. 50-52.

23. Si fa riferimento qui a varie proposte di legge presentate sia nella XVII sia nella XVIII Legislatura da rappresentanti di vari partiti, *funditus* E. Lehner, *Fake news e democrazia*, in *MediaLaws*, n. 1, 2019, pp. 7-8, criticate per possibili violazioni dei principi di determinatezza, tassatività e offensività delle fattispecie di reato da M. Fumo, *Bufale elettroniche, repressione penale e democrazia*, in *MediaLaws*, n. 1, 2018, pp. 83-91; N. Zanon, *Fake news e diffusione dei social media: abbiamo bisogno di un "Autorità Pubblica della Verità"?*, in *MediaLaws*, n. 1, 2018, pp. 12-17.

24. I. Spadaro, *Contrasto alle fake news e tutela della democrazia*, cit., pp. 49-50.

a rappresentare un attentato contro i diritti politici del cittadino sia posto in essere tramite sistemi di IA. Tali disposizioni sono affiancate nell'art. 24, comma 5, da una delega al Governo a predisporre «strumenti, anche cautelari, finalizzati a inibire la diffusione e a rimuovere contenuti generati illecitamente anche con sistemi di intelligenza artificiale, assistiti da un sistema di sanzioni effettive, proporzionate e dissuasive».

Pur non trattando direttamente la materia elettorale, appare rilevante la previsione all'art. 612-*quater* della procedibilità d'ufficio, eccezionale rispetto alla generale punibilità a querela della persona offesa, qualora il delitto sia commesso nei confronti di «una pubblica autorità a causa delle funzioni esercitate». Questa disposizione sembra pertanto orientata alla diffusione del fenomeno dei *deepfake sub specie* di fittizia comunicazione istituzionale, come già accaduto recentemente con la diffusione di false interviste televisive e video manipolati di esponenti del Governo, tra cui la Presidente del Consiglio, *leader* politici e alti funzionari pubblici al fine di promuovere piattaforme abusive di *trading online*: le nozioni di «pubblica autorità» e di «a causa delle funzioni esercitate» appaiono difficili da inquadrare²⁵, ma al contempo non troppo idonee a essere interpretate estensivamente, considerando il divieto di applicazione analogica delle norme penali, ai casi di disinformazione elettorale posti in essere tramite *deepfake* che non riguardino direttamente coloro che ricoprono cariche pubbliche. Pertanto, non pare nella sfera applicativa di tale disposizione possa rientrare la diffusione di contenuti fittizi generati con IA. aventi come soggetto semplici parlamentari o esponenti di partiti politici che non esercitano funzioni di governo.

Ben più interessante, e forse idonea a porre le basi per una disciplina italiana anti-*fake news* in materia elettorale, sebbene sempre nei limiti imposti dal divieto di analogia in materia penale, è la modifica dell'art. 294 c.p.: nella fattispecie di attentato contro i diritti politici del cittadino posto in essere mediante sistemi di intelligenza artificiale potrebbero ben rientrare le attività di disinformazione in prossimità di votazioni. Questo intervento di riforma si scontra con la giurisprudenza di legittimità in tema: fino a oggi, infatti, la disposizione è sem-

25. Pare, infatti, che per «pubblica autorità» si debba intendere qualsiasi ente pubblico che eserciti specifiche funzioni di natura amministrativa, anche se il termine in quanto tale ricorre nell'art. 403 c.c. (intervento della pubblica autorità a favore dei minori moralmente o materialmente abbandonati o in pericolo nell'ambiente familiare), riferendosi solo a sindaci, servizi sociosanitari e autorità di pubblica sicurezza, mentre nell'art. 504 c.p., sebbene solo in rubrica, in tema di «coazione alla pubblica Autorità mediante serrata o sciopero» il destinatario sembra essere più generalmente individuabile in tutti gli organi pubblici con poteri deliberativi, indipendentemente dalla loro natura monocratica o collegiale, nazionale o locale. La formulazione testuale non agevola pertanto l'esatta delimitazione dei concetti.

pre stata applicata con riferimento a fattispecie caratterizzate dalla compressione del libero esercizio dei diritti di elettorato attivo e, soprattutto, passivo mediante violenza, minaccia o inganno²⁶. La Cassazione, infatti, oltre ad aver chiarito che «la nozione di “diritti politici”²⁷, il cui libero esercizio è tutelato dall’art. 294 c.p., deve essere intesa in senso stretto, con esclusione di qualsiasi manifestazione del proprio pensiero che possa riguardare scelte politiche» (Cass. pen., Sez. VI, sent. 09/11/2016, n. 51722), si è in passato pronunciata sulla necessità che la pressione esercitata sull’elettore ai fini di modificare la propria determinazione in sede di votazione in modo contrario alla sua volontà sia effettivamente intensa, non essendo sufficienti a tal fine «le semplici suggestioni, le promesse chimeriche, le forzature dialettiche, le prospettazioni incomplete o tendenziose di situazioni nazionali o locali e le interpretazioni faziose di eventi» (Cass. pen., Sez. I, sent. 20/12/2018, n. 16381; v. anche Cass. pen., Sez. I, sent. 26/06/1989, n. 11835, con specifico riferimento al caso in cui, nel corso di una trasmissione televisiva, il cantante Adriano Celentano invitò gli elettori ad aggiungere la frase «La caccia è contro l’amore: non la vogliamo», esprimendo scarso interesse per i quesiti sottoposti a referendum abrogativo e ritenendo prioritario il tema della caccia, non oggetto della consultazione, anche a costo di invalidare la scheda e influenzare così l’esito della votazione.

Mantenendo però immutata l’attuale interpretazione della nozione di “attentato”, non è agevole comprendere quale portata applicativa possa avere l’introduzione del secondo comma dell’art. 294 c.p., in quanto difficilmente si comprende come tramite sistemi di intelligenza artificiale si possa «esercitare sull’elettore una pressione di tale intensità da indurlo a determinarsi, nell’esercizio di un diritto politico in modo contrario alla sua reale volontà» (Cass. pen., Sez. I, sent. 26/06/1989, n. 11835). Infatti, se sicuramente mediante I.A. si può tentare di suggestionare l’elettore o di modificarne la manifestazione del pensiero politico e la sua espressione alle urne, molto più arduo sembra produrre i medesimi effetti di violenza e minaccia.

26. Sul tema si veda anche L. Buscema, *Reati elettorali e principio di democraticità dell’ordinamento: profili assiologici e ricostruttivi*, in *Diritto Penale Contemporaneo*, 28 ottobre 2013, spec. pp. 7-9.

27. Per «diritti politici» si intendono «quelli che permettono al cittadino di partecipare all’organizzazione e al funzionamento dello Stato e degli altri enti di rilevanza costituzionale, come le regioni, le province e i comuni, ai quali è attribuita la funzione di indirizzo politico in relazione a un determinato aggregato di persone stanziate su una parte del territorio» (Cass. pen., Sez. I, sent. 14/10/1993, n. 11055).

4. Una valutazione dell'efficacia dei due modelli in prospettiva comparata

Confrontando i due modelli, emergono chiari punti di forza e di debolezza che influenzano direttamente la loro efficacia nel contrastare la disinformazione elettorale. Come si è notato *supra*, il modello spagnolo è caratterizzato da una gestione centralizzata, integrata e specializzata a livello istituzionale, con la *Junta Electoral Central* come baricentro del sistema, a fronte della frammentarietà, parcellizzazione e plurisoggettività dell'ordinamento italiano, che sovente non riesce a intervenire in modo reattivo e tempestivo a fronte della commissione di illeciti, soprattutto se paragonato alla celerità e alla pervasività degli interventi della Jec. Tale divario si apprezza anche considerando la natura e l'efficacia delle decisioni: se quelle spagnole sono caratterizzate da una fortissima rilevanza a livello politico e da una sistematicità tale da poter consolidare una vera e propria giurisprudenza in materia di contrasto alla disinformazione, con conseguente maggior prevedibilità delle pronunce della Jec, quelle italiane sono generalmente caratterizzate dall'assenza di vincolatività e dalla difficoltà di garantire una rapida attuazione di delibere e protocolli, lasciando spesso tanto le istituzioni quanto i soggetti colpiti dalla diffusione di *fake news* privi di effettiva tutela. La giustizia ordinaria è, infatti, inevitabilmente troppo lenta, e gli interventi dell'Agcom non sono sempre sufficientemente rapidi per un contesto scandito da scadenze assai ravvicinate quale quello elettorale, creando un senso di impunità per i diffusori di disinformazione. Al contempo, nel modello italiano vengono meno anche l'effetto deterrente nei confronti di quanti intendano promuovere campagne di divulgazione di *fake news*, nonché la pressione nei confronti delle piattaforme affinché approntino meccanismi interni idonei ad agevolare il contrasto a tali dinamiche, dal momento che, prescindendo dagli obblighi imposti dal *Digital Services Act*, i semplici protocolli volontari non risultano sufficienti a tal fine, dal momento che spostano parte della responsabilità di contrasto alla disinformazione a soggetti privati i cui interessi non sempre coincidono con quelli pubblici. Al contrario, pur non potendosi trascurare la natura transnazionale dell'operato dei *social media*, risulta indubbiamente più efficace sotto tale profilo la presenza di istruzioni chiare da parte di un'autorità nazionale centrale univocamente definita, con maggior certezza sotto il profilo degli organi competenti, delle tempistiche e dei rimedi disponibili. L'assenza di un'unica autorità referente porta pertanto a sovrapposizioni o, più spesso, a vuoti di intervento; la mancanza di sanzioni rende poi il sistema più debole.

La capacità della Jec di pronunciarsi in poche ore è il suo principale punto di forza: in una campagna elettorale, infatti, un intervento correttivo dopo

sole 48 ore può essere sufficiente a limitare il danno di una *fake news* virale, a maggior ragione considerando il funzionamento degli algoritmi che gestiscono la presentazione dei contenuti nei *feed*, i quali danno la priorità a contenuti pubblicati più recentemente. La composizione indipendente e altamente qualificata della *Junta* conferisce poi alle sue decisioni una dimensione istituzionale e tecnica difficilmente contestabile, in grado di fornire linee guida chiare a tutti gli attori sulle forme e le modalità lecite e illecite di propaganda elettorale. Pur in assenza di una giurisprudenza consolidata in materia, l'ordinamento spagnolo appare maggiormente preparato rispetto a quello italiano in tema di contrasto alla disinformazione, essendo dotato di un quadro normativo più definito e di un'autorità centrale in possesso delle competenze necessarie per intervenire.

L'analisi comparata tra Italia e Spagna dimostra che, di fronte alla minaccia della disinformazione elettorale, la presenza di un'autorità centrale, indipendente e specializzata costituisce un fattore di resilienza di cruciale importanza. L'entrata in vigore del *Digital Services Act* offre a entrambi i paesi strumenti assai incisivi, ma non è idonea a colmare autonomamente le carenze strutturali. Per l'Italia, la sfida più urgente nel futuro prossimo sarà colmare il suo *deficit* di *governance* in materia, non necessariamente creando un clone della Jec, ma sicuramente predisponendo una chiara attribuzione di competenze, un coordinamento rafforzato e permanente tra Agcom, Garante Privacy, Acn e autorità giudiziaria, e forse l'estensione dei poteri di uno di questi soggetti per raggiungere, per quanto possibile, la tempestività e l'autorevolezza del modello spagnolo. In parallelo, resta del tutto aperta la questione delle modalità di contrasto alla disinformazione elettorale sotto il profilo penalistico, non risultando risolutive le recenti modifiche apportate con la legge n. 132/2025. In un'epoca in cui la fiducia nelle istituzioni è sempre meno scontata, la capacità di garantire elezioni libere, eque e informate non è più un'opzione, ma una condizione essenziale per la sopravvivenza della democrazia stessa.

COMUNICAZIONE E NUOVE TECNOLOGIE: LA PROPAGANDA POLITICA ONLINE NELLA COSTITUZIONE

*Allegra Dominici**

SOMMARIO: 1. L'essenza costituzionale della comunicazione politica: il diritto all'informazione – 2. *Social media* e Costituzione: brevi cenni – 3. L'attualità della normativa italiana attorno alle problematiche dell'era digitale – 4. Riflessioni conclusive: una risposta condivisa a un problema comune.

1. L'essenza costituzionale della comunicazione politica: il diritto all'informazione

L'evoluzione digitale, caratterizzata dall'affermazione di internet quale sistema comunicativo globale, ha profondamente inciso nelle dinamiche di interazione tra istituzioni politiche e cittadinanza.

Una metamorfosi tecnologica che ha determinato una significativa ridefinizione degli assetti comunicativi tradizionali, generando un duplice esito: da un lato, l'ampliamento e la capillarizzazione dei canali di comunicazione politica, dall'altro, una complessa ridefinizione degli equilibri istituzionali e costituzionali¹. Le caratteristiche proprie dei *social media* – la facilità e la velocità con cui i contenuti raggiungono gli utenti, meccanismi di profilazione individuale e tecniche comunicative persuasive – impiegati per finalità di rilievo politico-istituzionale producono un effetto distorsivo sul circuito rappresentativo previsto dalla Costituzione.

Con il presente contributo si tenterà dunque di fornire una panoramica

* Dottoranda di ricerca in Scienze giuridiche, Università degli Studi di Urbino Carlo Bo.

1. Cfr. G. Pitruzzella, *La libertà di informazione nell'era di internet*, in *MediaLaws*, n. 1/2018; secondo l'A., citando Yochai Benkler, sono tre i caratteri principali di innovazione di questa "nuova era dell'informazione": a) produzione dell'informazione decentrata che consente a chiunque di divenire produttore di informazioni; b) l'accessibilità dei device elettronici; c) il carattere globale dei flussi di informazioni.

delle disposizioni costituzionali maggiormente coinvolte (2), alla luce di una normativa interna che limita le sue attenzioni a tematiche non più attuali (3), fornendo infine delle riflessioni conclusive sul ruolo e lo spazio riservato al decisore politico nell'ambito della comunicazione politica (4).

Prima di volgere lo sguardo all'analisi circa le disposizioni costituzionali coinvolte dall'utilizzo di tali strumenti in ambito di comunicazione politica, pare utile soffermarsi, seppur brevemente, sull'essenza costituzionale che possiede tale settore, il cui rilievo prescinde le implicazioni derivanti dall'utilizzo delle nuove tecnologie.

Sebbene non reso esplicito nel Testo², una costante giurisprudenza³ e una copiosa dottrina⁴, hanno rinvenuto nella disposizione di cui l'articolo 21 della Costituzione il fondamento di ciò che è stato definito come libertà-diritto all'informazione, giocoforza una lettura in senso funzionalista della disposizione in forza della quale si è giunti a delineare una situazione soggettiva declinabile nel diritto a ricevere un'informazione corretta, quale presupposto indeffettibile per il funzionamento del circuito rappresentativo⁵.

2. Necessario sottolineare la rilevanza che il diritto all'informazione possiede in riferimenti normativi sovranazionali, i quali entrano a far parte della Costituzione attraverso gli articoli 10 e 117, co. I Cost; si fa particolare riferimento all'articolo 19 della Dichiarazione universale dei diritti dell'uomo adottata il 10 dicembre 1948 dall'Onu nella parte in cui prevede accanto alla libertà di manifestare liberamente il proprio pensiero, il diritto di ricercare, ricevere e diffondere notizie e informazioni; e all'articolo 10 della Convenzione Europea dei Diritti dell'Uomo del 1950.

3. Sentt. nn. 59/1960; 122/1970; 202/1970; 105/1972; 1/1981; 194/1987; 826/1988.

4. «L'essere compiutamente e realmente informati» si configura come «premessa indispensabile del buon funzionamento del regime liberal-democratico» così L. Paladin, *Problemi e vicende della libertà di informazione nell'ordinamento italiano*, in L. Paladin (a cura di), *La libertà di informazione*, Utet, 1979; sul tema anche C. Esposito, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Giuffrè, 1958; P. Barile, *Libertà di manifestazione del pensiero*, Giuffrè, 1975.

5. Si parla in tal senso di “dimensione attiva” per definire la tutela che l'articolo 21 Costituzione riserva alle espressioni del pensiero individuale, assicurandone la privazione da ingerenze esterne. Rientra in tale filone la celebre sentenza d'esordio della Consulta, sent. n. 1 del 1956 Corte cost., in cui i Giudici, sancendo l'illegittimità dell'art. 113 r.d. 773/1931 nella parte in cui vietava «*Senza licenza dell'autorità locale di pubblica sicurezza distribuire o mettere in circolazione, in luogo pubblico o aperto al pubblico scritti o disegni*» assicurando in tal modo l'assoluta libertà in capo agli individui di esprimere, in qualsivoglia maniera, il proprio pensiero; sulla sentenza cfr. G. Grasso, *La sentenza n. 1 del 1956 Sessant'anni dopo*, in *Osservatorio costituzionale*, n. 1/2017; sul tema anche V. Crisafulli, *Problematica della libertà di informazione*, in *Il Politico*, 1964, secondo il quale «In regime democratico, la libertà di espressione (è) suscettibile di acquistare un più spiccato rilievo politico sino a potersi configurare [...] anche come strumentale rispetto all'esercizio di diritti, tradizionalmente annoverati tra quelli politici (diritto di voto, di referendum, di associarsi in partiti)». «In regime democratico, la libertà di espressione [è] suscettibile di acquistare un più spiccato rilievo politico sino a potersi configurare [...] anche

Il riconoscimento di tale situazione soggettiva di vantaggio al disposto di cui l'articolo 21 della Costituzione pone alcuni problemi critici di difficile risoluzione: l'assolutezza del tenore testuale della disposizione costituzionale di cui si discute, infatti, pare imporre il divieto di obblighi di verità o divieti di menzogna. Tuttavia, posta la particolare rilevanza della comunicazione nel settore politico, ci si chiede, se tale contesto possa continuare a ricevere un così ampio spettro di tutela dinanzi alle necessità di tutela portate dalla rivoluzione digitale.

Nel settore politico, infatti, la correttezza dell'informazione attingibile dagli utenti-elettori diviene il requisito essenziale per poter consentire il funzionamento della rappresentanza istituzionale e anche il funzionamento dell'intero circuito rappresentativo. La correttezza e la completezza dell'informazione pubblica, consentendo la formazione cosciente e completa del pensiero individuale attorno ai temi politici rilevanti, viene a essere un presupposto necessario di costituzionale rilevanza in quanto da quest'ultima deriva, conseguentemente, la determinazione elettorale degli individui.

L'imprescindibile esigenza di garantire un'informazione politica quanto più possibile cristallina si fa ancora più stringente a seguito dell'utilizzo dei nuovi sistemi comunicativi, i quali, fondati sull'applicazione di sistemi di *micro-targeting* dell'elettorato⁶, ripropongono al medesimo utente contenuti simili a quelli con cui ha interagito in precedenza⁷.

Un espediente dapprima confinato nel settore commerciale la cui applicazione in ambito politico risulta particolarmente pericolosa, in quanto produce il rafforzamento di convincimenti e idee già radicate, annullando il pluralismo informativo, risultato dell'incontro di idee e pensieri tra loro differenti⁸. In tal

come strumentale rispetto all'esercizio di diritti, tradizionalmente annoverati tra quelli politici (diritto di voto, di referendum, di associarsi in partiti)», così V. Crisafulli, *Problematica della libertà di informazione*, in *Il Politico*, 1964.

6. Il marketing politico-elettorale quale insieme di tecniche utilizzate al fine di far conoscere un candidato, distinguendolo dai concorrenti e favorirne il successo elettorale si è sviluppato prendendo le mosse dalle tecniche di marketing commerciale, «con l'effetto, però, di produrre un'equiparazione tra prodotto politico e prodotto da vendere», così M.R. Allegri, *Oltre la par condicio. Comunicazione politico-elettorale nei social media, fra diritto e autodisciplina*, FrancoAngeli, 2020.

7. I. Rizzato, *Le nuove frontiere del "digital marketing": dalla profilazione alla manipolazione "online" nell'ambito politico alla luce del Gdpr*, in *Cyberspazio e Diritto*, n. 1/2018; B. Caravita, *Di fronte a un mondo che cambia, chi è più pericoloso Trump o Zuckerberg?*, in A. Poggi, F. Fabrizzi, F. Savastano (a cura di), *Social Network, formazione del consenso, intelligenza artificiale. Itinerario di un percorso di ricerca di Beniamino Caravita*, Sapienza Università Editrice, 2023.

8. «La politica è l'arte di governare la cosa pubblica, non è pensabile che propria per questa arte prettamente umana, vi sia unanimità di vedute. Vi sono invece contrasti, dispareri e dissaccordi e allora è un bene che i cittadini possano, non solo esprimere la loro diversità di opi-

modo l'elettore viene a essere isolato nella sua bolla senza possibilità di ricevere informazioni alternative rispetto ai propri convincimenti⁹.

A fronte di un tessuto costituzionale fondato sulle garanzie poste a presidio della libertà di autodeterminazione elettorale, si paventa, a oggi, uno scenario quantomeno distopico. Lo svolgimento delle ultime campagne elettorali tramite *social* pare mostrare, sotto questo punto di vista, due specifiche criticità: da un lato, una proliferazione di contenuti distorti, diffusi dalle pagine dei partiti politici al fine di spostare consenso, e, dall'altro lato, la costante interazione tra eletti ed elettori, ha sfumato gli intervalli temporali delle campagne elettorali, creando ciò che viene a essere definito un regime di c.d. *permanent campaign*.

Un riferimento contestuale dai risvolti complessi che ci impone di riflettere circa la fisionomia critica che le novità digitali, calate nel panorama di rappresentanza costituzionale, possono assumere senza un coerente presidio normativo.

2. Social media e Costituzione: brevi cenni

Posto l'innegabile rilievo costituzionale del settore comunicativo in ambito politico, le caratteristiche intrinseche ai nuovi sistemi comunicativi rischiano di mettere in pericolo alcuni presidi costituzionali individuali e strutturali.

Tentando di ricondurre la problematica della propaganda *online* nell'ambito di precise disposizioni costituzionali potenzialmente coinvolte, pare utile concentrare l'attenzione su due profili di criticità: da un punto di vista delle garanzie individuali, e, dall'altro lato, sulla struttura e il funzionamento dei tradizionali paradigmi della rappresentanza.

Nella misura in cui l'informazione politica viene a essere individuata quale presupposto necessario alla formazione della coscienza dell'elettorato e, dunque, all'autodeterminazione dei singoli elettori, si ritiene possibile individuare l'esercizio del diritto di voto quale situazione soggettiva potenzialmente idonea a subire le criticità derivanti dall'utilizzo s-regolato dei nuovi canali comunicativi.

L'articolo 48 della Costituzione definisce il voto nella sua duplice accezione di "diritto e dovere", il cui esercizio viene «circondato da una serie di garan-

nione sui mezzi più opportuni per raggiungere il bene comune», così U. Merlin, P. Mancini, *Relazione sulle Libertà Politiche, Commissione per la Costituzione, I Sottocommissione*.

9. L. Califano, *Autodeterminazione vs. eterodeterminazione dell'elettore: voto "privacy" e "social network"*, in *federalismi.it*, n. 16/2019.

zie costituzionali che ne connotano il significato in termini sostanziali»¹⁰; si fa riferimento al disposto di cui il secondo comma della disposizione che caratterizza il voto quale personale, eguale, libero e segreto¹¹.

L'affermazione della libertà del voto, in particolare, definisce quest'ultimo quale esito di un processo determinativo privo di intromissioni potenzialmente idonee a incidere o deviare la formazione della volontà dell'individuo. A tal fine, si profilano due differenti categorie di norme giuridiche volte ad assicurare l'effettività della libertà del voto. Da un lato, infatti, tale requisito viene garantito da un sistema di sanzioni penali volte a contrastare comportamenti e/o atteggiamenti più o meno intrusivi e incidenti nella sfera decisionale dell'individuo¹²; dall'altro lato, si struttura l'apparato normativo volto a tutelare la «corretta prospettazione delle opportunità all'interno delle quali l'avente diritto al voto è invitato a scegliere»¹³, ed è proprio in tale contesto che trovano ragion d'essere le norme volte a disciplinare il fenomeno della comunicazione politica¹⁴ di cui si dirà nel paragrafo successivo.

Come brevemente accennato, la portata di tali innovative modalità comunicative rende impossibile relegare le conseguenze della propaganda politica via *social* a una dimensione esclusivamente riferita al singolo, parendo invero essere idonea a produrre effetti anche per ciò che riguarda il sistema delle forme di rappresentanza così come previsto dalla Costituzione¹⁵.

Muovendo dalla tradizionale finalità affidatagli dal Costituente, l'articolo

10. Così E. Grosso, *Articolo 48*, cit.; M. Rubechi, *Articolo 48*, in F. Clementi, L. Cuocolo, F. Rosa, G.E. Vigevari, *op. ult. cit.*

11. F. Lanchester, *Voto (diritto di)*, in *Enciclopedia del diritto*, XLVI, Giuffrè, 1993, 1106 ss.; M. Rubechi, *Il diritto di voto. Profili costituzionali e prospettive evolutive*, Giappichelli, 2016, p. 209 ss.

12. In tal senso le pene previste dall'art 294 c.p. e dagli artt. 97 e 100 del T.U. n. 361 del 1957, si v. sul punto M. Mazzanti, *Reati elettorali*, in *Enciclopedia del diritto*, XIV, Milano, 1965, 806 s. in cui l'A. suole suddividere il *genus* dei reati elettorali in diverse categorie quali «a) ipotesi generali di reato; b) reati concernenti la formazione delle liste elettorali e altri atti elettorali; c) reati riguardanti l'esercizio del diritto elettorale; d) reati relativi alla genuinità e all'esattezza dei risultati elettorali; d) reati concernenti la propaganda elettorale».

13. Così F. Lanchester, *Voto (diritto di)*, cit.; secondo l'A. tali opportunità di scelta non devono essere prospettate né in maniera distorta né in modo contraddittorio «sibbene [devono essere] chiare e comprensibili»; «nel concetto di libertà di voto» prosegue l'A. «si inserisce, dunque, a pieno titolo la fase preparatoria dello stesso».

14. *Infra*, § 3.

15. «La disciplina costituzionale dei partiti incide direttamente sullo svolgimento del principio di sovranità popolare espresso dall'articolo 1 e sull'opzione politica contenuta nel secondo comma dell'art. 3 Cost. [...] Il collegamento tra tali disposizioni e la garanzia dell'art. 49 offre uno dei principali tramiti per interpretare il principio della sovranità popolare secondo una opzione pluralista fondata sul libero confronto-competizione», così G. Rizzoni, *Articolo 49*, in R. Bifulco, A. Celotto, M. Olivetti, *Commentario alla Costituzione*, Utet, 2006, 981 ss.

49 della Costituzione italiana designa il partito politico quale associazione a cui liberamente i cittadini possono aderire, nonché sede qualificata per l'intermediazione tra società e istituzioni politiche. Sebbene nell'epoca in cui tale disposizione si colloca, i partiti assolvevano a tale indispensabile compito di intermediazione, pare potersi dire che, a oggi, con l'avvento dei nuovi canali di comunicazione, le necessità che hanno condotto a siffatta previsione costituzionale siano quantomeno mutate¹⁶.

Una «democrazia assolutamente diretta», infatti, si paventa quale contrapposizione alla fisionomia rappresentativa dei corpi intermedi¹⁷.

Alla luce di una più attenta analisi due paiono essere le modalità attraverso cui i partiti hanno “interagito” con le nuove tecnologie nel corso degli ultimi anni.

Da un lato, lo scenario tipico delle realtà partitiche di matrice tradizionale è caratterizzato da un utilizzo strumentale dei nuovi canali comunicativi; i *social media* vengono utilizzati quali meri metodi di interlocuzione controllata con i propri elettori, sfruttando la vicinanza che tali strumenti consentono, al fine di un maggior coinvolgimento con l'elettorato che non si esaurisce nei periodi elettorali ma si estende nell'ambito di tutti i procedimenti decisionali e programmatici del partito. La storicità di tali realtà partitiche è rinvenibile, in tal senso, nel mantenimento di una tradizionale dinamica *top-down* che non presume una modifica degli equilibri di potere, cosa che invece accade con i partiti politici di nuovo conio¹⁸. Dall'altro lato, infatti, a proposito delle nuo-

16. La dottrina in materia suole distinguere la tradizionale figura dei partiti di massa, quali rappresentanti di gruppi portatori di ideali comuni, tipici del periodo liberale; con la nuova fisionomia assunta da tale peculiare associazione nel corso del XXI secolo, giungendosi a parlare di “partiti elettorali”, nati dalla rottura tra idee e ideologie nonché dalla neo-fluidità del corpo sociale, non più suddiviso in classi portatrici di interessi differenti e contrapposti, caratterizzati dal rafforzamento del ruolo e dalla centralità assunta dalla figura del leader che impersonifica l'intera identità del partito; foriero di tale evoluzione, l'avvento dei nuovi mezzi comunicativi quali la radio e la rete televisiva; cfr. sul punto F. Raniolo, V. Tarditi, *La rivoluzione digitale e la trasformazione dei partiti*, in *Rivista di Digital Politics*, n. 2/2021; M. Calise, *Il destino dei partiti*, in *Quaderni Costituzionali*, n. 1/2023; M. Calise, *I tecnopartiti*, in *Rivista di Digital Politics*, n. 2/2021. S. Rodotà, *Iperdemocrazia: come cambia la sovranità democratica con il web*, cit. indica come «il partito politico nell'era preteleviva, si presentava come il protagonista di una comunicazione politica diretta (assemblee pubbliche, comizi, contatti continui dei membri del partito con i cittadini) e soprattutto corale e, quindi, in certa misura spersonalizzata. Il partito politico»; aggiunge l'autore «aveva bisogno di stabilire una molteplicità di contatti nello spazio e nel tempo» a differenza di oggi, in cui le molteplici tecniche che consentono al politico «di disporre di strumenti che cancellano quella dimensione spaziale e temporale».

17. Così F. Raniolo V. Tarditi, *La rivoluzione digitale e la trasformazione dei partiti*, cit.

18. Si parla, in tal senso, di partiti digitali per descrivere i partiti antisistema o movimenti politici nati nel corso degli ultimi decenni, facendo leva su propagande di opposizione fondate

ve realtà partitiche, è possibile individuare un utilizzo strutturale delle nuove tecnologie, la cui finalità trascende l'ampiamiento delle modalità attraverso cui si raggiunge l'elettore, giungendo sino a diventare lo strumento attraverso cui il partito si genera e si rinnova, in una costante dinamica di dialogo con il corpo elettorale. In tali contesti, le nuove tecnologie diventano gli strumenti strategici attraverso cui si attua la visione di partito¹⁹. A differenza di ciò che avviene rispetto al modello partitico più ancorato alle modalità tradizionali, la comunicazione non si esaurisce in una dinamica verticale classica, ma si fonda nella strutturazione di una dialettica orizzontale che fa sentire l'elettore-utente parte integrante del processo identitario e decisionale del partito-movimento politico. Dinamiche comunicative che ridefiniscono le modalità di interazione tra rappresentanti e rappresentati attraverso un utilizzo strutturale dei canali di raffronto, sempre più fruibili e meno controllati.

3. L'attualità della normativa italiana attorno alle problematiche dell'era digitale

Nonostante, dunque, la rilevanza costituzionale delle implicazioni derivanti dall'utilizzo dei social ai fini di propaganda elettorale, la disciplina italiana in materia di comunicazione politica presenta un quadro normativo ormai risalente, inidonea a rispondere alle esigenze portate dall'utilizzo di questi strumenti.

La «problematica convivenza tra politica e media»²⁰ fino a qualche decennio fa esauriva la sua potenzialità lesiva nella necessità di regolamentare l'utilizzo dell'unico canale mediatico – quale la televisione – attraverso il quale i partiti politici potevano svolgere la loro naturale attività pubblicitaria²¹.

su modalità comunicative incentrate sui canali web e social; cfr. sul punto M. Bassini, *Partiti, tecnologie e crisi della rappresentanza democratica. Brevi osservazioni introduttive*, cit.

19. «Nella e-democracy la digitalizzazione ha mutato la percezione degli individui verso le nuove tecnologie ormai considerate parte dei processi decisionali fino a condizionarli; ed è in questo passaggio di natura delle nuove tecnologie, da mero supporto tecnologico dell'azione amministrativa a co-artefice della decisione politica, che si cela l'insidia per la quale si pone la delicata questione del rapporto tra modalità di formazione del consenso, impatto delle tecniche di profilazione algoritmica e proprietà privata delle piattaforme telematiche», così M. Calamo Specchia, *Dalla democrazia indiretta alla democrazia in diretta: l'impatto delle nuove tecnologie sulla crisi della democrazia rappresentativa*, in *Diritto Pubblico Europeo Rassegna online*, n. 2, 2023.

20. G. Sirianni, *Par condicio: i complessi rapporti tra potere politico e potere televisivo*, in *Politica del diritto*, n. 3, 2005.

21. Particolare meritevole di attenzione è la sentenza della Corte costituzionale n. 48 del

Posto che l'apparato normativo allora predisposto si inseriva in un contesto tecnologico che vedeva nella programmazione televisiva l'unico e principale canale attraverso cui si svolgeva il dibattito politico; esso si occupava principalmente di garantire le contrapposte esigenze di «assicurare una morale parità delle opinioni»²², prevedendo la necessaria compartecipazione di esponenti di forze politiche opposte all'interno dei programmi televisivi, pur lasciando spazi di libertà nelle decisioni relative ai programmi televisivi delle emittenti. La normativa si occupava dunque di garantire che l'espressione del voto avvenisse in maniera "libera", ossia attraverso la formazione di convincimenti personali costruiti in maniera corretta e consapevole.

Due le principali leggi adottate a tal fine: la normativa n. 28 del 2000 e la precedente n. 515 del 1993²³; entrambe accomunate dalla finalità di regolamentare l'accesso ai mezzi di comunicazione radio-televisivi da parte delle forze politiche, ma che non considera – in ragione dell'epoca del suo intervento – le problematiche relative a internet.

Nonostante l'omogeneità dei fini, le due leggi divergono in alcuni aspetti contenutistici, in particolare, solo la legge n. 28 del 2000, nel tentativo di sistematizzare la materia, si riferisce anche a periodi lontani dalle elezioni²⁴.

La disciplina derivante dalle due leggi a oggi in vigore affida il ruolo di vigi-

1964 con cui i Giudici giungono ad affermare la medesima derivazione costituzionale della libera manifestazione del pensiero e dell'utilizzo dei mezzi di comunicazione, conseguentemente legittimando, in tal modo, l'intervento legislativo volto a regolare il settore della comunicazione al fine di garantire il pluralismo informativo.

22. G. Sirianni, *Par condicio. I complessi rapporti tra potere politico e potere televisivo*, cit.

23. Tale norma costituisce uno sviluppo della precedente disciplina contenuta nella legge n. 103 del 1975 che per prima ha incluso il pluralismo politico tra i principi informatori del servizio pubblico ex art. 1, co. 2, affidandone la garanzia alla Commissione parlamentare per l'indirizzo e la vigilanza sui servizi radiotelevisivi, la quale (ex art. 4) aveva il potere di disciplinare direttamente le rubriche di "tribuna politica", "tribuna sindacale", "tribuna elettorale", "tribuna stampa" che la Rai era obbligata a trasmettere; tuttavia, la necessità di un aggiornamento non più rinviabile che ha portato l'emanazione della legge del 1993 si ravvisa nella nascita delle emittenze private che sino a quel momento risultavano essere prive di una regolamentazione.

24. Le disposizioni di cui la legge n. 28 del 2000 sono state oggetto di particolari critiche in dottrina fondate sull'assunto secondo cui tale norma «ponendo una disciplina di estremo dettaglio lederebbe un principio di pari valore quale quello della libertà di manifestazione del pensiero, mirando a salvaguardare un interesse di dubbio radicamento costituzionale, ovvero l'interesse che l'elettore esprima in modo corretto la propria volontà», così F. Furlan, *Articolo 48*, in S. Bartole, R. Bin (a cura di), *Commentario breve alla Costituzione*, Cedam, 2008, 495 ss.; cfr. anche sul punto G. Bognetti, *La par condicio alla luce dell'esperienza americana*, in *Dir. Pubbl. Com. eur.*, 2000, p. 485 ss.; P. Caretti, *I diritti fondamentali libertà e diritti sociali*, Giappichelli, 2005; per una completa analisi della normativa si v. anche F. Biondi Del Monte, *Esiste ancora la "par condicio"? Contenuto e limiti della legge n. 28/2000*, in *Osservatorio Costituzionale*, n. 1, 2018.

lanza sul rispetto della *par condicio* e i relativi limiti e divieti da queste previsti, a due diverse istituzioni a seconda dei canali comunicativi utilizzati. Se per il servizio pubblico l'autorità competente è la Commissione parlamentare per l'indirizzo generale e la vigilanza dei servizi radiotelevisivi, per i canali privati il medesimo ruolo viene affidato all'Autorità Garante delle comunicazioni (Agcom). Provvedendo oneri comunicativi tra le due autorità in modo da assicurare l'effettività del controllo che, di fatto, risulta lacunoso.

Un sistema «bicefalo» e «schizofrenico»²⁵ che a oggi pare mostrare la sua inadeguatezza al recepimento di istanze di differente portata derivanti non solo dall'utilizzo di canali che, per le loro intrinseche caratteristiche²⁶, rischiano di porre ancor di più a rischio i precetti costituzionali in cui si articola il «diritto all'informazione»²⁷; ma soprattutto dalla necessità di adattare il piano normativo a un piano fattuale caratterizzato dalla moltiplicazione dei canali comunicativi²⁸.

Nonostante la legge del 2000 abbia tentato di portare la disciplina relativa alla propaganda politica nei media a un livello successivo, l'articolo 2 della norma n. 28/2000, riferendosi esplicitamente alla diffusione su «mezzi radiotelevisivi di programmi contenenti opinioni e valutazioni politiche», rende impossibile la sua applicazione in senso estensivo all'ambiente del *web*.

Sia la normativa che le diverse sentenze della Corte²⁹ che si inseriscono in questa tematica si pongono l'obiettivo di tutelare il pluralismo informativo, tanto nella dimensione interna – assicurando a tutte le forze politiche pari condizioni di visibilità – quanto nella dimensione esterna, la quale si concretizza attraverso la garanzia della diversità delle fonti informative a cui poter attingere.

Sul punto pare utile precisare che l'incontrollabilità dell'eccesso delle fonti

25. Così G. Sirianni, *Par condicio. I complessi rapporti tra potere politico e potere televisivo*, cit.; per una trattazione completa della legislazione in materia di *par condicio* cfr. C. Pinelli, *Può la pari opportunità di voci convertirsi nell'imposizione del silenzio?*, in *Giur. cost.*, n. 1/1995; E. Lamarque, *Propaganda e pubblicità elettorale nella sentenza n. 161/95 della Corte costituzionale*, in *Dir. Inf.*

26. Si fa qui riferimento alla capacità di tali sistemi comunicativi di dissuadere l'utente attraverso la riproposizione a quest'ultimo di contenuti a lui compiacenti, basati su informazioni dallo stesso prodotte, ricevute e poi riutilizzate dalla piattaforma consultata; cfr. E. Longo, *Dai "big data" alle "bolle filtro": nuovi rischi per i sistemi democratici*, in *Percorsi costituzionali*, n. 1, 2019.

27. Cfr. B. Caravita, *Social network, formazione del consenso, istituzioni politiche: quale regolamentazione possibile?*, in *federalismi.it*, n. 2/2019.

28. O. Pollicino, *Tutela del pluralismo nell'era digitale: ruolo e responsabilità degli Internet Service Provider*, cit.

29. Cfr. sent. n. 105 del 1972 Corte cost.

informative attingibili costituisce, a oggi, la ragione fondativa delle richieste di coloro che reclamano l'intervento legislativo. Un pluralismo di fonti che, proprio la legislazione vigente si era posta l'obiettivo di tutelare e garantire, rischia a oggi di essere distorto nella massimizzazione del contesto³⁰.

Posta la poca attualità della disciplina normativa dell'ordinamento italiano dinanzi all'innovazione tecnologica e tutti i rischi a essa connessi, un ruolo di supplenza è stato assunto dall'Agcom le cui competenze in ambito di comunicazione politica risalgono alla legge istruttiva della stessa autorità n. 249 del 1997.

Nonostante l'articolo 1 della suddetta legge istitutiva ricollega il ruolo dell'Autorità a mezzi di comunicazione specifici, l'Agcom, nel corso degli ultimi anni, si è auto-attribuita «un ruolo di impulso e coordinamento tra i diversi attori operanti nel settore dell'informazione online per favorire l'autoregolamentazione su base volontaria ai fini di contrasto dei fenomeni di disinformazione generale e, nello specifico, nella lesione della correttezza, dell'imparzialità e del pluralismo dell'informazione»³¹ attraverso l'emanazione della delibera n. 423 del 2017 con cui viene istituito il Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali a cui si deve l'instaurazione di un dialogo tra esperti e attori economici interessati dal quale è risultata l'elaborazione di soluzioni condivise per contrastare la disinformazione e sostenere iniziative di autoregolamentazione della comunicazione politica attraverso internet.

Il concreto risultato di questo Tavolo tecnico è l'elaborazione delle Linee guida per la parità di accesso alle piattaforme online durante la campagna elet-

30. Secondo Donati la pressoché illimitata «moltiplicazione delle piattaforme trasmissive di informazioni» non può di fatto rendere esaurita la necessità di interventi statali volti a tutelare il pluralismo informativo poiché quest'ultimo viene a essere messo a rischio proprio dal funzionamento delle piattaforme digitali, fondato su «situazioni di concentrazione e dominanza nella diffusione di contenuti informativi tali da consentire a uno o pochi soggetti di influenzare e alterare il corretto svolgimento del dibattito politico», così F. Donati, *Democrazia, pluralismo delle fonti di informazione e rivoluzione digitale*, in *federalismi.it*, n. 23/2013.

31. Delibera 423/2017/Cons "Istituzione di un Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali" a cui viene affidato il compito di favorire «lo scambio e il consolidamento di buone pratiche tra i soggetti operanti nel settore dell'informazione anche online, con riferimento: alle metodologie di rilevazione, alle strategie di prevenzione, strumenti di blocco o rimozione dei contenuti online che siano qualificati come lesivi della correttezza e imparzialità e del pluralismo dell'informazione; monitoraggio dei flussi economici volti al finanziamento di tali contenuti lesivi; definizione di tecniche di contrasto delle strategie di disinformazione online con particolare riferimento a quelle di account falsi o inesistenti»; disponibile su: www.agcom.it/sites/default/files/migration/delibera/Delibera_423-17-CONS.pdf.

torale per le elezioni politiche 2018³², le quali esplicitamente richiamano, quale loro base fondativa, i principi espressi dalla legge n. 28/2000 ossia imparzialità ed equità nelle condizioni di accesso ai mezzi di comunicazione per tutti i soggetti politici.

Alla disciplina del 2017 fa seguito quella introdotta con la delibera del 2019 n. 94³³ emanata in vista delle elezioni europee dello stesso anno, che richiama nella quasi totalità le indicazioni fornite dal Tavolo tecnico adattandole a necessità concrete del periodo elettorale. Tra queste, sicuramente meritevoli di menzione sono: la richiesta di indicare, nell'ambito di messaggi pubblicitari elettorali online, la natura di "messaggio elettorale" e specificarne il soggetto politico committente; il divieto di pubblicare sondaggi nei quindici giorni precedenti al voto; il richiamo al silenzio elettorale nel giorno della votazione e in quello precedente anche nell'ambito delle piattaforme social.

La strumentalizzazione dei dati degli utenti-elettori in ambito delle tecniche riconducibili al c.d. marketing politico³⁴ ha posto le basi per l'intervento dell'Autorità preposta alla tutela dei dati personali e della privacy. Un intervento che ha assunto le vesti di una Dichiarazione del 2019, adottata dal Comitato europeo per la protezione dei dati, in cui si richiede alle Autorità nazionali di «monitorare e, se necessario, fa rispettare l'applicazione dei principi di protezione dati, quali la trasparenza, la limitazione delle finalità, la proporzionalità e la sicurezza, nonché l'esercizio dei diritti dell'interessato, nel contesto delle elezioni e delle campagne politiche»³⁵.

32. Nella premessa delle Linee Guida per la parità di accesso alle piattaforme online durante la campagna elettorale per le elezioni del 2018 troviamo indicato che «Sulla base delle iniziative discusse nell'ambito del Tavolo tecnico, e tenuto conto dei compiti istituzionali che la legge affida all'Autorità in materia di par condicio elettorale, sono stati identificati i principi generali che informano la materia e che, in quanto tali, sono applicabili a tutti i mezzi di informazione, comprese le piattaforme digitali»; disponibile su: www.agcom.it/sites/default/files/migration/documento_generico/Documento_generico_01-02-2018_0.pdf.

33. Gli impegni assunti dalle società esercenti le piattaforme online per garantire la parità di accesso dei soggetti politici alle piattaforme digitali durante la campagna elettorale per le elezioni dei membri del Parlamento europeo spettanti all'Italia – 2019 costituiscono il secondo intervento di autoregolamentazione per le campagne elettorali promosso dal "Tavolo tecnico" per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali", documento disponibile sul sito dell'Autorità.

34. G. Buttarelli, *Marketing politico e tutela dei dati personali*, "26Th International Conference on Privacy and Personal Data Protection", Wrocław, 14-16 settembre 2004.

35. Dichiarazione 2/2019 "Sull'uso dei dati personali nel corso di campagne politiche" adottata il 13 marzo 2019 dal Comitato europeo per la protezione dei dati personali in cui si legge che «il rispetto delle norme in materia di protezione dei dati, anche nel contesto delle attività elettorali e delle campagne politiche, è essenziale per proteggere la democrazia e rappresenta uno strumento utile a mantenere la fiducia dei cittadini e l'integrità delle elezioni».

Emblematico, dal punto di vista del ragionamento di cui si discute, l'emissione del provvedimento Meta del 2022³⁶ da parte del Garante per la Protezione dei Dati Personali, volto a sanzionare il gruppo Meta, in particolare Facebook Italia, per le modalità poco trasparenti attraverso cui la piattaforma coinvolgeva gli utenti maggiorenni al sito del Ministro degli Interni, al fine di assicurare una maggiore conoscenza e conoscibilità delle informazioni relative alle elezioni politiche che di lì a poco si sarebbero tenute.

La scarsa informazione relativa al trattamento e alla seguente profilazione dei dati degli utenti-elettori, nonché l'assenza di qualsivoglia incarico formale tra Ministro degli Interni e Gruppo Meta, hanno comportato l'intervento del Garante al fine di riequilibrare le modalità attraverso le quali gli elettori venivano informati³⁷. Un intervento non isolato nel panorama europeo, a cui infatti ha fatto seguito il maxi provvedimento sanzionatorio del Garante irlandese.

Nonostante lo sforzo e le attività delle Autorità Indipendenti nel settore siano utili e quanto meno adatte a recepire le istanze relative alle problematiche più attuali³⁸, queste ultime peccano di efficacia nella misura in cui si tratta di approcci invitanti e privi del carattere precettivo, generale e astratto caratterizzanti il solo intervento legislativo.

Da più voci è stato a lungo reclamato un intervento comunitario nel settore della comunicazione politica, posto l'utilizzo di uno strumento quale l'*online* che per definizione travalica i singoli confini nazionali sia come effetti che come conseguenze.

Dopo un primo approccio caratterizzato dall'emissione di direttive comunitarie e atti di *soft law* la cui incidenza effettiva si è rivelata abbastanza fumosa, l'Unione Europea tenta di rispondere alla problematica attraverso l'adozione del recente Regolamento 900/2024³⁹ attraverso il quale si introducono, sulla base della necessità di una maggior trasparenza, doveri di etichettatura dei messaggi di natura politica *online* e del preventivo consenso dell'utente che diviene necessario alla profilazione del medesimo.

36. Si fa riferimento al provvedimento adottato in data 21 dicembre 2022, disponibile su: www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9853406.

37. Cfr. Garante per la Protezione dei Dati Personali, *Provvedimento del 20 giugno 2024*, disponibile su: www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10112606.

38. Così B. Caravita, *Social network, formazione del consenso, istituzioni politiche: quale regolamentazione possibile?*, cit.

39. V. Iaia, *I complessi contorni della nozione di pubblicità politica alla prova delle elezioni europee 2024*, in *MediaLaws*.

4. Riflessioni conclusive: una risposta condivisa a un problema comune

Il significato costituzionale della comunicazione politica, ricavabile dall'articolo 21, comporta alcune criticità legate alla rigidità delle garanzie previste per la tutela della libertà di manifestazione del pensiero. Le garanzie riservate dal Testo alla libertà di cui si discute, infatti, possono porsi in contrasto con esigenze di regolazione di un settore in cui le manifestazioni del pensiero si trasformano in informazione politica, presupposto necessario alla formazione dell'autodeterminazione del singolo elettore attorno a tematiche politiche di rilievo.

La rilevanza costituzionale delle implicazioni derivanti dall'utilizzo s-regolato del settore si fa ancora più forte a seguito dell'intervento delle nuove tecnologie quali strumenti di propaganda politica. A fronte di tale situazione, il quadro legislativo nazionale presenta una disciplina inidonea a rispondere a criticità portate da meccanismi più moderni, relegando la disciplina della comunicazione politica alla regolamentazione della *par condicio* delle diverse forze politiche nelle trasmissioni radio-televisive.

Dinanzi a un quadro regolatorio datato, il panorama normativo si amplia e si innova con il recente Regolamento europeo 900/2024, specificamente volto a disciplinare il settore della pubblicità politica, adattando la disciplina previamente sperimentata in ambito di *marketing*. Un intervento del legislatore sovranazionale necessitato dalla dimensione del fenomeno che si è tentato di regolamentare, nonché – come si è tentato di evidenziare nel paragrafo precedente – dal silenzio del legislatore interno sul tema.

A fronte delle brevi riflessioni sin qui svolte pare opportuno dunque interrogarsi sulla sede prescelta a definire i contorni di tale situazione di incertezza normativa: da un lato, la scelta di un intervento del legislatore comunitario pare essere suffragata dal rilievo sovranazionale del fenomeno di cui si discute, dall'altro, invero, la necessità di una legislazione nazionale si ritiene possa essere confermata sulla base del rilievo costituzionale che possiedono le implicazioni della materia della comunicazione politica; da ultimo, il ruolo delle Autorità amministrative indipendenti, a oggi sintomo patologico di un sistema normativo inidoneo a rispondere a istanze più recenti, ma il cui intervento nel settore può essere giustificato dal requisito dell'indipendenza strutturale e funzionale che potrebbe garantire un ruolo di revisore e controllore della correttezza dell'attività *online* svolta dagli esponenti politici.

IL SISTEMA DI *GOVERNANCE* ED *ENFORCEMENT* DEL REGOLAMENTO (UE) 2024/900 ALLA PROVA DELL'EFFETTIVITÀ: UN'ANALISI COMPARATIVA CON IL *DIGITAL SERVICES ACT*

Emanuela Palomba*

SOMMARIO: 1. La disciplina della “pubblicità politica 4.0” nel Regolamento (UE) 2024/900 – 2. L'apparato di *governance* e i meccanismi di *enforcement*: Regolamento (UE) 2024/900 e *Digital Services Act* a confronto – 3. Conclusioni.

1. La disciplina della “pubblicità politica 4.0” nel Regolamento (UE) 2024/900

Alla luce del moltiplicarsi degli spazi e dei mezzi di informazione nello sconfinato «mare *schmittiano* di *Internet*»¹, l'Unione europea ha approvato nel marzo 2024 il Regolamento (UE) 2024/900 relativo alla trasparenza e al *targeting* della pubblicità politica (Ttpa)², segnando così una nuova tappa di quel lungo cammino che la vede impegnata nella ricerca del punto di equilibrio della triangolazione «tecnologia – democrazia – diritti»³. Se, da un lato, infat-

* Dottoranda di ricerca in IA – Diritto costituzionale e pubblico, Università degli Studi di Napoli Federico II.

1. In questi termini, E. Cocchiara, *La regolazione del digitale: tra self-regulation, codici di condotta e authorities*, in *Nuove Autonomie*, n. 1, 2025, p. 386, «Internet è un nuovo mare *schmittiano* e perciò con un proprio *nomos* diverso da quello della terra». Cfr. M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista del Gruppo di Pisa*, n. 2, 2021, pp. 166-167; A. Papa, *Disintermediazione dell'informazione e della comunicazione politica nelle democrazie digitali*, in *federalismi.it*, n. 15, 2025, pp. 25-26.

2. Regolamento (UE) 2024/900 del Parlamento europeo e del Consiglio del 13 marzo 2024, relativo alla trasparenza e al *targeting* della pubblicità politica. Tale Regolamento è entrato in vigore nell'aprile 2024 e si applica a decorrere dal 10 ottobre 2025.

3. Così A. Festa, *La disinformazione online come “minaccia ibrida” alla democrazia nell'Unione europea: meccanismi di tutela e strumenti a contrasto per uno Spazio di libertà, sicurezza e giustizia*, in *Freedom, Security & Justice: European Legal Studies*, n. 2, 2025, p. 283. Per una disamina approfondita del lungo e tortuoso iter normativo della regolamentazione europea sulle

ti, l'utilizzo sempre più frequente delle piattaforme digitali e dei sistemi basati sull'Intelligenza artificiale (IA)⁴ consente agli attori politici di aumentare la propria visibilità e di affinare le proprie tecniche comunicative nella "pubblicità politica 4.0"⁵, dall'altro lato l'impiego dei nuovi strumenti tecnologici può compromettere l'affidabilità delle informazioni diffuse e, dunque, l'integrità del processo elettorale⁶.

Quale *lex specialis* del Regolamento (UE) 2022/2065 relativo a un mercato unico dei servizi digitali (Dsa)⁷, il Ttpa stabilisce norme armonizzate nel territorio dell'UE per la fornitura di pubblicità politica e dei servizi connessi e sull'uso delle tecniche di *targeting* o di consegna del messaggio pubblicitario in

piattaforme digitali, si rinvia a M.C. Girardi, *Libertà e limiti della comunicazione nello spazio pubblico digitale*, in *federalismi.it*, n. 17, 2024.

4. Sull'evoluzione delle campagne politiche alla luce dei nuovi mezzi tecnologici, J. Chester, K.C. Montgomery, *The role of digital marketing in political campaigns*, in *Internet Policy Review*, vol. 6, n. 4, 2017; V. Iaia, *I complessi contorni della nozione di pubblicità politica alla prova delle elezioni europee 2024*, in *MediaLaws*, 19 giugno 2024: «gli ambienti digitali diventano sempre più il "place to be" per ottenere il sostegno del pubblico». In tal senso, I. Khan, *Freedom of expression and elections in the digital age: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, United Nations, Human Rights Council, A/HRC/59/50, 11th June 2025, p. 4: «Politicians increasingly use direct communication and social media to reach a wider audience and convey the image of being "one of the people"».

5. La qualificazione qui proposta della pubblicità politica contemporanea come "pubblicità politica 4.0" deriva dall'inquadramento dell'attuale periodo storico nella c.d. "Quarta Rivoluzione scientifica". Cfr. F. Laviola, *Regolazione della tecnologia e dimensione del tempo*, in *Osservatorio sulle fonti*, n. 3, 2021, p. 1168, che delinea le quattro fasi storiche della Rivoluzione scientifica: Prima Rivoluzione scientifica-copernicana, Seconda Rivoluzione scientifica-darwiniana, Terza Rivoluzione scientifica-freudiana, Quarta Rivoluzione scientifica-turingiana; K. Schwab, P. Pyka, *Die Vierte Industrielle Revolution*, Pantheon, 2016; L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, 2017.

6. Le elezioni presidenziali rumene del 2024 rappresentano un esempio recente di come l'impiego delle tecnologie digitali e dell'IA possa incidere sullo svolgimento del processo elettorale. Nel dicembre 2024, la Corte costituzionale della Romania ha annullato il primo turno elettorale a causa dell'intervenuta manipolazione del voto e della distorsione del principio di pari opportunità tra i candidati, derivante dall'utilizzo illecito e non trasparente delle tecnologie digitali e dell'IA da parte del candidato indipendente Călin Georgescu durante la propria campagna elettorale. Così Curtea Constituțională a României, sent. n. 32/2024 del 6 dicembre 2024. Cfr. A. Simoncini, *L'annullamento delle elezioni presidenziali in Romania. Luci e ombre di una divisiva decisione costituzionale*, in *Quaderni costituzionali*, vol. 45, n. 1, 2025; S. Sassi, A. Sterpa, *La Corte costituzionale della Romania difende la democrazia liberale dalla disinformazione. Prime note sulla sentenza n. 32 del 6 dicembre 2024*, in *federalismi.it*, n. 4, 2025; F. Rosa, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025.

7. 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).

ambito di pubblicità politica *online* qualora comportino il trattamento di dati personali (*ex art. 1, par. 1*), al fine di assicurare il corretto funzionamento del relativo mercato interno, nel rispetto e a tutela dei diritti e delle libertà fondamentali garantiti dalla Carta dei diritti fondamentali dell'Unione europea (*ex art. 1, par. 4*).

Tale Atto mira, dunque, a bilanciare la libertà di manifestazione del pensiero dell'attore politico con il parallelo diritto dei cittadini a una completa e obiettiva informazione politica, premessa essenziale al corretto esercizio del diritto di voto e, dunque, presupposto ineludibile della democraticità dell'ordinamento⁸.

Per assicurare l'applicazione degli obblighi scaturenti dal Ttpa⁹, rivestono fondamentale importanza i meccanismi di vigilanza e di esecuzione predisposti dal Legislatore europeo, dacché solo un solido apparato di controllo e repressione delle infrazioni è idoneo a garantire l'effettiva tutela dei fondamentali diritti politici dei destinatari dei *political advertisement*.

Questo contributo si propone di analizzare, pertanto, l'apparato di *governance* e i meccanismi di *enforcement* previsti dal Ttpa, interrogandosi sulla loro adeguatezza ed efficacia alla luce del coordinamento e della comparazione con la parallela, e talvolta sovrapposta, normativa del Dsa¹⁰.

8. La Corte costituzionale italiana, secondo un orientamento ormai consolidato, interpreta l'art. 21 Cost. come fondamento costituzionale non solo del "diritto di informare", ma anche del "diritto a essere informati" (Corte cost., nn. 202/1976; 148/1981; 826/1988). Sul diritto dei cittadini a essere informati come presupposto per il corretto esercizio del diritto di voto *ex art. 48 Cost* e, dunque, come «pietra angolare dell'ordine democratico», *precipue* Corte cost. sent. n. 84/1969, ma anche sent. nn. 126/1985 e 112/1993. Cfr. *ex pluribus*, A. Loiodice, *Situazioni costituzionali e diritto all'informazione*, Cedam, 1990, pp. 35 ss.; L. Cianci, *Il diritto a essere informati alla prova delle strategie di microtargeting per la comunicazione politica*, in *Nomos – Le attualità nel diritto*, n. 1, 2023, p. 3; P. Villaschi, *The regulation of political targeting in the Italian and European Union legal framework*, in *Rivista di diritto internazionale privato e processuale*, n. 1, 2024, p. 140; F. Fabrizzi, *La formazione del consenso, la rappresentanza, il Parlamento nell'era della digitalizzazione*, in *Italian Papers On Federalism*, n. 1, 2025, pp. 124-142.

9. Per l'analisi della disciplina sostanziale del Regolamento (UE) 2024/900 si permetta il rinvio a un proprio scritto, *Le neurotecnologie nella comunicazione politica: luci e ombre del Regolamento (UE) 2024/900*, in *Lecostituzionaliste.it*, dicembre 2024.

10. Il quadro normativo sui sistemi di moderazione e raccomandazione dei contenuti *online* è ulteriormente complicato dalla intersezione di numerosi atti dell'UE, fra cui il Regolamento (UE) 2016/679 sulla protezione dei dati personali, il Regolamento (UE) 2024/829 sull'intelligenza artificiale, nonché il recente Regolamento Europeo (UE) 2024/1083 sulla libertà dei media. Sull'interazione fra i diversi interventi legislativi unionali nel settore digitale, A. Licastro, *La regolazione dello spazio informativo disintermediato dinanzi alla progressiva avanzata della "synthetic media technology" in tempo di tornate elettorali*, in *Diritto pubblico*, n. 3, 2024; L. Califano, *La strategia normativa dell'Unione europea per un nuovo ordine digitale*, in *federalismi.it*, n. 15, 2025; E.A. Rossi, *Problemi di coordinamento normativo nel contrasto alla disinforma-*

2. L'apparato di governance e i meccanismi di enforcement: Regolamento (UE) 2024/900 e Digital Services Act a confronto

Il sistema di vigilanza elaborato dal Legislatore europeo nel Ttpa presenta una struttura “a due fattori” analoga a quella prevista nel Dsa.

Spetta, innanzitutto, ai singoli utenti (persone fisiche o giuridiche) segnalare agli editori di pubblicità politica i *political advertisement* potenzialmente non conformi al Ttpa¹¹, che essi dovranno esaminare in modo diligente, non arbitrario e obiettivo, dandovi seguito tempestivamente (*ex art. 15 Ttpa*)¹². La funzione di vigilanza viene, dunque, attribuita a soggetti privati, sia con riferimento alla segnalazione delle condotte illecite, sia per quanto concerne la successiva attività di valutazione e sanzionamento.

Parallelamente anche le autorità nazionali sono chiamate a trattare, ai sensi dell'art. 24, ogni notifica loro presentata riguardante violazioni del Ttpa e a monitorarne la corretta applicazione, secondo il rigoroso schema “tripartito”¹³ di delimitazione delle competenze delineato all'art. 22: per il controllo sulla corretta applicazione degli artt. 18 e 19 in materia di *targeting* e di consegna del messaggio di pubblicità politica *online* sono competenti le autorità nazionali per la protezione dei dati personali e il Garante europeo della protezione dei dati (*ex art. 22, par. 1*); per vigilare sull'osservanza degli obblighi da parte dei prestatori di servizi intermediari, è demandato agli Stati membri il compito di designare le autorità competenti, che possono anche coincidere con le autorità designate a norma del Dsa (*ex art. 22, par. 3*); infine, per l'esecuzione e l'applicazione di tutti gli aspetti residuali, ogni Stato membro deve designare una o più autorità

zione politica e commerciale, in *federalismi.it*, n. 15, 2025. Proprio al fine di semplificare e armonizzare il complesso ecosistema normativo sul digitale in ossequio a quanto evidenziato dall'ex Presidente della Banca centrale europea, Mario Draghi, nel suo *Report on EU competitiveness* del settembre 2024, la Commissione europea ha presentato il 19 novembre 2025 due nuove proposte di regolamento (COM(2025) 836 e COM(2025) 837) rientranti nel pacchetto *Digital Omnibus*. Tali Atti mirano a ridurre quegli oneri amministrativi e quelle incoerenze normative individuate nel Rapporto Draghi come due dei principali *competitiveness deficit* che ostacolano le *startup* europee in crescita. Cfr. M. Draghi, *The future of European competitiveness – September 2024*, Luxembourg: Publications Office of the European Union, 2025, in part., *Part B – In-depth analysis and recommendations*, pp. 317-326.

11. E. Caterina, *Verso il nuovo regolamento UE in materia di pubblicità politica: mercato delle idee o della propaganda?*, in *Quaderni costituzionali*, n. 1, 2024, p. 209.

12. Si tratta del noto meccanismo di *notice and action*, ripreso dall'art. 16 Dsa. Cfr., *ex pluribus*, E. Birritteri, *Contrasto alla disinformazione*, Digital Services Act e attività di private enforcement: *fondamento, contenuti e limiti degli obblighi di compliance e dei poteri di autonormazione degli operatori*, in *MediaLaws*, n. 2, 2023, pp. 60 ss.

13. C. Acocella, *Il sistema di enforcement della disciplina in materia di pubblicità politica: le autorità competenti e le sanzioni*, in *AmbienteDiritto.it*, n. 4, 2025, pp. 10-13.

competenti, anche non coincidenti con quelle di cui ai parr. 1 e 3, che devono in ogni caso godere di piena indipendenza strutturale «sia dal settore che da qualsiasi intervento esterno o pressione politica» (*ex art. 22, par. 4*).

Pertanto, al controllo “privato” sulla corretta applicazione degli obblighi scaturenti dal Ttpa, si affianca il controllo “pubblico” delle autorità nazionali competenti, che a tal fine godono dei poteri di cui all’art. 22, par. 5.

In entrambe le fasi di vigilanza si deve constatare l’attuale assenza di un chiaro e definito *framework* normativo che ne delimiti gli attributi e i requisiti.

Innanzitutto, non risultano ancora nitidi i criteri da utilizzare per identificare e valutare i messaggi di pubblicità politica “occulta”, vale a dire non espressamente dichiarata dai prestatori di servizi di pubblicità politica in violazione della normativa del Ttpa¹⁴.

A fronte di una ampia formulazione di *political advertising* in sede di apparato definitorio (*ex art. 3, punto 2, Ttpa*)¹⁵, l’art. 8 reca un elenco non tassativo di indici utili all’individuazione dei messaggi di pubblicità politica – il contenuto, lo sponsor, la lingua, il contesto, i mezzi, i destinatari, l’obiettivo – la cui concretizzazione è demandata alla Commissione in sede di adozione di orientamenti comuni.

I primi orientamenti pubblicati dalla Commissione europea il 10 ottobre 2025¹⁶ non sembrano tuttavia dirimere i preesistenti nodi interpretativi relativi all’apparato definitorio e ai criteri di individuazione dei messaggi di pubblicità politica. Quanto all’apparato definitorio, la Commissione non fornisce infatti soluzioni interpretative di carattere definitivo, limitandosi per lo più a riprodurre in via tautologica il contenuto normativo dell’articolato o ad aggiungere indicazioni che, per la loro portata, non possono considerarsi pienamente esaustive¹⁷. Quanto ai criteri di identificazione dei messaggi di pubblicità politica, l’intervento della Commissione si risolve invece nella mera formulazione

14. A. Festa, *La disinformazione online come “minaccia ibrida” alla democrazia nell’Unione europea: meccanismi di tutela e strumenti a contrasto per uno Spazio di libertà, sicurezza e giustizia*, cit., p. 290.

15. Cfr. C. Giunta, *L’espressione del pensiero politico: comunicazione, propaganda, pubblicità*, in *federalismi.it*, n. 14, 2025, pp. 22 ss.; V. Iaia, *I complessi contorni della nozione di pubblicità politica alla prova delle elezioni europee 2024*, cit.

16. Commissione europea, Comunicazione della Commissione, *Orientamenti per sostenere l’attuazione del regolamento (UE) 2024/900 relativo alla trasparenza e al targeting della pubblicità politica*, 10 ottobre 2025, C/2025/5514.

17. Si consideri, a titolo esemplificativo, la distinzione tra i messaggi di pubblicità politica, ricompresi nell’ambito di applicazione del Ttpa, e i messaggi di natura meramente privata o commerciale, cui si applica la deroga: la Commissione si limita a fornire alcuni indici sintomatici della finalità politica, personale o commerciale – quali il contenuto, il linguaggio, il tono e il contesto generale – senza però elaborare parametri uniformi o criteri di portata generale.

di alcune domande orientative che gli operatori dovrebbero porsi nel valutare i potenziali *political advertisement*¹⁸. La perdurante vaghezza dell'apparato definitorio, non superata neppure dai più recenti orientamenti della Commissione, rischia pertanto di tradursi in applicazioni eterogenee e, conseguentemente, in potenziali disarmonie tra gli ordinamenti degli Stati membri.

Con specifico riguardo, al controllo pubblico, se da un lato appare sicuramente degna di nota l'attribuzione di ampi poteri alle autorità nazionali e il vincolo di indipendenza "strutturale" delle stesse¹⁹, dall'altro va rilevato un significativo irrigidimento nel riparto delle competenze che rischia di compromettere l'armonia del sistema, generando ambiguità applicative e potenziali conflitti di competenza.

Consapevole del suddetto rischio, il Legislatore europeo ha rimarcato più volte l'esigenza di collaborazione fra le diverse autorità nazionali e fra queste e le istituzioni europee in applicazione del principio di leale collaborazione *ex art. 4, par. 3, TUE*²⁰.

Tuttavia, come si avrà modo di evidenziare, i meccanismi di collaborazione concretamente predisposti non appaiono ancora in grado di risolvere in via definitiva i rischi di un *enforcement* inefficace e di una tutela ineffettiva.

L'art. 22, par. 9, Ttpa impone a ciascuno Stato membro di designare un'autorità competente come "punto di contatto nazionale" a livello dell'UE al fine di assicurare una cooperazione e un coordinamento efficaci e strutturati a livello nazionale tra tutte le pertinenti autorità competenti e agevolare la cooperazione tra queste e i punti di contatto nazionali degli altri Stati membri.

18. Commissione europea, Comunicazione della Commissione, *Orientamenti per sostenere l'attuazione del regolamento (UE) 2024/900 relativo alla trasparenza e al targeting della pubblicità politica*, cit., *Tabella 1 – Elementi da prendere in considerazione ai fini dell'identificazione di un messaggio di pubblicità politica*, pp. 19-21.

19. Tale scelta regolatoria appare particolarmente significativa se comparata alla parallela disciplina dell'*AI Act* di poco successivo che, all'art. 70, par. 1, si limita a prescrivere la mera indipendenza "funzionale" delle autorità nazionali competenti, lasciando agli Stati membri ampia discrezionalità nella loro designazione. Per un'analisi critica della *governance* dell'*AI Act*, C. Novelli, P. Hacker, J. Morley, J. Trondal, L. Floridi, *A robust governance for the AI act: AI office, AI Board, scientific panel, and national authorities*, in *European Journal of Risk Regulation*, pp. 1-25, 2024. In Italia, S. Calzolaio, *Autorità indipendenti e di governo della società digitale*, in F. Pizzetti (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, pp. 100-104.

20. Sull'esigenza di una maggiore collaborazione e di un più proficuo coordinamento tra le diverse autorità nazionali competenti nell'ecosistema digitale, G. De Minico, *The Challenge of the Virtual World for Independent Authorities*, in *European Public Law*, vol. 29, n. 1, 2023, in part. p. 2 e pp. 23-24. Di recente anche M. Guerra, *Le prospettive di collaborazione tra Autorità indipendenti nell'era digitale*, in *Rivista italiana di informatica e diritto*, vol. 7, n. 1, 2025, pp. 1-30.

A livello unionale, i punti di contatto nazionali si riuniscono periodicamente nella “rete dei punti di contatto nazionali” per il reciproco scambio periodico di informazioni e prassi eccellenti, nonché per una cooperazione strutturata con la Commissione, facilitando l’elaborazione di orientamenti volti ad aiutare gli sponsor e i prestatori di servizi di pubblicità politica a conformarsi ai requisiti del Ttpa (*ex art. 22, par. 8*).

Se, dunque, a livello teorico, sembrerebbe assicurata la cooperazione tra le autorità nazionali competenti (attraverso i punti di contatto nazionali), e fra queste e i punti di contatto nazionali degli altri Stati membri (attraverso la rete dei punti di contatto nazionali)²¹, a livello pratico, il controllo sull’esecuzione delle norme del Ttpa e sull’applicazione delle relative misure sanzionatorie risulta ancora inefficace. Sebbene, infatti, la competenza spetti alle autorità nazionali dello Stato membro in cui il prestatore di servizi è stabilito (*ex art. 23, par. 1, primo periodo*), nei casi in cui il prestatore risulti formalmente stabilito in più Stati membri essa sarà attribuita all’autorità dello Stato membro dove si trova lo stabilimento principale (*ex art. 23, par. 1, secondo periodo*), che però sarà tenuta a collaborare con le autorità competenti degli altri Stati membri coinvolti (*ex art. 23, parr. 2 e 3*). Pertanto, ai sensi dell’art. 23, par. 4, se un’autorità nazionale competente di uno Stato membro dovesse sospettare che il Ttpa sia stato violato sul proprio territorio, dovrà informare l’autorità competente sede dello stabilimento principale del prestatore, chiedendole, laddove necessario, di valutare la questione e di adottare le opportune misure di indagine e di esecuzione. Spetterà, poi, all’autorità competente dello stabilimento principale comunicare, senza indebito ritardo e non oltre un mese dal ricevimento della segnalazione, all’autorità competente che ha effettuato la notifica e alla rete dei punti di contatto nazionali la propria valutazione sulla presunta violazione e le misure di indagine o di esecuzione adottate/che intende adottare (*ex art. 23, par. 7*).

Dinanzi, quindi, a una presunta violazione del Ttpa e alla possibile lesione dei fondamentali diritti politici dei propri cittadini, l’autorità nazionale dello Stato membro in cui il prestatore di servizi avrebbe commesso la violazione, ma nel quale non detiene il proprio stabilimento principale, dovrà attendere la valutazione e le eventuali misure di indagine o di esecuzione adottate dall’autorità nazionale di un altro Stato membro (segnatamente lo Stato dove il prestatore detiene il proprio stabilimento principale).

Benché siano consentite indagini congiunte (ai sensi dell’art. 23,

21. L. Ullrich, *The European Regulation on the Transparency and Targeting of Political Advertising*, in B. Steinrötter, C. Heinze, M. Denga (eds.), *EU Platform Law. A Handbook*, Nomos, vol. I, 2025, pp. 657 ss.

parr. 8-10), tale possibilità risulta notevolmente attenuata dalla previsione normativa secondo cui spetta pur sempre all'autorità nazionale dello Stato membro sede di stabilimento principale la scelta in ordine al loro avvio e alla loro conduzione²².

Il rischio, pertanto, è quello di dar luogo a un meccanismo di cooperazione transfrontaliera meramente orizzontale tra autorità nazionali che, affidandosi esclusivamente al principio di leale collaborazione, non si “verticalizzi” mediante il coinvolgimento di organi o istituzioni dell'Unione²³. In assenza di un soggetto terzo idoneo a dirimere gli eventuali conflitti tra tali autorità, potrebbe infatti instaurarsi una pericolosa *impasse* istituzionale, suscettibile di tradursi, in ultima analisi, nella mancata tutela dei fondamentali diritti politici dei cittadini dello Stato membro in cui si è verificata la violazione.

Per le autorità di controllo indipendenti di cui al Regolamento (UE) 2016/679, tale rischio è neutralizzato dall'esplicito riconoscimento della possibilità di «avvalersi della gamma di strumenti previsti dal regolamento (UE) 2016/679 per cooperare e prestarsi assistenza reciproca, compresa, se del caso, la procedura d'urgenza di cui all'articolo 66 di tale regolamento», sì da «agire tempestivamente per far rispettare i diritti degli interessati» (Considerando 94 Ttpa).

Diversamente, la medesima garanzia non può dirsi assicurata con riferimento alla cooperazione transfrontaliera tra le altre autorità nazionali e, in particolare, nell'ambito dei rapporti cooperativi tra coordinatori dei servizi digitali. Ai sensi del Considerando 91 Ttpa, infatti, la cooperazione transfrontaliera fra coordinatori dei servizi digitali «dovrebbe limitarsi alla cooperazione nazionale tra coordinatori dei servizi digitali, senza sconfinare a livello dell'Unione come prevede il regolamento (UE) 2022/2065».

Proprio con riferimento a tale profilo, l'apparato di *enforcement* previsto dal Ttpa si differenzia profondamente da quello del Dsa.

A norma dell'art. 58, par. 1, Dsa, qualora il coordinatore dei servizi digitali dello Stato membro in cui si trovano i destinatari del servizio abbia motivo di sospettare che un fornitore di servizi intermediari abbia violato il Regolamento con ripercussioni negative sui destinatari del servizio nel proprio Stato membro, questi può chiedere al coordinatore dei servizi digitali dello Stato membro sede di stabilimento del fornitore di esaminare la questione e adottare le misu-

22. Ivi, pp. 657 ss.

23. L. Lionello, *La reazione europea alle interferenze nei processi elettorali: il nuovo regolamento relativo alla trasparenza e al targeting della pubblicità politica*, in C. Schepisi (a cura di), *Unione europea, pluralismo e libertà dei media nell'era digitale*, Editoriale scientifica, pp. 277-297, in part. pp. 295-296.

re di indagine ed esecuzione necessarie, a condizione però che la Commissione non abbia già avviato un'indagine per la stessa violazione²⁴. A seguito della segnalazione, al pari di quanto previsto dal Ttpa, il coordinatore dei servizi digitali del luogo di stabilimento deve comunicare, senza indebito ritardo e in ogni caso entro due mesi dal ricevimento della richiesta, al coordinatore dei servizi digitali interessato e al comitato europeo per i servizi digitali la propria valutazione sulla presunta violazione, corredata da una spiegazione delle eventuali misure di indagine o di esecuzione adottate o che intende adottare.

Tuttavia, ai sensi dell'art. 59 Dsa, in mancanza di una comunicazione entro i termini previsti, o qualora il comitato non concordi con la valutazione o le misure adottate, il comitato ha facoltà di deferire la questione alla Commissione, fornendole tutte le informazioni necessarie²⁵.

Il deferimento alla Commissione non esautorà il coordinatore dei servizi digitali dello Stato membro di stabilimento, ma si configura piuttosto come un *referral*²⁶ dinanzi a un'istituzione terza²⁷ che, valutata la questione, si esprime comunicando il proprio parere al comitato e al coordinatore dei servizi digitali del luogo di stabilimento, chiedendo eventualmente a quest'ultimo di riesaminare la questione. Il coordinatore dei servizi digitali del luogo di stabilimento adotterà, quindi, le misure di indagine o di esecuzione necessarie per garantire la conformità al Dsa, tenendo «nella massima considerazione» ex art. 59, par. 3, il parere e la richiesta di riesame della Commissione e, entro due mesi dalla richiesta di riesame, sarà tenuto a informare quest'ultima, nonché l'autorità nazionale o il comitato (a seconda di chi abbia inoltrato la richiesta), in merito alle misure adottate o che intende adottare.

Analogamente, in materia di indagini comuni ex art. 60 Dsa, una volta che il coordinatore dei servizi digitali del luogo di stabilimento del prestatore abbia

24. Contrariamente a quanto stabilito nel Ttpa, la stessa possibilità (subordinata alle medesime condizioni) è concessa anche al comitato europeo dei servizi digitali, su domanda di almeno tre coordinatori dei servizi digitali del luogo di destinazione (ex art. 58, par. 2, Dsa).

25. R.A. Achleitner, *Enforcement: Powers of and cooperation with authorities*, in B. Steinrötter, C. Heinze, M. Deng (eds.), *EU Platform Law. A Handbook*, cit., pp. 203 ss.

26. Così R. Sabia, *L'enforcement pubblico del Digital Services Act tra Stati membri e Commissione europea: implementazione, monitoraggio e sanzioni*, in *MediaLaws*, n. 2, 2023, pp. 88-113, in part. pp. 100-102.

27. Sul punto, occorre tener conto della natura della Commissione europea che, in sede di *referral*, opera come soggetto "terzo" rispetto alle parti, ma non necessariamente come istituzione "imparziale" svolgendo pur sempre la funzione di organo esecutivo dell'Unione europea. Cfr. I. Buri, *A Regulator Caught Between Conflicting Policy Objectives: Reflections on the European Commission's Role as DSA Enforcer*, in J. van Hoboken et al. (eds.), *Putting the Digital Services Act into Practice Enforcement, Access to Justice, and Global Implications*, *Verfassungsblog*, 2023, pp. 75-89, in part. 79-82.

comunicato la propria posizione preliminare sulla presunta violazione a tutti i coordinatori dei servizi digitali, alla Commissione e al comitato, quest'ultimo può deferire la questione alla Commissione secondo il meccanismo di cui all'art. 59 Dsa se:

1. il coordinatore dei servizi digitali del luogo di stabilimento abbia omesso di comunicare la propria posizione preliminare entro il termine stabilito;
2. il comitato sia sostanzialmente in disaccordo con la posizione preliminare;
3. il coordinatore dei servizi digitali del luogo di stabilimento abbia omesso di avviare prontamente l'indagine congiunta a seguito della raccomandazione del comitato.

Tanto nella disciplina della cooperazione transnazionale, quanto in quella delle indagini comuni²⁸, dunque, l'apparato di *enforcement* del Dsa appare più completo ed efficace rispetto a quello del Ttpa in quanto prevede l'intervento della Commissione quale organo terzo chiamato a esprimersi sugli eventuali contrasti o divergenze che dovessero insorgere fra le diverse autorità nazionali, non pregiudicando tuttavia la scelta finale del coordinatore dei servizi digitali in ordine a se e quali misure di indagine o di esecuzione adottare.

3. Conclusioni

Dalle considerazioni svolte appare evidente che l'apparato di *governance* ed *enforcement* previsto dal Ttpa risulti ancora inadeguato e scarsamente efficace ai fini dell'effettiva tutela dei diritti politici dei cittadini europei.

L'analisi condotta ha rilevato, infatti, come non risultino ancora sufficientemente chiari i contorni della definizione di pubblicità politica, né adeguatamente definiti gli indici di identificazione dei messaggi di pubblicità politica "occulta" in violazione degli obblighi scaturenti dal Ttpa.

Sarebbe opportuno, pertanto, che la Commissione delinei con maggiore chiarezza, in sede di elaborazione degli orientamenti comuni ex art. 8 Ttpa, gli elementi costitutivi della pubblicità politica al fine di rafforzare la certezza giuridica e favorire l'uniformità e la coerenza normativa a livello europeo.

Inoltre, si è evidenziato come la previsione di un rigido e settorializzato riparto di competenze per materia fra le diverse autorità nazionali – secondo

28. Più in generale sulla *governance* del Dsa, M. Orofino, *Il Digital Service Act tra continuità (solo apparente) e innovazione*, in F. Pizzetti (a cura di), *La regolazione europea della società digitale*, cit., pp. 164-170.

quanto delineato nel Ttpa e poi specificato dalle normative dei singoli Stati membri – rischi di frammentare le discipline, a scapito dell’armonizzazione europea, ingessando un settore, quale quello digitale, caratterizzato da un così elevato grado di interconnessione da richiedere necessariamente una vigilanza e un controllo condiviso.

Nonostante l’intenzione del Legislatore unionale di favorire la cooperazione fra le autorità a livello nazionale, transnazionale ed europeo, dalla presente analisi è emersa un’involuzione rispetto alla precedente normativa europea in materia di servizi digitali: il Ttpa, infatti, a differenza del Dsa, non prevede omogenei ed efficaci meccanismi di risoluzione dei contrasti a livello unionale fra le autorità nazionali dei diversi Stati membri.

La suddetta criticità potrebbe tuttavia essere superata, almeno per le piattaforme e i motori di ricerca *online* di grandi dimensioni, qualificando i rischi posti dai servizi di pubblicità politica da essi forniti quali “rischi sistemici” ai sensi degli artt. 34 e 35 Dsa²⁹, consentendo così alla Commissione europea di intervenire, laddove necessario, attivando il *crisis response mechanism* ai sensi dell’art. 36 Dsa³⁰. Tuttavia, questo approccio risulta pur sempre insufficiente a garantire una tutela adeguata contro i rischi sistemici dei servizi di pubblicità politica forniti da quelle piattaforme e da quei motori di ricerca *online* che, non superando la soglia dimensionale minima per poter essere classificati come *Very large online platforms* (Vlop) o *Very large online search engines* (Vlose), necessitano di autonomi e specifici meccanismi di garanzia.

Dinanzi a uno scenario digitale privo di confini, dove l’*overlapping* normativo fra i diversi atti unionali e la sovrapposizione di competenze fra le diverse autorità rappresentano una caratteristica fisiologica intrinseca, risulta in ogni caso fondamentale assicurare che il sistema di *governance* ed *enforcement* sia chiaro, definito e soprattutto effettivo perché «the best rules on paper will not achieve much if regulators are not willing or able to enforce them»³¹.

Il Legislatore europeo deve, pertanto, favorire l’interconnessione e la cooperazione fra le autorità nazionali e fra queste e le istituzioni europee attraverso

29. In tal senso, Commissione europea, Comunicazione della Commissione, *Orientamenti della Commissione 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 a norma dell’articolo 35, paragrafo 3, del regolamento (UE) 2022/2065*, 26 aprile 2024, C/2024/3014. Cfr. G. Contissa, F. Galli, *La governance della “disinformazione aumentata” tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025, p. 143.

30. L. Lionello, *La reazione europea alle interferenze nei processi elettorali: il nuovo regolamento relativo alla trasparenza e al targeting della pubblicità politica*, cit., p. 296.

31. Così J. Jaursch, *Platform Oversight: Here is what a Strong Digital Services Coordinator Should Look Like*, in J. van Hoboken et al. (eds.), *Putting the Digital Services Act into Practice Enforcement, Access to Justice, and Global Implications*, cit., p. 93.

una *governance* strutturata che non resti confinata in una mera dichiarazione di intenti, ma si traduca in procedure e meccanismi collaborativi concreti, in linea di continuità con la precedente strategia europea culminata nella disciplina del Dsa. Soltanto a queste condizioni sarà davvero possibile assicurare la piena inclusività della sfera pubblica³², sì da tutelare effettivamente la libera formazione della volontà politica dei cittadini e preservare il carattere democratico dell'ordinamento in quanto la creazione del consenso è «presupposto non aggirabile per il funzionamento degli ordinamenti democratici»³³.

32. J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Raffaello Cortina Editore, 2023.

33. In questi termini, B. Caravita, *Il consenso ai tempi dei social*, in *federalismi.it*, n. 13, 2018, p. 3; anche in A. Poggi, F. Fabrizzi, F. Savastano (a cura di), *Social Network, formazione del consenso, intelligenza artificiale. Itinerario di un percorso di ricerca di Beniamino Caravita*, Sapienza University Press, 2023, p. 25.

Sezione IV

**SOLUZIONI REGOLATORIE E TECNICHE
NEL MONDO DEL DIGITALE**

THE GOVERNANCE OF “AUGMENTED DISINFORMATION” BETWEEN THE DIGITAL SERVICES ACT AND THE AI ACT

Federico Galli, Giuseppe Contissa***

SOMMARIO: 1. Misinformation and the problem with AI – 2. The solution with AI? – 3. The risks of “augmented disinformation” – 4. Augmented disinformation in the Digital Services Act – 4.1. Disinformation content as “illegal content” – 4.2. Automated moderation of disinformation content – 4.3. Disinformation and systemic risks – 4.4. Codes of conduct and disinformation – 5. Augmented disinformation in the AI Act – 5.1. Prohibited manipulation and disinformation – 5.2. Disinformation in systems posing a high risk to electoral influence – 5.3. Transparency and disinformation obligations – 5.4. General-purpose AI models: disinformation as a systemic risk – 6. Conclusions.

1. Misinformation and the problem with AI

The integration of AI into digital communication has revolutionised the way information is produced, distributed and received. This transformation has also brought considerable risks, especially in terms of the amplification and dissemination of false or manipulative content.

Today, AI plays a crucial role in amplifying disinformation¹ through the use

* Ricercatore di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

** Professore associato di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

1. We will not enter the debate of defining “disinformation” as well as of the inadequacy of the concept of “fake news”. See O. Pollicino, *Delegation of Regulatory Functions and Combating Systemic Risks Caused by Disinformation in the Digital Services Act: What Risks to Freedom of Expression?*, in *MediaLaws*, n. 3, 2023, pp. 114-43. As many now do, we prefer to approach the issue using the definition provided by the High-Level Expert Group, namely «false, inaccurate or misleading information, designed, presented and promoted with the intent to cause public harm or generate profit». High Level Expert Group on Fake News and Online Disinformation, *A Multi-Dimensional Approach to Disinformation: Report of the Independent High-Level Group on Fake News and Online Disinformation*, European Commission, 2018.

of recommendation algorithms, which structure the visibility of content based on engagement and emotional response metrics². Even when such algorithms are not designed with the intent to spread falsehoods, their operation incentivises the visibility of controversial, polarising or sensationalist content that may coincide with disinformation.

According to the University of Oxford's Computational Propaganda Research Project report, many platforms use algorithms that can unknowingly promote the spread of disinformation by rewarding content that generates more interactions, regardless of its accuracy³. Another study, published in *Science*, provided empirical evidence that fake news spreads faster and more widely than real news, precisely because of its emotional charge and novelty, which are two dimensions that AI algorithms leverage⁴.

Another critical vector of disinformation is the use of AI for user profiling and *micro-targeting* of disinformation content⁵. Digital platforms use AI and machine learning techniques to collect behavioural, demographic and psychometric data to build users' profiles, which are then used to target tailored content. This capability can be exploited for the benefit of end users (e.g., when positive and relevant content is targeted) or it can also be exploited for malicious purposes. In this case, for example, disinformation content can be "personalised" based on the biases, fears and prejudices of individuals, thus making it more credible and persuasive. The most well-known case is certainly that of Cambridge Analytica, where data collected via Facebook was used to influence political opinions through targeted campaigns⁶.

2. See, for example, V. Bakir, A. McStay, *Empathic Media, Emotional AI, and the Optimisation of Disinformation*, in M. Boler, E. Davis (eds.), *Affective Politics of Digital Media*, Routledge, 2020, pp. 263-79, where the authors analyse how "empathic media", driven by AI, allow biometric and behavioural signals – such as facial expressions, tone of voice, typing rhythm or social media interactions – to be exploited to spread disinformation.

3. S. Bradshaw, P.N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*, Oxford Internet Institute, University of Oxford, 2019, p. 20.

4. S. Vosoughi, D. Roy, S. Aral, *The Spread of True and False News Online*, in *Science*, n. 6380, 2018, pp. 1146-1151.

5. F. Lagioia, G. Sartor, *Profilazione e Decisione Algoritmica: Dal Mercato Alla Sfera Pubblica*, in *federalismi.it*, n. 11, 2020, pp. 85-110.

6. Among the various contributions that have analysed this case, see the interview conducted by Carole Cadwalladr with Christopher Wylie, the Cambridge Analytica whistleblower, who revealed how data from over 50 million Facebook profiles was collected without consent to feed a psychometric profiling system aimed at large-scale electoral manipulation. The investigation, published by The Guardian, highlighted the central role of the misuse of data and AI-assisted behavioural targeting techniques in influencing the political opinions of American voters during the 2016 elections, and represented a turning point in public awareness of the systemic risks associated with algorithmic surveillance and personalised disinformation. C. Cadwalladr,

In electoral or healthcare contexts, personalised disinformation can have a direct effect on individual choices, fuelling abstentionism, radicalisation or distrust of institutions. This is a delicate area, where AI does not simply spread disinformation, but also adapts it to the individual user, reinforcing *echo chambers* and further polarising public debate⁷.

AI techniques are now also used to create advanced bots, i.e. programmes capable of interacting with other online users by simulating human behaviour. These bots, which can operate on a large scale, are used to spread disinformation, manipulate trends, create fake “engagement” or fuel coordinated campaigns.

For example, it has been shown that bots played a central role in spreading disinformation during political events such as the 2016 American elections or the Brexit referendum⁸. Bots can also amplify the effect of “astroturfing”, i.e. simulating spontaneous grassroots consensus that is actually artificial and planned. In addition to spreading disinformation, bots can also play an active role in attacking legitimate sources, delegitimising journalists, scientists or activists through automated harassment campaigns, discrediting them or spreading conspiracy theories.

Disinformation is therefore not only conveyed through AI, but is directly produced by AI systems, contributing to the creation of an “information epidemic”⁹ or “information disorder”¹⁰ where media noise outweighs the merit of communication.

Finally, in the field of fake content production, the recent spread of generative AI models, such as Gpt, Claude or Gemini, in large-scale applications, such as ChatGpt, has taken the relationship between AI and disinformation to a new level. These systems, which enable the autonomous production of textual, visual and audiovisual content, are capable of creating fake news, manipulated narratives, realistic but non-existent images, and even videos in which public figures appear to say or do things that never happened (so-called *deepfakes*)¹¹.

E. Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, in *The Guardian*, 2018, www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

7. C.R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, 2017, p. 20.

8. E. Ferrara *et al.*, *The Rise of Social Bots*, in *Communications of the ACM*, n. 7, 2016, pp. 96-104.

9. C. Wardle, H. Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*, Council of Europe, 2017.

10. R. Bracciale, F. Grisolia, *Information Disorder: Acceleratori Tecnologici e Dinamiche Sociali*, in *federalismi.it*, n. 11, 2020, pp. 58-72.

11. A.M. Włorska, *Deepfakes and Disinformation*, Friedrich Naumann Foundation for Freedom, 2020.

According to Chesney and Citron, *deepfakes* represent an “epistemic threat”¹², as they undermine the very concept of “shared reality”: if everything can be realistically faked, it becomes difficult to believe even the truth. This effect is called the “*liar’s dividend*”, i.e. the possibility for anyone to deny documented facts by claiming that they are manipulated¹³.

At the same time, *large language models* (Llm) can be used to produce large-scale disinformation text content consistently and in a matter of seconds. As Luciano Floridi observes, the ability of these models to simulate human language without understanding the content generates a structural risk: the texts produced may be grammatically correct and stylistically persuasive, but the semantics of the messages and their veracity are completely detached from reality¹⁴. When used maliciously, they therefore become powerful tools for the large-scale production of propaganda or fake news.

2. The solution with AI?

Although artificial intelligence is in the spotlight for its role in spreading and creating disinformation, it is also a tool with extraordinary potential to counteract its negative effects. Today, most initiatives aimed at combating the spread of disinformation take place thanks to AI technologies.

One of the main contributions of AI to the fight against disinformation is its ability to automatically detect false or misleading content. Thanks to Natural Language Processing (Nlp) models, AI can analyse huge amounts of textual data to identify linguistic signals characteristic of disinformation: excessive generalisations, sensationalist statements, alarmist tones or arguments without verifiable sources.

These systems mainly rely on supervised classification models, which learn to distinguish between true and false news from large manually labelled datasets¹⁵. During training, a learning algorithm receives examples of

12. R. Chesney, D. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in *California Law Review*, n. 6, 2019, pp. 1753-1819.

13. K.J. Schiff, D.S. Schiff, N.S. Bueno, *The Liar’s Dividend: Can Politicians Claim Misinformation to Evade Accountability?*, in *American Political Science Review*, n. 1, 2025, pp. 71-90. The study highlights how this practice can compromise politicians’ accountability. In particular, the authors identify two main strategies: invoking informational uncertainty and encouraging opposition support from their supporters.

14. L. Floridi, *The False Promise of ChatGpt*, in *Philosophy & Technology*, n. 3, 2023, pp. 1-6.

15. For an up-to-date overview of *fake news* analysis models, see X. Zhou, R. Zafarani, *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*, in *ACM Computing Surveys*, n. 5, 2020, pp. 1-40.

previously verified content (labelled as “true” or “false”) and learns to identify correlations between textual characteristics (such as syntactic structure, the presence of hyperbolic terms, the absence of references to reliable sources, or the repeated use of polarising rhetoric) and the veracity of the content. Once trained, the model can be used in the operational phase to analyse new texts and estimate, with a certain degree of probability, whether they are potentially disinformation.

In addition, AI can compare the content of a message with databases of verified news or official statements to identify its plausibility. Several studies have confirmed the effectiveness of these approaches. For example, it has been shown that machine learning can be used to recognise fake news on social media with a good level of accuracy¹⁶. This preventive screening function, especially when integrated into platform moderation systems, is a powerful tool for limiting the viral spread of disinformation.

In this field, AI often plays a supporting role to professionals in the fact-checking industry, speeding up manual verification. AI systems can automatically retrieve official sources, previous articles or public data to compare with suspicious claims, significantly reducing the human workload. Some tools, such as ClaimBuster or Full Fact AI, already used by news outlets or in academic settings, allow verifiable statements to be identified in real time during political events or interviews.

Another area where AI has proven particularly effective is in mapping disinformation networks, i.e. identifying accounts, pages and groups that act in a coordinated manner to spread false or manipulative content. Through social network analysis and clustering techniques, algorithms can analyse content sharing patterns, publication timings and message repetitiveness to uncover suspicious connections between seemingly independent subjects¹⁷.

Finally, recent developments in generative AI have also found application in the field of disinformation control. For example, Large Language Models (Llm) have been progressively used as tools to verify suspicious claims, synthesise information from authoritative sources, and provide clear and contextualised explanations on controversial issues. Some of these models are capable of

16. K. Shu *et al.*, *Fake News Detection on Social Media: A Data Mining Perspective*, in *ACM SIGKDD Explorations Newsletter*, n. 1, 2017, pp. 22-36. The study highlights how the combined analysis of textual content, social data (such as interactions and retweets) and metadata related to sources can significantly improve the ability of predictive models to distinguish real news from fake news, paving the way for the integration of these tools into digital platform monitoring systems.

17. T. Papanastasiou, *Fake News Propagation and Detection: A Case Study*, in *Online Social Networks and Media*, n. 20, 2020, p. 100103.

processing a request in natural language, comparing it with reliable sources and producing a response that indicates whether or not a given statement is reliable.

An interesting case in point is Grok, the Llm-based chatbot integrated into the X platform (formerly Twitter), developed by xAI. Starting in 2024, several users began using Grok as a sort of on-demand fact-checker, querying it directly under controversial posts to obtain immediate verification of the content¹⁸. This practice has raised significant debates: on the one hand, it demonstrates growing confidence in the potential of generative AI as a tool against misinformation; on the other, it raises concerns about its reliability, transparency and possible spread of errors.

3. The risks of “augmented disinformation”

By “augmented disinformation” we mean the set of transformations that the phenomenon of disinformation undergoes as a result of the integration of AI into the processes of production, dissemination and management of information content. AI acts as a powerful multiplier, capable of making disinformation faster, more personalised and more convincing, while also offering effective tools to combat it. In this sense, it represents a true “Janus face” of the contemporary (dis)information system: on the one hand, it enables new threats, while on the other, it supports the defence of truthful information.

It is precisely this ambivalence that calls for careful reflection on how to govern the use of AI in combating disinformation, since the technologies used to protect the information ecosystem can, in turn, generate unintended, sometimes serious, distortive effects on the quality of information itself, fundamental rights, pluralism and the democratic quality of public debate.

AI systems used to combat disinformation are mainly based on machine learning techniques and natural language models, designed to identify recurring patterns in content and classify them according to their reliability. However, these systems do not operate through a deep semantic understanding of content, but through statistical correlations and probabilistic models trained on large volumes of data. This implies a number of significant risks ranging from technical errors to broader systemic effects.

18. J. Singh, *X Users Treating Grok Like a Fact-Checker Spark Concerns over Misinformation*, in *TechCrunch*, 2025, techcrunch.com/2025/03/19/x-users-treating-grok-like-a-fact-checker-spark-concerns-over-misinformation/.

One of the main risks concerns classification errors, in particular the possibility that legitimate content may be mistakenly labelled as false or harmful. Automatic moderation systems, even if advanced in terms of semantic language analysis (e.g. through word embeddings), do not understand the semantic referents of words, let alone the intention behind the dissemination of content. This could lead, for example, to the removal of satirical, critical or simply controversial content that is perfectly lawful; or to the removal of false content that the user believes to be true in good faith (which would be a case of misinformation). This is a phenomenon known as *overblocking* or *collateral censorship*, which can seriously compromise users' freedom of expression, especially in politically sensitive areas¹⁹.

Furthermore, AI's ability to grasp cultural, linguistic or rhetorical nuances is still limited, making it difficult to distinguish between misinformation and unconventional forms of opinion. In a digital space where the margin for error can have systemic effects, relying too heavily on automated decisions can be problematic, if not harmful.

Added to this is the well-documented risk of algorithmic bias²⁰. AI tools are trained on large amounts of data, and if this data reflects pre-existing biases (cultural, linguistic, ideological) it is likely that these distortions will be replicated or even amplified by the models. In the context of disinformation, this can mean that certain social groups or linguistic communities are disproportionately penalised²¹. Content conveyed through low-resource languages, by political or religious minorities, or by individuals who use non-standard registers of communication, may be considered less reliable, if not directly disinformative. This raises questions in terms of fairness and inclusivity, as algorithmic moderation could end up further marginalising voices that are already underrepresented in public debate.

19. J. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, in *University of California Davis Law Review*, n. 3, 2018, p. 1149. See also M. Durante, *Computational Power: The Impact of Ict on Law, Society and Knowledge*, Routledge, 2021, p. 124 where the Author analyses the various ways of "regulating" information by Internet intermediaries.

20. The literature on this subject is vast. Reference should be made to one of the first writings on the subject, C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, 2017; S. Quintarelli, et al., *AI: Ethical Profiles. An Ethical Perspective on Artificial Intelligence: Principles, Rights and Recommendations*, in *BioLaw Journal*, n. 3, 2019, pp. 159-177.

21. See S. Park et al., *AI Chatbots and Linguistic Injustice*, in *Journal of Universal Language*, n. 1, 2024, pp. 99-119, which highlights how AI-based chatbots can reinforce linguistic injustices, penalising speakers of minority or non-standard languages in interactions mediated by artificial intelligence.

A further problem is the opacity that characterises many AI systems, particularly those based on deep neural networks. These technologies are often described as “black boxes” precisely because, despite offering high performance, they do not make it easy to understand the logical path that led to a particular decision²². This means that if content is automatically flagged as “misinformation”, a user may see their post removed or penalised without knowing exactly why and without having the means to effectively challenge the decision. This raises not only transparency issues, but also principles of procedural justice, as any control system should provide for the possibility of review and appeal.

The use of AI to select, filter or remove disinformation content can have a significant impact on information pluralism, even in the absence of malicious intent. Where artificial intelligence is trained to promote content considered more “safe” or “authoritative”, there is a risk that the variety of opinions will be drastically reduced, especially in contexts where the truth is not unambiguous but subject to debate²³. This is particularly true in areas such as political communication, science dissemination practices or economic theories where there are legitimate differences in the interpretation of facts. If algorithms tend to favour institutional or “mainstream” sources, critical content or alternative voices risk being suppressed even when they are not examples of disinformation at all. The danger, therefore, is that AI will become a tool for standardising public discourse, where complexity is sacrificed in the name of information security.

Finally, the automation of the fight against disinformation presents the risk (in today’s political context, far from hypothetical) of instrumental use by private or public actors. On the one hand, online platform providers could use AI systems to moderate content in accordance with their economic or commercial interests, penalising information that hinders their political agenda or that of their investors. On the other hand, authoritarian governments (but not only) could exploit the automatic labelling of disinformation to target political opponents, activists or journalists deemed inconvenient, masking what is in reality likely to be selective censorship as “online environment security”. In this sense, AI is an ambiguous technology: it can protect democratic debate, but it can also be used to manipulate it if it is not accompanied by appropriate legal safeguards, independent oversight bodies and civil society participation.

22. F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.

23. G. De Gregorio, N. Stremlau, *Inequalities and Content Moderation*, in *Global Policy*, n. 5, 2023, pp. 870-879.

4. Augmented disinformation in the Digital Services Act

EU Regulation 2022/2065, known as the “Digital Services Act” (Dsa), was officially adopted on 19 October 2022 and entered into force on 16 November 2022, with full application from 17 February 2024, with the exception of certain obligations for large platforms applied from August 2023²⁴. Included alongside the Digital Markets Act (Dma) in the broader European strategic plan for regulating the digital space, the Dsa aims to provide a regulatory framework to ensure a safer, more transparent online ecosystem that respects fundamental rights²⁵.

In public debate, the Dsa is often presented as one of, if not the, new legal frameworks created by the EU to combat online disinformation²⁶. On closer inspection, however, the Dsa does not address the issue of automatic moderation of disinformation in a direct and systematic manner²⁷.

Firstly, the term “disinformation” never appears in the text of the Dsa. The only occurrences are contained in the recitals, which refer to the concept in combination with other semantically related terms: “online disinformation” (Recital 2), “disinformation campaigns” (Recital 69) and “coordinated disinformation campaigns” (Recital 83).

Moreover, in these cases, the uses of the term “disinformation” seem to reflect different meanings. For example, Recital 2 refers to “online disinformation” as a “social risk” and Recital 83 mentions “coordinated disinformation campaigns” as an example of “systemic risks” related to public health. Meanwhile, Recital 69 refers to “manipulative techniques” that contribute to disinformation campaigns and «can negatively impact entire groups and amplify societal harms».

In reality, as we will see in the following paragraphs, the phenomenon of disinformation, particularly in the sense discussed in the previous section, has an indirect relevance to the application of the Dsa.

24. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277, 27.10.2022, pp. 1-102.

25. F. Casolari, *Il “Digital Services Act” e La Costituzionalizzazione Dello Spazio Digitale Europeo*, in *Giurisprudenza Italiana*, n. 2, 2024, pp. 462-465.

26. Z. Meyers, *Will the Digital Services Act Save Europe from Disinformation?*, Centre for European Reform, 20 April 2023, www.cer.eu/insights/will-digital-services-act-save-europe-disinformation; Deutsche Telekom Ag, *Digital Services Act (Dsa)*, Deutsche Telekom, 13 July 2023, www.telekom.com/en/company/details/digital-services-act-dsa-1060552; C. Goujard, *EU Wants Big Tech to Help Deepfake-Proof the 2024 Election*, Politico, 14 February 2024, www.politico.eu/article/eu-big-tech-help-deepfake-proof-election-2024.

27. M. Husovec, *The Digital Services Act’s Red Line: What the Commission Can and Cannot Do about Disinformation*, in *Journal of Media Law*, n. 1, 2024, pp. 47-56.

The starting point for this line of reasoning is that the phenomenon can be examined in the Dsa, first and foremost, in terms of its material component, as it manifests itself in the publication of unlawful, or rather “illegal”, content, which intermediary service providers are required to remove diligently. We will see, in fact, that this possibility may also apply to disinformation content that is not unlawful but is incompatible with the terms of use of the service.

In such cases, service providers may – and in a sense are encouraged to – lawfully initiate voluntary investigations to locate, identify and remove content, provided that this is done in compliance with the due diligence obligations set out in Chapter III of the Regulation. We will see that these specifically consider cases where “content moderation” is carried out with the aid of automated systems and, in this regard, provide certain safeguards for users.

Finally, for service providers that are Very Large Online Platforms (Vlop) or Very Large Online Search Engines (Vlose), disinformation is considered a potential systemic risk, particularly in relation to the design or operation of their service and its related systems. In such cases, Vlop and Vlose must carry out risk assessment and mitigation activities and comply with a number of specific obligations.

4.1. Disinformation content as “illegal content”

Article 3(h) of the Dsa defines “illegal content” as «any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law».

Given this definition, a key point in establishing the scope of the Dsa is to understand the extent to which the law of a Member State classifies disinformation as illegal content.

We know, for example, that there is no legislation in the Italian legal system that directly classifies disinformation as illegal content. However, there are sectoral regulations that could be relevant. A case in point is the dissemination of false news likely to disturb public order (e.g. Article 656 of the Italian Criminal Code)²⁸, or the legislation on misleading advertising, i.e. advertising

28. See S. Flaminio, *Lotta Alle Fake News: Dallo Stato Dell'arte a Una Prospettiva Di Regolamentazione Per Il “Vivere Digitale” a Margine Del Digital Services Act*, in *Rivista Italiana di Informatica e Diritto*, n. 2, 2022, p. 82, on other criminal offences that may involve the dissemination of false news.

that provides false information that may distort the behaviour of the average consumer (Article 20 of the Consumer Code). In such contexts, disinformation may take on legal significance and constitute unlawful content.

Another relevant example concerns recent legislation in some European countries regarding the dissemination of false information about the effects of Covid-19 vaccines. For example, in Germany, specific measures have been introduced to combat health misinformation, particularly on social media, with penalties for the dissemination of content that incites mistrust of vaccination campaigns. In countries such as Poland, on the other hand, the approach has been more cautious and less restrictive, leaving more room for freedom of expression even on controversial issues.

Phenomena such as these raise considerable difficulties in the uniform implementation of Dsa regulations, as online platforms may find themselves in a position where they have to remove content in one country while allowing it to be disseminated in another.

If misinformation is classified as illegal content under national law, then the provisions of the Dsa apply, relating, on the one hand, to the exemption from liability for intermediary service providers and, on the other, to specific obligations in combating illegal content.

With regard to liability, Articles 4, 5 and 6 apply, which reproduce, essentially unchanged from the E-Commerce Directive²⁹, the conditions under which intermediary service providers (in particular providers of mere conduit, temporary storage and hosting services, including online platforms) may benefit from immunity from non-contractual liability for illegal content transmitted or stored through their services. In the case of hosting service providers, for example, immunity applies provided that the provider is not aware of the illegal activity or content or, as soon as it becomes aware of the illegality, acts promptly to remove it or disable access to it.

In the case of illegal disinformation content, this means that a provider can only be exempt from liability if it acts promptly once it has received a report from a user or a notification from the competent authority. In the case of an order issued by a competent authority (whether judicial or administrative), Article 9 of the Dsa applies in particular, requiring intermediary service providers to cooperate with the authorities in combating illegal content. In particular, they must designate a contact point to receive removal orders issued by the authorities of Member States and act in accordance with such requests.

29. G. Giordano, *La responsabilità degli intermediari digitali nell'architettura del "Digital Services Act": è necessario che tutto cambi affinché tutto rimanga com'è?*, in *Comparazione e Diritto Civile*, n. 1, 2023, pp. 193-221.

In addition to the obligations to remove content following receipt of an order from the authorities, the Dsa has introduced a “notice and action” mechanism for hosting service providers, which was not included in the 2000 Directive³⁰. This procedure should allow any user of the service to report content that is considered illegal. The report must contain a sufficiently detailed explanation, allowing the provider to assess the illegal nature of the content, and must be accompanied by the information necessary to accurately identify the subject of the request. If the report meets all the formal criteria, it has the legal effect of giving the provider actual knowledge of the unlawful content, within the meaning of Article 6 of the Dsa, thereby triggering its obligation to take prompt action.

The effectiveness of the reporting system is reinforced by Article 22, which introduces the concept of “trusted flaggers”. These generally are bodies with proven experience and expertise in specific areas, which are recognised by the Member State’s digital services coordinator as having the capacity to report illegal content through a priority and preferential channel. Once online platforms receive a report from a trusted flagger, they are obliged to follow up on it as a matter of priority, ensuring faster response times than usual.

The provision of trusted flaggers also plays a central role in combating disinformation that is classified as illegal, as it allows specialised bodies (such as consumer protection organisations, accredited fact-checkers or health authorities) to promptly activate platforms in the event of false content that is potentially harmful to the public interest. However, the effectiveness of the mechanism depends on the quality and impartiality of the designated entities, as well as on the platforms’ ability to manage the volume of reports received efficiently and non-arbitrarily.

4.2. Automated moderation of disinformation content

One of the main changes in the Dsa concerning the liability of Isp with respect to the e-commerce Directive is the so-called “good Samaritan clause” provided for in Article 7³¹. This provision establishes that Isp do not automatically lose the possibility of claiming exemption from liability if they voluntarily decide to undertake content moderation activities, provided that

30. In particular, Article 17 of the Dsa requires hosting service providers to make available an easily accessible and usable mechanism that allows any natural or legal person to report content deemed illegal.

31. M.A. Astone, “*Digital Services Act*” e nuovo quadro di esenzione dalla responsabilità dei prestatori di servizi intermediari: quali prospettive?, in *Contratto e Impresa*, n. 4/2022, pp. 1050-1063.

such activities are carried out in good faith and in a diligent, proportionate and non-discriminatory manner. This provision therefore makes the immunity regime compatible with voluntary practices that Isp may implement to moderate user content, when carried out in good faith and in compliance with due diligence obligations.

The Dsa pays particular attention to this last aspect, which – it can be said – represents the real novelty of the Regulation compared to the E-Commerce Directive³². The Regulation, in fact, provides for a series of due diligence obligations applicable to service providers, aimed at ensuring that the exercise of moderation power is not arbitrary, opaque or detrimental to the fundamental rights of users³³. In this sense, content moderation is not only a technical faculty, but a regulated function subject to principles of transparency and contestability of appeal.

Content moderation is defined in Article 3(t) as a set of «the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service». These include all «measures taken that affect the availability, visibility and accessibility of such illegal content or information, such as its demotion, demonetisation or removal, or the disabling of access to it, or that affect the ability of service recipients to provide such information, such as the termination or suspension of a service recipient's account».

A first important aspect to consider is that moderation practices extend not only to illegal content but also to content that violates the general terms and conditions³⁴. These are the content guidelines that the intermediary service sets out in its terms of use, which often include ethical criteria, community standards or specific policies (e.g. against hate speech, violent content, ecc.). This means that disinformation content may fall within the scope of the Dsa, not only as we have seen above, because it is illegal under EU or Member State law, but also when it violates the internal rules defined unilaterally by the platform³⁵.

32. On the paradigm shift from directive to regulation, see M.C. Buiten, *The Digital Services Act: From Intermediary Liability to Platform Regulation*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, n. 12, 2022, pp. 361-380.

33. L. D'Agostino, *Disinformazione e obblighi di compliance degli operatori del mercato digitale alla luce del nuovo Digital Services Act*, in *MediaLaw*, n. 2, 2023.

34. See the contribution by A. Michinelli, *La gestione dei contenuti: illegali e non, la loro moderazione*, in L. Bolognini, E. Pelino, M. Scialdone (a cura di), *Digital Services Act e Digital Markets Act*, Giuffrè, 2023, pp. 131-163.

35. We refer to J.P. Quintais, N. Appleman, R.Ó. Fathaigh, *Using Terms and Conditions to*

In fact, in recent years, many platforms have adopted more or less stringent policies for moderating disinformation content. Facebook/Meta states in its “community standards” that it will remove various types of disinformation (defined as «a statement that has been judged false by an authoritative third party’), including disinformation about vaccines (e.g., that vaccines cause autism), disinformation during public health emergencies, and disinformation that interferes with voters (e.g., disinformation that a person is a candidate in an election)»³⁶. YouTube’s “community standards” also focus on similar instances of misinformation, such as videos that deny the effectiveness of vaccines or promote false cures for serious diseases, or manipulated videos, i.e. old footage passed off as current or videos altered to make politicians appear to say or do things that never happened³⁷.

Secondly, from the definition of content moderation, it is clear that the Dsa explicitly recognises («automated or non-automated activities») the role of automated processes in identifying and managing potentially harmful content. More specifically, the Dsa not only recognises the legitimate use of automated technologies in moderation, but also openly acknowledges the risks in terms of errors³⁸, distortions and unjustified impacts on users’ rights³⁹.

In this sense, due diligence obligations are an important point that concerns not only moderation practices as such, but also how service providers design and use the systems involved in moderation⁴⁰.

Apply Fundamental Rights to Content Moderation, German Law Journal, n. 5, 2023, pp. 881-911, which raises the issue of the compatibility with the protection of fundamental rights of online content moderation processes based on platform terms of use. The Authors propose an innovative approach that should lead providers to integrate principles of public law into private law instruments, offering a guarantee for users without the need for direct state intervention.

36. Meta, *Community Standards* (April 2025).

37. YouTube, *Community Guidelines* (April 2025).

38. Recital 26: «The condition of acting in good faith and in a diligent manner should include acting in an objective, non-discriminatory and proportionate manner, with due regard to the rights and legitimate interests of all parties involved, and providing the necessary safeguards against unjustified removal of legal content, in accordance with the objective and requirements of this Regulation. To that aim, the providers concerned should, for example, take reasonable measures to ensure that, where automated tools are used to conduct such activities, the relevant technology is sufficiently reliable to limit to the maximum extent possible the rate of errors».

39. Recital 58: «providers of online platforms should be required to provide for internal complaint-handling systems, which meet certain conditions that aim to ensure that the systems are easily accessible and lead to swift, non-discriminatory, non-arbitrary and fair outcomes, and are subject to human review where automated means are used. Such systems should enable all recipients of the service to lodge a complaint and should not set formal requirements, such as referral to specific, relevant legal provisions or elaborate legal explanations».

40. See, for example, K. Söderlund *et al.*, *Regulating High-Reach AI: On Transparency*

For example, Article 14 requires all service providers to include in their terms and conditions information on the restrictions they impose on the use of their services. Such information must cover the procedures, measures and “tools” used for moderation, including “algorithmic decision-making” and “human verification”.

Further information on automated processes must be provided in accordance with Article 15, which requires the periodic publication of transparency reports. All intermediary service providers must publish meaningful and understandable information concerning moderation activities, including the use of automated tools (letter *c*). This description must also include a qualitative description, a description of the precise purposes, accuracy indicators and possible error rates of the automated tools used in pursuit of these purposes, and any safeguards applied (letter *e*).

For example, in its latest transparency report covering the first half of 2024, Pinterest states that it removed most content that violated community rules using automated systems⁴¹. For example, in the “medical misinformation” category, 100 pieces of content were removed automatically and 70.137 were removed using a hybrid approach, i.e. with subsequent human review. In the “climate misinformation” category, on the other hand, there were no fully automated removals, but 27 were removed entirely by humans and 4.116 with the involvement of an automated system. Pinterest specifically states that it uses «cutting-edge *machine learning* models, continuously updated through the addition of new data and the adoption of innovative technical solutions», and subjects them to «rigorous offline and online testing», without providing details on accuracy.

The issue of information on automatic moderation is linked to the long-standing and still debated issue of the transparency of automated decisions, which, for example, has been addressed in the General Data Protection Regulation (Gdpr)⁴². In that context, there has been discussion, in particular, of when and to what extent the data controller should provide information on the logic used in automated decision-making processes⁴³, as well as on the

Directions in the Digital Services Act, in *Internet Policy Review*, n. 1, 2024, pp. 1-31, which analyses the impact of the Dsa rules on the transparency of algorithmic systems and highlights how these obligations clash with the technical complexity of AI systems.

41. Pinterest, *Digital Services Act Transparency Report* (April 2025).

42. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

43. The reference is to Article 13(2)(f) in conjunction with Article 22(3). For a commentary on Article 22 of the Gdpr, see F. Lagioia, G. Sartor, A. Simoncini, *Article 22*, in *Codice Della*

importance of ensuring that such information is understandable and accessible to the data subject.

In addition to *ex ante* transparency obligations, i.e. at the stage of preparing and publishing the terms of service, the Dsa also deals with protecting users when they are subject to a specific moderation decision. In particular, Article 17 introduces an obligation for storage service providers (hosting), including online platforms, to provide a clear statement of reason for any decision to remove or restrict access to content. In addition to providing information on the type of removal and the facts or content on which the decision is based, this statement must indicate whether the decision is taken by automated means and provide an «explanation of the reasons why the information is considered illegal content» (letter *d*) or «incompatible with the terms of the contract» (letter *e*). For online platforms only, there is also an obligation under Article 20 to provide access to a complaint management system that allows the removal decision to be challenged. The decision on the complaint cannot be made solely through automated means (Article 20(6)).

Therefore, even when specific disinformation content is removed through automated tools, the platform should inform the end user of this. This raises the issue of providing a detailed explanation of the automated decision in relation to the specific case. This is relevant to the debate on the explainability of automated decisions, which is a central theme of both the Gdpr (Article 22) and the AI Act (Article 86). The key question is whether and to what extent platforms should be able to provide a comprehensible and meaningful explanation of how the algorithms used to moderate content work. The problem is not only regulatory but also technical: as mentioned above, many AI systems, particularly those based on deep learning models, operate as “black boxes” and are not easily interpretable even by their developers.

The letter of the Dsa seems to lean towards a functionalist interpretation of the obligation to explain: it does not necessarily require a technical description of the algorithm, but rather information useful for understanding the impact of automation on the decision made. This may include, for example, an indication of the type of model used (e.g. text classifier, linguistic pattern recognition), the degree of automation of the process (fully automated or with human supervision), or the associated statistical

Privacy e Data Protection, R. D’Orazio *et al.* (eds.), Giuffrè, Milan, 2021, pp. 388-403. See also M. Brkan, G. Bonnet, *Legal and Technical Feasibility of the Gdpr’s Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas*, in *European Journal of Risk Regulation*, n. 1, 2020, pp. 18-50.

margin of error. In other words, the explanation must be sufficiently detailed to allow the user to challenge the decision in an informed manner, especially when it comes to *borderline* content, such as disinformation that is not manifestly unlawful.

4.3. Disinformation and systemic risks

In addition to moderation obligations, some platforms are required to prevent the dissemination of content that does not necessarily constitute illegal content under EU law or the national laws of Member States, nor content that is incompatible with the terms of service. This applies in particular to online intermediaries that have been designated by the European Commission as Vlop or Vlose, as they have an average of at least 45 million monthly active users in the Union and have therefore been classified as such by the European Commission⁴⁴.

The main obligation for Vlop and Vlose is to periodically assess the “systemic risks” arising from the operation or use of their services, including algorithmic systems.

The systemic risks mentioned include:

1. the dissemination of illegal content;
2. negative effects (current or foreseeable) on the exercise of fundamental rights enshrined in the Charter;
3. negative effects (current or foreseeable) on civic debate, electoral processes and public security;
4. negative effects (current or foreseeable) on gender-based violence, the protection of public health and minors, and serious negative consequences for the physical and mental well-being of individuals.

In the risk assessment process, Vlop and Vlose must consider in particular how various aspects of the service affect these risks, such as the structure and functioning of their recommendation systems, as well as other relevant algorithmic systems; how they manage and moderate content on the platform; the terms of service applicable to users and how they are actually enforced; the criteria by which advertisements are selected and displayed; the policies adopted for the collection and use of user data.

In addition, the assessment must include an analysis of the impact that may result from manipulative use of the service by users.

44. Article 33 Dsa.

The phenomenon of “augmented disinformation” appears to be a central element in the assessment of systemic risks⁴⁵. Among these, it can be abstractly traced back to all four categories of systemic risks provided for in Article 33 of the Dsa.

The first case (letter *a*) concerns the specific situation, discussed in the previous subparagraphs, in which disinformation is examined as a hypothesis of illegal content, as it violates unlawful provisions of national or EU law.

Furthermore, disinformation can be identified as a systemic risk in that it is a phenomenon that has a negative effect, even if only potential, on socially relevant interests (letters *c* and *d*), such as civic debate and electoral processes, but also the protection of public health and mental well-being. Examples include the spread of coordinated anti-vaccine campaigns during the Covid-19 pandemic or the massive circulation of false or misleading content during electoral events, which can polarise public opinion or undermine trust in institutions.

Finally, the moderation of disinformation may also be considered detrimental to the exercise of certain fundamental rights recognised by the Charter (letter *b*).

Obviously, the main point of friction is with freedom of expression and information, including pluralism of information (Article 11), given the detrimental effects of disinformation on public debate and the potential distortion of the perception of reality⁴⁶. But consider also Article 8, which protects the right to personal data protection, which can be compromised when disinformation content concerns a natural person and causes damage to their reputation. Or Article 14, which protects the right to education, jeopardised by the spread of false news on scientific and educational issues.

In the latter case, moreover, it is not only the problem of disinformation as such that is highlighted, but also the impact on fundamental rights of moderation systems on lawful forms of expression. This is the case, for example, with controversial opinions, fringe theories or institutional criticism which, although not based on false data, could be removed by inaccurate algorithms or automated systems unable to distinguish between harmful content and content that is simply not aligned.

Once systemic risks have been identified, Vlop and Vlose are required

45. See E. Longo, *Freedom of Information and the Fight Against Disinformation in the Digital Services Act*, in *Giornale di Diritto Amministrativo*, n. 6, 2023, pp. 737-745.

46. See on this point, A. Palumbo, J. Piemonte, *Delega di funzioni regolamentari e lotta ai rischi sistemici causati dalla disinformazione nel Digital Services Act: quali rischi per la libertà di espressione?*, in *MediaLaw*, n. 2, 2023, which raise doubts about how adequate this co-regulation model is in ensuring freedom of expression for platform users.

to take reasonable, proportionate and effective measures to mitigate those risks. Such measures may include, among other things, adjusting content moderation procedures and all relevant decision-making processes, allocating resources dedicated to moderation, reviewing the general terms and conditions of service and their effective enforcement, and testing and adjusting the algorithmic systems used for content recommendation, ranking and dissemination.

In this regard, the Commission's Guidelines on mitigating systemic risks to electoral processes, published in the run-up to the European elections on 26 April 2024, were particularly important. The document provided operational guidance for Vlop and Vlose relevant to identifying and mitigating risks related to disinformation before, during and after the European election period.

Among other things, platforms were invited to strengthen their internal processes by setting up specialised teams to detect risks specific to national contexts and assess how users use their services during the pre-election period. In addition, platforms were required to take specific measures for each national and electoral context to promote election information from official sources, literacy campaigns, and the adaptation of recommendation systems to enable users to form their own opinions freely and knowledgeably.

Cooperation with national and European authorities, independent experts and civil society is also essential, as is the establishment of rapid response mechanisms in the event of significant incidents. Finally, non-confidential post-election reviews will be published to promote transparency and public feedback.

Finally, specific attention has been paid to the risks associated with generative AI. Platforms and search engines that enable the creation or dissemination of AI-generated content, such as deepfake videos or automated texts, are required to proactively assess the associated risks and take appropriate mitigation measures. These include clear and visible labelling of AI-generated content, updating terms of use to include these aspects, and consistent and effective enforcement of these rules.

Another key tool for ensuring Vlop and Vlose comply with systemic risk management obligations is the independent audits required by Article 37 of the Dsa. These must be conducted at least annually by qualified and independent third parties, who are tasked with assessing the effectiveness of the measures taken to identify, analyse and mitigate systemic risks. The audits cover the entire risk governance system, including content moderation processes, algorithmic mechanisms, advertising policies and interaction with terms of service.

In close connection with these obligations, Article 42 of the Dsa requires Vlop and Vlose to publish transparency reports on the audits conducted⁴⁷. In these reports, platforms must clearly indicate the results of independent audits, any critical issues identified and the corrective measures taken or planned. These reports must be written in a clear, accessible and understandable manner in order to ensure public and institutional oversight of the activities of large digital platforms⁴⁸.

4.4. Codes of conduct and disinformation

Article 45 of the Digital Services Act provides for the possibility for the European Commission to promote and approve voluntary codes of conduct, intended in particular for Vlop and Vlose as tools to contribute to the identification, assessment and mitigation of systemic risks. The risks covered expressly include disinformation, especially where it adversely affects public debate, electoral processes or public health.

As mentioned above, one of the first significant instruments in this area was the 2018 Code of Practice on Disinformation, promoted by the European Commission and signed by several large digital platforms, such as Facebook, Google, Twitter and Mozilla⁴⁹. Although voluntary and non-binding, this code marked a first attempt at coordinated action to combat the spread of disinformation online, through measures such as advertising transparency, the closure of fake accounts and the promotion of reliable sources.

In 2022, the code was strengthened and transformed into the new “Strengthened Code of Practice on Disinformation”, with a larger number of signatories, including TikTok, LinkedIn, Adobe and other industry players, and a more structured framework⁵⁰. The new code introduced measurable targets, periodic reporting obligations, and a more rigorous monitoring mechanism. It incorporated specific recommendations to address the use of artificial

47. Added to this obligation is that of Article 40, which requires platforms to provide access to meaningful and relevant data to independent research bodies designated by the Commission or Member States in order to enable the monitoring of systemic risks.

48. Of interest in this regard is the position taken by P. Leerssen, *Outside the Black Box: From Algorithmic Transparency to Platform Observability in the Digital Services Act*, in *Weizenbaum Journal of the Digital Society*, n. 2, 2024, which discusses how the defining objective of the Dsa’s regulatory philosophy is the “observability” of platforms rather than the transparency of the algorithms they use in the provision of services.

49. European Commission, Code of Practice on Disinformation, October 2018.

50. European Commission, Strengthened Code of Practice on Disinformation, June 2022. See also: European Commission, *Questions and Answers: 2022 Strengthened Code of Practice on Disinformation*, 16 June 2022, ec.europa.eu/commission/presscorner/detail/en/qanda_22_3665.

intelligence in the production of disinformation, algorithmic transparency, and the protection of electoral integrity.

After the Dsa came into force, it was assumed that the Code of Practice would soon evolve into an officially recognised code of conduct under Article 45 of the Regulation. This transformation recently took place on 13 February 2025, when the European Commission and the European Committee for Digital Services approved the integration of the 2022 Code into the regulatory framework of the Dsa, giving it the status of Code of Conduct on Disinformation.

The Code of Practices is now the main reference point for the assessment of systemic risks in relation to disinformation. It follows that adherence to the Code of Practice on Disinformation may constitute, for Vlop and Vlose, a parameter for competent authorities (such as the Commission or national digital services coordinators) to assess compliance with the due diligence obligations under the Dsa in relation to systemic risk management.

5. Augmented disinformation in the AI Act

Regulation 2024/1689, known as the Artificial Intelligence Act (AI Act), was approved on 13 March 2024 and entered into force on 2 August 2024, with its provisions to be phased in over the following 6 to 36 months⁵¹.

The AI Act represents the first comprehensive legislative proposal at global level aimed at regulating artificial intelligence systems according to a risk-based approach: in particular, based on the risk that different types of use may pose to the safety, health and fundamental rights of European citizens, and outlines a system of more or less stringent rules for each class⁵².

From the point of view of the architecture of the regulation, this approach takes the form of a multifaceted classification of AI systems into four risk levels:

51. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) n. 300/2008, (EU) n. 167/2013, (EU) n. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Regulation on artificial intelligence), OJ L, 2024/1689 of 12.7.2024.

52. An initial analysis of the risk approach in the Regulation T. Mahler, *Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal*, in L. Colonna S. Greenstein (a cura di), *Nordic Yearbook of Law and Informatics 2020-2021*, The Swedish Law and Informatics Research Institute, Stockholm, 2021, pp. 247-270. For critical observations, see C. Novelli, *The European Artificial Intelligence Act: Some Implementation Issues*, in *federalismi.it*, n. 2, 2024, pp. 95-113.

1. systems that exhibit unacceptable forms of risk, the development or use of which is prohibited;
2. high-risk systems, for which a series of compliance requirements are established, accompanied by obligations for suppliers and deployers;
3. certain AI systems, for which transparency obligations are established (associated with limited risk);
4. systems with minimal risk. There is an additional special risk category for general-purpose AI models (with stricter rules for general-purpose AI models with systemic risk, which differs from that of general-purpose AI models without systemic risk).

Like the Dsa, the AI Act does not systematically address the relationship between AI and disinformation. In particular, the AI Act does not seem to be concerned with AI as a tool for combating disinformation, focusing instead on preventing the harmful effects of AI in the production of disinformation.

Among the objectives of the AI Act listed in Article 1, the European legislator indicates the intention to guarantee fundamental rights «including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union».

In this context, disinformation – particularly that mediated by advanced technologies such as deepfakes or generative AI models – is recognised as one of the main systemic challenges that AI can amplify. The European legislator has therefore introduced a series of specific provisions into the regulation aimed at mitigating its impact, particularly in relation to the protection of the democratic process, public security and fundamental rights.

5.1. Prohibited manipulation and disinformation

The first case is the use of AI to generate disinformation as part of a practice that aims to manipulate human behaviour, compromising the ability to make informed decisions and potentially causing harm. In this regard, Article 5 of the AI Act, which lists prohibited practices, is relevant, and in particular Article 5(1)(a), which states the following:

1. The following AI practices shall be prohibited: (a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or

is reasonably likely to cause that person, another person or group of persons significant harm.

In relation to the issue of disinformation, the techniques that the legislator refers to as “deceptive” are particularly relevant. Although the AI Act does not provide a definition of “deceptive techniques”, Recital 29 specifies that deceptive techniques are those that «subvert or impair person’s autonomy, decision-making or free choice in ways that people are not consciously aware of those techniques or, where they are aware of them, can still be deceived or are not able to control or resist them».

Furthermore, the Guidelines on Prohibited Practices, recently adopted by the European Commission⁵³, clarify in paragraph 70 that “deceptive techniques” implemented by AI systems should be understood as the presentation of false or misleading information with the aim or effect of deceiving individuals and influencing their behaviour in a way that undermines their autonomy, decision-making and free choices.

The same Guidelines also emphasise the connection between the regulation of misleading practices under Article 5(1)(a) and the information requirements under Article 50 (see section 5.4 below). In particular, it is emphasised that the obligation under Article 50(4) for the deployer to label “deep fakes” and AI-generated texts when they concern matters of public interest, the obligation under Article 50(1) to ensure that AI systems that interact with people are designed in such a way as to inform people that they are interacting with AI and not with a human being, and, finally, the technical measures under Article 50(2) that enable the detection of AI-generated and manipulated content, are all to be considered as mitigation measures against deceptive practices, measures that should also be made effective through appropriate interventions by the AI system provider on the design of the system itself.

In particular, according to the Guidelines (paragraph 71), the explicit labelling of “deep fakes” and chatbots reduces the risk of deception that could arise once AI-generated content is disseminated to the public and the associated and consequent risk of harmful distortive effects on the formation of opinions and beliefs and on the behaviour of individuals.

Among the prohibited practices, another provision potentially relevant to the issue of disinformation is that of Article 5(1)(b), which expressly prohibits the use of AI systems that exploit vulnerabilities related to the age, disability

53. European Commission, *Annex to the Communication to the Commission – Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, Brussels, 4.2.2025 C(2025) 884 final.

or socio-economic situation of specific groups, with the aim – or effect – of materially distorting their behaviour in a harmful way.

This wording can also be interpreted in terms of disinformation, as much misleading content aims precisely to manipulate the most vulnerable sections of the population, compromising their ability to make independent decisions.

5.2. Disinformation in systems posing a high risk to electoral influence

Article 6(2) of the AI Act and Annex III identify specific AI systems as “high risk”. Providers of such systems are subject to strict obligations, outlined in Articles 9 to 15, which aim to ensure compliance, including requirements for transparency, traceability, data management and human oversight.

In particular, Annex III, which classifies specific cases of AI system use as high risk, includes systems used in the administration of justice and democratic processes (point 8). These include systems intended to influence the outcome of elections or referendums, or to influence the voting behaviour of voters, as described in Annex III, 8(b):

AI systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda. This does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view.

The inclusion of this use case is understandable and in line with the aims of the AI Act, which include protecting democracy and the rule of law from the harmful effects of AI systems.

In fact, Recital 62 reiterates the need to classify such systems as high risk, «in order to address the risks of undue external interference with the right to vote enshrined in Article 39 of the Charter, and of adverse effects on democracy and the rule of law», while Recitals 120 and 136 link the information obligations imposed on suppliers and deployers of AI systems to the need to mitigate the so-called “systemic” risk of negative impacts on electoral processes, including through disinformation (see section 5.5 below).

It is conceivable that classifying these systems as high risk will have significant implications for suppliers’ activities in combating disinformation. Firstly, providers are required to conduct a conformity assessment on systems before placing them on the market and to ensure that they meet high standards of accuracy, robustness and security, as well as to implement effective human oversight mechanisms that allow for prompt intervention in the event of misuse, undesirable effects or the generation of manipulative content.

Deployers – such as political parties, communication agencies or other actors using such systems in election campaigns or political communication activities – will also be subject to specific obligations. In particular, they will have to conduct a fundamental rights impact assessment, pursuant to Article 29 of the AI Act, to verify that the use of the system does not pose disproportionate risks to freedom of expression, privacy or the right to free and pluralistic information.

5.3. Transparency and disinformation obligations

Outside the cases mentioned above, the issue of disinformation is regulated in the AI Act by imposing a series of transparency obligations on providers and deployers, which remedy the information asymmetry between these actors and users, as indicated in Recital 132 («Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception irrespective of whether they qualify as high-risk or not. In certain circumstances, the use of these systems should therefore be subject to specific transparency obligations»).

The rules specifying these obligations are contained in Article 50 («Transparency obligations for providers and deployers of certain AI systems»).

The first paragraph requires suppliers to design and develop AI systems intended to interact directly with natural persons in such a way that those persons are informed that they are interacting with an AI system, unless this is obvious to an informed, attentive and circumspect person, taking into account the circumstances and context of use. This rule does not apply to AI systems authorised by law for crime-fighting activities.

The second paragraph requires suppliers of AI systems, including general-purpose AI systems, that generate synthetic audio, image, video or text content to ensure that the outputs of the AI system are marked in a machine-readable format and detectable as being artificially generated or manipulated.

The third paragraph requires deployers of AI systems for emotion recognition or biometric categorisation to inform individuals exposed to them about how the system works.

The fourth paragraph requires deployers of AI systems for generating or manipulating deepfake content (audio and video, but not text) to disclose that such content has been artificially produced. However, there are some exceptions, such as when such systems are used for legitimate purposes, including criminal investigations, or in artistic, satirical or fictional contexts. In such cases, the transparency obligation is limited to adequate labelling that

does not compromise the integrity or enjoyment of the work. This regulatory balance aims to protect freedom of expression and creativity by preventing the regulation of digital content from having a deterrent effect on cultural activities. Finally, deployers of an AI system that generates or manipulates published text for the purpose of informing the public on matters of public interest are required to disclose that the text has been artificially generated or manipulated.

A key aspect of the information requirements is the focus on the problem of deepfakes. Article 3(60) of the Regulation defines deepfake as «AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful». This definition, which plays a central role in the regulation of AI systems used for disinformation purposes, is based on a series of elements or conditions whose interpretation raises some difficulties:

1. the artificial or manipulated nature of the content;
2. the similarity to existing elements;
3. the potential to deceive human perception, leading people to believe that the content is authentic⁵⁴.

With regard to the artificial or manipulated nature of the content, a difficult issue to resolve is the threshold of manipulation required for such content to be considered a deepfake: for example, in the case of images, given that an image taken by a digital camera never shows (true) reality, but is the result of the chain of algorithmic image processing and the settings chosen by the photographer, the question is whether the application of any AI-based digital filter (widely available on common smartphones) to the image would be sufficient for it to be considered manipulated, and therefore a deepfake, if the other conditions were also met.

The second controversial element is the similarity to existing people, objects, places, entities or events. The question is whether, in order to classify content as a deepfake, it is necessary to ascertain the similarity of that content to a specific element that actually exists. For example, in order to classify a synthetically generated face as a deepfake, is it necessary to find a similarity with the face of a specific person who actually exists?

With regard to the third requirement, namely the potential to deceive

54. K. Meding, C. Sorge, *What Constitutes a Deep Fake? The Blurry Line Between Legitimate Processing and Manipulation Under the EU AI Act*, 2025, in *Proceedings of the 2025 Symposium on Computer Science and Law*, ACM, 2025.

human perception, the wording would ultimately seem to exclude situations in which audio or video content appears falsely authentic only to a computer system (e.g., following an adversarial attack).

In addition to these interpretative issues, there is a higher-level problem linked to the definition of deployer contained in Article 3(4), which excludes individuals who use the AI system in the course of a personal, non-professional activity.

This excludes all non-professional actors from the information obligations under Article 50, in particular those in paragraph 4. For example, all non-professional users who flood online platforms on a daily basis with content that qualifies as deepfake according to the definition contained in the AI Act would not be subject to the information obligations.

5.4. General-purpose AI models: disinformation as a systemic risk

The Regulation contains a separate chapter (Chapter V) dealing with “general-purpose AI models”⁵⁵. These are defined in Article 4 as AI models that exhibit significant generality and are capable of competently performing a wide range of distinct tasks, regardless of how the model is placed on the market, and which can be integrated into a variety of downstream systems or applications. This definition excludes AI models used for research and development of prototypes before being placed on the market⁵⁶.

Chapter V provides for two legal regimes for general-purpose AI models: one for all models, the other specific only to models that present a “systemic risk”.

According to Article 51, a general-purpose AI model presents a systemic risk when it meets any of the following requirements:

55. As is well known, the legal regime governing such systems was extensively debated during the final stages of the Regulation’s approval, proving to be one of the most controversial points in the final text. See, for example, L. Bertuzzi, *EU’s AI Act Negotiations Hit the Brakes over Foundation Models*, Euractiv, 2023, www.euractiv.com/section/tech/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/; L. Bertuzzi, *Commission Discloses Disagreements Between General-Purpose AI Providers and Other Stakeholders*, Euractiv, 2024, www.euractiv.com/section/tech/news/commission-discloses-disagreements-between-general-purpose-ai-providers-and-other-stakeholders/.

56. Different from *the* general-purpose AI *model* is the general-purpose AI *system*. A general-purpose AI system exists when a general-purpose AI model is integrated into or is a component of an AI system. The latter system thus differs from other specific-purpose systems covered by the Regulation (such as those in Annex III) in that it has the capacity to serve a variety of purposes. It can be used directly or integrated into other AI systems, including specific-purpose ones.

1. it has “high impact capacity”, assessed on the basis of appropriate technical tools and methodologies⁵⁷;
2. based on a decision by the Commission, either *ex officio* or on the request of the Scientific Panel, it has equivalent capability or impact, taking into account the criteria set out in Annex XIII.

The quantitative thresholds for systemic risk are accompanied by a “qualitative explanation” of what is meant by “systemic risk” in Recital 110. This concerns any *actual* or reasonably foreseeable negative effects in relation to serious incidents, disruptions to critical sectors and serious consequences for public health and safety, democratic processes, public and economic security, as well as the dissemination of illegal, false or discriminatory content. Attention should also be paid to the risks arising from potential intentional misuse or unintended problems, including the “facilitation of disinformation”.

It can therefore be said that disinformation represents, as in the Dsa, a systemic risk introduced by general-purpose AI models, as it has the potential to compromise the quality of public information, influence democratic processes and fuel social polarisation. The generalist nature and scalability of such models amplify the risk that automatically generated content – such as deepfakes, falsified or manipulative texts, and realistic but misleading images – will be disseminated on a massive scale and at such speed that traditional containment and verification measures become ineffective.

This ability to multiply and automate the production of disinformation raises crucial governance issues. Although the Regulation does not directly intervene on the content generated, it imposes obligations on providers of high-impact models to assess, monitor and, where necessary, mitigate systemic risks. These obligations include: the preparation of adequate technical documentation, transparent communication with downstream providers, the adoption of measures to prevent misuse and, in the most critical cases, mandatory notification to the Commission.

57. In particular, this requirement refers to the concept of “high impact capability”, which Article 4(64) defines as «capabilities that match or exceed those recorded in the most advanced general-purpose AI models». Following the most common *benchmarks* at the time of the Regulation’s approval, the legislator referred to Flops (acronym for Floating point Operations Per Second), i.e. the number of floating-point operations performed in one second by a Cpu, as a possible indicator of “high impact capability”. In particular, it has included a presumption that a general AI model has high impact capabilities if it exceeds the measure of 10^{25} . However, it is not excluded that alternative benchmarks may be used in this assessment.

6. Conclusions

The Digital Services Act and the AI Act are two pillars of the European strategy for regulating the digital space and, although they were created with distinct objectives and scopes, both touch on, albeit from different perspectives, the phenomenon we have defined as “augmented disinformation”.

Both regulations share certain general objectives: the protection of fundamental rights (in particular, freedom of expression and the right to information), the promotion of transparency, and the responsible management of risks associated with the use of digital technologies. However, neither addresses the issue of disinformation in a systematic manner.

In the Dsa, disinformation is potentially absorbed, from a material perspective, into the broader concept of “illegal content” or “content incompatible with the terms of service”. From a structural perspective, its various facets can be traced back to the concept of “systemic risk”, deriving from the structure and functioning of the systems used by platforms, which platforms must manage.

In the AI Act, on the other hand, disinformation is largely considered to be an effect attributable to artificial intelligence systems considered in terms of various levels and dimensions of risk: systems capable of manipulating people and causing harm (unacceptable risk); systems capable of influencing the outcome of democratic elections (high risk); systems capable of generating false content (limited risk); general-purpose AI systems with high impact capabilities. However, “disinformation” is never defined or treated as a separate, cross-cutting category.

On the other hand, although the two regulations operate in a shared digital ecosystem, they develop along regulatory axes that are not always compatible, with the risk of producing overlaps, gaps or tensions in their application⁵⁸.

A first aspect concerns the fact that the rules relevant to the fight against disinformation contained in the AI Act apply only to “AI systems” as defined in the regulation itself. As is well known, the definition of AI includes automated systems designed to operate with varying levels of autonomy and which may be adaptable after deployment and which, for explicit or implicit purposes, deduce from the input they receive how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments⁵⁹.

The definition chosen by the European legislator, with the aim of covering all

58. See J. Calvet-Bademunt, J. Barata, *The Digital Services Act Meets the AI Act: Bridging Platform and AI Governance*, Tech Policy Press, May 2024.

59. The AI Act has thus aligned with the revised definition of the Oecd, which was approved

possible current and future applications of AI systems, is broad and inclusive, but at the same time contains numerous ambiguities and does not provide clear guidance on how to distinguish artificial intelligence systems from other IT systems. Some interpretative issues have been partially resolved by the “Guidelines on the definition of an AI system” published by the European Commission last February⁶⁰.

However, this approach may not be suitable for reflecting the peculiarities of different types of artificial intelligence, as it applies the same regulatory framework to all systems, including deterministic ones (i.e. those that always produce the same output given a certain input)⁶¹. The approach adopted seems to be modelled primarily on machine learning-based systems (i.e. those that produce different outputs depending on the training data provided as input). Such uniform treatment does not seem consistent with risk-based regulation and the principle of technological neutrality: deterministic AI, which is predictable by its very nature, does not pose the same levels of risk as unpredictable machine learning-based AI.

On the other hand, the Dsa deals with all content moderation systems, regardless of whether they are based on AI as defined by the AI Act or on simpler algorithmic techniques (e.g. semantic filters, heuristic rules). In principle, the Regulation focuses on platforms and their procedural obligations in terms of transparency, accountability and user protection.

In reality, as we have seen, when it comes to disinformation, the obligation on Vlop and Vlose to identify and mitigate systemic risks largely concerns the functioning of the IT systems that determine how users access the service. There is talk of “recommendation systems”, “moderation systems” and “advertisement selection and presentation systems” as examples of the more general category of “algorithmic systems”; but there is no mention of the term “artificial intelligence”, only “algorithmic systems” or “automated decision-making processes”.

This choice is understandable and reflects a functionalist approach, focused on the effects and risks of the systems as such, regardless of their technical classification. The emphasis is on how content is processed, the impact of algorithmic tools, the information experience of users and the balance between freedom of expression and platform responsibility.

by the members of that organisation in November 2023. Recital 12 of the AI Act provides some additional elements for interpreting the rule.

60. European Commission, *Guidelines on the definition of an artificial intelligence system established by AI Act*, English version of 6 February 2025, C(2025) 924 final.

61. Cf. with the brief joint note by G. Contissa, F. Galli, *AI Act and Fundamental Rights: Technological Assumptions and Regulatory Implications*, in *Quaderni Costituzionali*, n. 3, 2024, pp. 738-740.

However, this means that there may be cases where algorithmic tools used by platforms to classify or moderate disinformation content fall within the scope of the Dsa but not within that of the AI Act.

Conversely, some systems that can be classified as AI under the AI Act but are not used by online platforms may not fall within the control mechanisms provided for by the Dsa. For example, questions have been raised about the extent to which platforms that provide chatbots powered by general-purpose AI models, such as ChatGpt, are regulated by the Dsa⁶².

Article 3 of the Dsa defines the concepts of intermediation service, online platform and online search engine, outlining the operational scope of the legislation. In particular, an intermediation service means any information society service that falls within the categories of mere conduit, caching or hosting. An online platform, on the other hand, is defined as a hosting service that, at the request of the user, stores and disseminates information to the public, unless this activity is marginal or ancillary to other services. Finally, an online search engine is a service that allows large-scale searches of websites through text, voice or other inputs, returning results in various formats.

None of these categories seems entirely satisfactory for covering chatbot applications based on generative AI.

It does not seem correct to classify such systems as online platforms, as there is no act of dissemination to the public at the request of the service recipient. Interactions with a chatbot are typically individual and private: the output generated is not automatically published or shared with other users, nor is it stored and made publicly accessible by the platform.

Similarly, it is difficult to argue that a chatbot falls within the definition of an online search engine, which presupposes the possibility for the user to “search” websites. Although some chatbots now integrate information retrieval or web browsing functions, the basic architecture is not designed to return results in the form of structured hyperlinks, but rather to generate autonomous content from textual input.

A possible, albeit partial, regulatory classification could be found within the concept of hosting, defined in Article 3(g) as an information society service «consisting of the storage of information provided by, and at the request of, a recipient of the service». In this case, the storage of user prompts or interactions by the chatbot provider could theoretically constitute a hosting activity. However, here too, the classification is uncertain: if such data is not stored permanently but only temporarily and for functional purposes (to

62. B. Botero Arcila, *Is It a Platform? Is It a Search Engine? It's Chatgpt! The European Liability Regime for Large Language Models*, in *J. Free Speech L.*, n. 3, 2023, p. 455.

generate output), it is doubtful whether it can properly be considered hosting for the purposes of the Dsa.

Moving on to the concept that is perhaps most relevant to the issue of disinformation, a further area of potential friction lies in the different interpretations of the notion of “systemic risk”. We have seen that both regulations refer to this concept, but from partially divergent perspectives.

In the Dsa, systemic risk is linked to the potential impact of Vlop and Vlose on fundamental rights, public health, safety and the democratic process, including the spread of disinformation. The AI Act, on the other hand, introduces a similar notion but refers only to general-purpose AI systems and is linked to an idea of risk that essentially derives from the “high-impact capabilities” of such systems.

This “potential misalignment” seems to be recognised by the AI Act itself. Recitals 120 and 136 of the AI Act specify that compliance with the obligations imposed by the Regulation on suppliers and deployers of certain AI systems, which are designed to enable the detection and disclosure of the fact that the outputs of such systems are generated or manipulated artificially, are important in contributing to the effective implementation of Regulation (EU) 2022/2065. It is unclear whether this provision – which, we repeat, is included in a recital – should be interpreted to mean that compliance with the obligations of the AI Act can be presumed to satisfy the obligations of the Dsa, or vice versa. Furthermore, it is unclear which obligations are being referred to, i.e. whether such coordination concerns all relevant obligations relating to disinformation – as discussed in the previous section – or only those for providers of general-purpose AI models in relation to the obligation to identify and mitigate systemic risks.

In short, even in this case, the lack of explicit regulatory coordination between the AI Act and the Dsa could therefore result in fragmentation of responsibility, especially in cases where a generative AI system operates through a publicly accessible online interface. Who is responsible, for example, for the unintentional dissemination of disinformation generated by such models? The model provider, the platform operator, or the user?

While the AI Act and the Dsa share convergent regulatory objectives, their implementation requires more robust interpretative and operational coordination. Without explicit alignment between the two regulations, there is a risk of overlap, regulatory gaps and legal uncertainty for the actors involved. The growing integration of AI into online information dynamics therefore requires regulatory harmonisation that recognises the hybrid and systemic nature of these technologies.

FAKE NEWS E INTELLIGENZA ARTIFICIALE: SFIDE ETICHE E SOLUZIONI TECNOLOGICHE

*Andrea Ongarini**, *Federico Cerutti***, *Andrea Loreggia****

SOMMARIO: 1. Introduzione – 2. Related works – 3. Sfide etiche – 4. Valutazione sperimentale del framework Dkfind – 5. Setup sperimentale – 6. Conclusioni.

1. Introduzione

La crescente diffusione di contenuti falsi e fuorvianti, comunemente noti come fake news, rappresenta una delle sfide più urgenti e complesse dell'ecosistema informativo contemporaneo. Questo fenomeno, descritto da Wardle e Derakhshan come *information disorder*¹, si manifesta in forme sempre più sofisticate, rendendo difficile per i cittadini distinguere tra informazione autentica e manipolata. Sebbene la disinformazione non sia un fenomeno nuovo, la sua portata e velocità sono state radicalmente amplificate dall'avvento delle tecnologie digitali e, in particolare, dall'intelligenza artificiale (IA).

L'IA generativa ha introdotto un cambiamento paradigmatico nella produzione e diffusione dei contenuti. Sistemi avanzati come i Large Language Models (Llm) permettono oggi di generare testi, immagini e video falsi con un livello di realismo tale da renderli spesso indistinguibili dai contenuti autentici. Questo ha creato una tensione epistemologica: la stessa tecnologia che potrebbe essere impiegata per contrastare la disinformazione è anche uno dei suoi prin-

* Collaboratore di ricerca in Ingegneria industriale e dell'Informazione, Università degli Studi di Brescia.

** Professore ordinario di Ingegneria industriale e dell'Informazione, Università degli Studi di Brescia.

*** Professore associato di Ingegneria industriale e dell'informazione, Università degli Studi di Brescia.

1. C. Wardle, H. Derakhshan, *Information disorder: Toward an interdisciplinary framework for research and policymaking*, vol. 27, 2017.

cipali vettori. I metodi tradizionali di verifica, basati sull'intervento umano, risultano inadeguati di fronte alla scala e alla rapidità con cui i contenuti vengono generati e diffusi. Di conseguenza, emerge la necessità di sviluppare strumenti automatizzati e intelligenti per affrontare il problema in modo sistemico.

La tassonomia proposta da Wardle e Derakhshan distingue tre forme principali di distorsione informativa: la disinformazione, ovvero la diffusione di informazioni false senza intento ingannevole; la disinformazione, che implica la creazione e diffusione intenzionale di contenuti falsi per ingannare o danneggiare; e la malinformazione, che consiste nella condivisione di informazioni veritiere con l'intento di arrecare danno. Queste dinamiche si propagano attraverso una varietà di canali, dai social media alle app di messaggistica, fino ai media tradizionali, coinvolgendo attori eterogenei: governi, gruppi politici, individui, e sistemi automatizzati come bot e troll farm. Le conseguenze sono profonde e strutturali: polarizzazione sociale, erosione della fiducia nelle istituzioni, e indebolimento del dibattito pubblico.

Oltre alla creazione di contenuti falsi, l'IA contribuisce alla loro amplificazione attraverso algoritmi di raccomandazione che privilegiano contenuti emotivamente coinvolgenti e polarizzanti. Questo meccanismo favorisce la viralità delle fake news rispetto alle notizie verificate, contribuendo alla formazione di bolle informative e alla radicalizzazione delle opinioni. Studi recenti evidenziano come l'esposizione ripetuta alla disinformazione possa alterare significativamente le credenze e i comportamenti degli individui.

In questo contesto, diventa cruciale adottare un approccio multidisciplinare che integri competenze in IA, scienze cognitive, sociologia e comunicazione, al fine di progettare soluzioni efficaci per mitigare l'impatto della disinformazione e promuovere un ecosistema informativo più trasparente, resiliente e affidabile.

2. Related works

La lotta alla disinformazione ha generato un vasto corpus di ricerche interdisciplinari, che spaziano dall'informatica e dall'intelligenza artificiale fino alla psicologia cognitiva, alla comunicazione e alle scienze politiche. In questa sezione si propone una rassegna ragionata della letteratura scientifica esistente, con particolare attenzione ai sistemi di fact-checking automatici, alle metodologie di rilevamento della disinformazione e all'impiego di soluzioni basate su IA per contrastare la diffusione di notizie false.

I primi approcci al fact-checking si sono basati su metodologie manuali, sviluppate da organizzazioni come PolitiFact, FactCheck.org, GossipCop e Sno-

pes.com, che hanno definito protocolli rigorosi per la verifica delle affermazioni. Tali protocolli includono l'identificazione di affermazioni verificabili, la consultazione di fonti affidabili e di esperti di dominio, la sintesi delle evidenze e la pubblicazione di analisi dettagliate. Sebbene questi metodi offrano elevata accuratezza, la loro scalabilità è limitata. Pennycook e Rand² hanno evidenziato come la lentezza del fact-checking manuale riduca l'efficacia della correzione, che spesso non raggiunge la totalità del pubblico esposto alla notizia falsa. L'International Fact-Checking Network (Ifcn) ha definito uno standard globale basato su cinque principi fondamentali: imparzialità, trasparenza delle fonti e dei finanziamenti, chiarezza metodologica e impegno alla correzione degli errori.

Per superare i limiti del fact-checking manuale, la ricerca ha esplorato soluzioni automatizzate basate su tecniche di apprendimento automatico. Shu *et al.*³ hanno proposto una tassonomia che distingue tra approcci basati sul contenuto e quelli basati sul contesto sociale. I metodi content-based analizzano le caratteristiche linguistiche, stilistiche e visive delle notizie. Perez-Rosas *et al.*⁴ hanno sviluppato classificatori linguistici che sfruttano n-grams, pattern di punteggiatura, metriche di leggibilità e caratteristiche psicolinguistiche, dimostrando che le fake news tendono a utilizzare un linguaggio più soggettivo ed emotivo. Horne e Adali⁵ hanno osservato che i contenuti falsi sono spesso più sintetici, con strutture ripetitive e un vocabolario semplificato.

L'avvento del deep learning ha segnato una svolta nell'individuazione automatica delle fake news. Le Convolutional Neural Networks (Cnn) e le Recurrent Neural Networks (Rnn) sono state impiegate per rilevare pattern testuali e multimediali. Wang⁶ ha introdotto il dataset Liar⁷, dimostrando che modelli

2. G. Pennycook, D.G. Rand, *Fighting misinformation on social media using crowdsourced judgments of news source quality*, in *Proceedings of the National Academy of Sciences*, 2019, vol. 116, n. 7, pp. 2521-2526.

3. K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, *Fake news detection on social media: A data mining perspective*, in *ACM SIGKDD Explorations Newsletter*, 2017, vol. 19, n. 1, pp. 22-36.

4. V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, *Automatic detection of fake news*, in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3391-3401, Agosto 2018. Association for Computational Linguistics.

5. B. Horne, S. Adali, *This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news*, in *Proceedings of the International AAAI Conference on Web and Social Media*, 2017, vol. 11, pp. 759-766.

6. W. Yang Wang, *"Liar, liar pants on fire": A new benchmark dataset for fake news detection*, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (vol. 2: Short Papers), pp. 422-426, 2017.

7. D. Quan Vietnamca, *Liar dataset*, 2020, www.kaggle.com/datasets/doanquanvietnamca/liar-dataset.

ibridi, basati su analisi del testo e metadati relativi all'autore dell'affermazione, possono migliorare le performance di classificazione. Le reti Lstm (Long Short-Term Memory) si sono rivelate particolarmente efficaci nell'analisi delle dipendenze temporali nella propagazione delle notizie. Ma *et al.* (2016) hanno proposto un modello Rnn capace di analizzare la sequenza temporale dei post sui social media, evidenziando come le notizie vere e false seguano pattern evolutivi differenti. Shu *et al.*⁸ hanno introdotto Defend, un framework “explainable” per il rilevamento delle fake news, che non solo classifica i contenuti ma fornisce anche spiegazioni interpretabili sulle decisioni prese, contribuendo alla trasparenza e alla responsabilità algoritmica. Con l'emergere delle architetture basate su transformers e dei Large Language Models (Llm), il panorama è stato ulteriormente rivoluzionato. Modelli come Bert e le sue varianti hanno ottenuto risultati significativi su benchmark di fact-checking e rilevamento della disinformazione. Khandelwal *et al.*⁹ hanno dimostrato che il fine-tuning di Bert su dataset specifici migliora sensibilmente le prestazioni rispetto agli approcci tradizionali. Tuttavia, gli stessi modelli che eccellono nel rilevamento possono anche essere utilizzati per generare contenuti falsi altamente convincenti. Gpt-3 e i suoi successori hanno mostrato una capacità notevole nel produrre testo coerente e contestualmente plausibile, rendendo difficile distinguere tra contenuti generati artificialmente e quelli autentici.

Questa ambivalenza ha portato alla nascita di una nuova linea di ricerca dedicata al rilevamento del testo generato da IA. Mitchell *et al.*¹⁰ hanno proposto DetectGpt, un metodo che sfrutta l'analisi della curvatura di probabilità per identificare contenuti generati da Llm. Gehrmann *et al.*¹¹ hanno sviluppato Gltr (Giant Language Model Test Room), uno strumento che visualizza pat-

8. K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, *Defend: Explainable fake news detection*, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

9. A. Khandelwal, *Fine-tune longformer for jointly predicting rumor stance and veracity*, in *Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IJDD Cods & 26th Comad)*, Cods-Comad '21, pp. 10-19, 2021. Association for Computing Machinery.

10. E. Mitchell, Y. Lee, A. Khazatsky, C.D. Manning, C. Finn, *DetectGPT: Zero-shot machine-generated text detection using probability curvature*, in A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, J. Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, vol. 202 of *Proceedings of Machine Learning Research*, pp. 24950-24962. PMLR, 23-29 Jul 2023.

11. S. Gehrmann, H. Strobel, A. Rush, *Gltr: Statistical detection and visualization of generated text*, in M.R. Costa-jussà, E. Alfonseca (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 111-116, 2019. Association for Computational Linguistics.

tern linguistici tipici del testo sintetico, facilitando l'identificazione da parte di esperti e utenti.

Un'evoluzione promettente è rappresentata dai sistemi Rag (Retrieval-Augmented Generation), che combinano le capacità generative dei Llm con il recupero di documenti rilevanti. Lewis *et al.*¹² hanno introdotto questa architettura, che consente di generare risposte più fondate sui fatti. In parallelo, il dataset Fever¹³ ha rappresentato un benchmark fondamentale per la verifica automatica delle affermazioni. Nie *et al.*¹⁴ hanno sviluppato Nsmn, un sistema a tre stadi che integra recupero, selezione e verifica, migliorando significativamente l'accuratezza.

La ricerca più recente si è concentrata sul ragionamento multi-hop e sull'aggregazione delle prove. Zhao *et al.*¹⁵ e Fan *et al.* (2020) hanno proposto modelli basati su grafi per catturare relazioni tra fonti multiple. Cui *et al.* (2020) hanno integrato grafi di conoscenza come Wikidata nei sistemi di fact-checking, migliorando la spiegabilità e la robustezza delle decisioni.

La crescente complessità dell'ecosistema della disinformazione ha portato allo sviluppo di approcci multimodali. Khattar *et al.*¹⁶ hanno proposto Mvae, un framework che analizza incongruenze tra testo e immagini. Li *et al.*¹⁷ hanno classificato i metodi di rilevamento dei deepfake in base ad artefatti di compressione, incongruenze fisiologiche e coerenza temporale.

L'integrazione efficace delle modalità è una sfida aperta. Strategie come Early Fusion, Late Fusion e Hybrid Fusion sono state esplorate per combinare testo, immagini e audio. Wang *et al.*¹⁸ hanno sviluppato un modello di atten-

12. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, D. Kiela, *Retrieval-augmented generation for knowledge-intensive Nlp tasks*, in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS 2020. Curran Associates Inc.

13. J. Thorne, A. Vlachos, O. Cocarascu, C. Christodoulopoulos, A. Mittal, *The fact extraction and verification (fever) shared task*, 2018.

14. Y. Nie, H. Chen, M. Bansal, *Combining fact extraction and verification with neural semantic matching networks*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, n. 1, pp. 6859-6866.

15. Y. Zhao, K. Yan, F. Huang, J. Li, *Graph-based high-order relation discovery for fine-grained recognition*, in *Proceedings of the IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr)*, pp. 15079-15088, June 2021.

16. D. Khattar, J. Singh Goud, M. Gupta, V. Varma, *Mvae: Multimodal variational autoencoder for fake news detection*, in *The World Wide Web Conference, WWW '19*, pp. 2915-2921, 2019. Association for Computing Machinery.

17. X. Li, Y. Lang, Y. Chen, X. Mao, Y. He, S. Wang, H. Xue, Q. Lu, *Sharp multiple instance learning for deepfake video detection*, in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1864-1872. Association for Computing Machinery.

18. Y. Guo, H. Ge, J. Li, *A two-branch multimodal fake news detection model based on multimodal bilinear pooling and attention mechanism*, in *Frontiers in Computer Science*, vol. 5, 2023.

tion multimodale che pondera dinamicamente l'importanza delle modalità. Dataset come Fakeddit (Nakamura *et al.*, 2020) e Twitter-MFN (Singhal *et al.*, 2019) hanno facilitato la ricerca, sebbene la disponibilità di dati video manipolati rimanga limitata.

Le architetture transformer multimodali, come VL-Bert e Lxmert, stanno mostrando risultati promettenti nel cogliere relazioni complesse tra testo e immagini. Tuttavia, la detection multimodale deve ancora affrontare sfide legate alla generalizzazione cross-piattaforma, alla scalabilità computazionale e all'interpretabilità dei modelli.

3. Sfide etiche

Lo sviluppo di sistemi automatizzati di fact-checking solleva importanti questioni etiche relative alla libertà di espressione, alla trasparenza algoritmica e alla responsabilità delle decisioni automatizzate. I sistemi di intelligenza artificiale devono essere progettati con meccanismi di spiegabilità che consentano agli utenti di comprendere le basi delle decisioni di classificazione delle notizie.

La potenziale censura involontaria di contenuti legittimi rappresenta un rischio significativo, così come il rafforzamento di bias esistenti attraverso gli algoritmi di classificazione. Per affrontare queste sfide, alcuni ricercatori hanno proposto framework etici per lo sviluppo e l'implementazione di sistemi di fact-checking automatizzati, sottolineando l'importanza della supervisione umana, della diversità nei dati di addestramento e della trasparenza nei processi decisionali.

Il rischio di censura involontaria è particolarmente rilevante: algoritmi di moderazione e filtraggio possono erroneamente escludere contenuti legittimi, compromettendo la libertà di espressione, l'accesso all'informazione e il dialogo democratico.

Secondo uno studio del Parlamento Europeo¹⁹, la regolazione del filtraggio automatizzato deve bilanciare due obiettivi spesso in tensione:

1. prevenire e mitigare i danni individuali e sociali causati da contenuti illeciti o inappropriati;
2. favorire la diffusione di contenuti benefici e promuovere interazioni libere e civili tra gli utenti.

19. G. Sartor, A. Loreggia, *The impact of algorithms for online content filtering or moderation. upload filters*, European Commission, 2020.

La moderazione algoritmica²⁰, se non correttamente progettata, può rafforzare bias preesistenti, penalizzando gruppi vulnerabili o opinioni minoritarie. Inoltre, la mancanza di trasparenza algoritmica e di meccanismi di spiegabilità rende difficile per gli utenti comprendere le ragioni delle decisioni prese dai sistemi di IA.

Per affrontare queste sfide, è necessario adottare framework etici che includano:

- supervisione umana nei processi decisionali automatizzati;
- diversità e rappresentatività nei dati di addestramento;
- meccanismi di spiegabilità e accountability;
- trasparenza nei criteri di classificazione e nelle fonti utilizzate;
- procedure di ricorso e revisione per contenuti rimossi o penalizzati ingiustamente.

Infine, una regolazione inadeguata può indurre filtraggi eccessivi o insufficienti, minando la fiducia nelle piattaforme e nella loro capacità di gestire responsabilmente l'informazione online. È quindi essenziale che le politiche pubbliche promuovano un approccio equilibrato, che tuteli i diritti fondamentali e garantisca un ecosistema informativo sano e pluralistico.

4. Valutazione sperimentale del framework Dkfn

La valutazione sperimentale di questo lavoro è concepita per offrire spunti utili in merito alle implicazioni etiche e regolatorie. L'obiettivo è fornire uno strumento pratico e interpretabile anche per stakeholder privi di competenze tecniche, ma coinvolti nella supervisione o nella regolazione dell'adozione di tali sistemi. Evidenziando il comportamento dei diversi modelli in termini di accuratezza, ed efficienza, la valutazione contribuisce a chiarire come queste tecnologie debbano essere utilizzate con attenzione e responsabilità, affinché la loro implementazione sia coerente con i valori democratici e con la tutela dei diritti fondamentali. I risultati presentati sono tuttavia ancora preliminari e tuttora in fase di approfondimento.

Questa sezione presenta la valutazione sperimentale preliminare condotta per analizzare le prestazioni del framework Dkfn (Detect, Know, Filter, Neutralize

20. F. Galli, A. Loreggia, G. Sartor, *The regulation of content moderation*, in *International conference on the legal challenges of the fourth industrial revolution*, Springer International Publishing, 2022, pp. 63-87.

Disinformation)²¹ nella rilevazione automatica di fake news, con particolare riferimento agli approcci few-shot e Retrieval-Augmented Generation (Rag).

L’obiettivo principale dell’esperimento è quello di esplorare l’efficacia del framework Dkfind nel classificare contenuti informativi in scenari caratterizzati da una disponibilità limitata di dati etichettati. In particolare, si intende valutare come le performance del sistema varino in funzione del modello linguistico e del motore di ricerca adottati, mettendo in luce il ruolo dell’architettura e della capacità di generalizzazione dei modelli in contesti reali e dinamici.

5. Setup sperimentale

Per questa analisi sono stati selezionati sette modelli linguistici di grandi dimensioni (Large Language Models), scelti per la loro eterogeneità in termini di architettura, dimensione parametrica e data di rilascio. La *Tabella 1* riassume le caratteristiche principali dei modelli impiegati.

Tab. 1 – Lista dei modelli utilizzati negli esperimenti.

Modello	Parametri (B)	Cutoff conoscenza	Data di rilascio
Amazon-Nova-Pro	-	-	-
Claude-3-Haiku	~20	luglio 2024	13 marzo 2024
Claude-3.5-Sonnet	~70	aprile 2024	22 ottobre 2024
Gpt-Oss-20b	20	1° giugno 2024	5 agosto 2025
Llama3-3-70b	70	dicembre 2023	18 dicembre 2024
Mistral-8x7b	12	< 27 settembre 2023	1° marzo 2024
Qwen3-32b (Dense)	32	< data di rilascio	18 settembre 2025

Sono stati impiegati due dataset di riferimento per la classificazione binaria delle notizie, PolitiFact e GossipCop, presentati in Shu *et al.*²² e spesso utilizzati in letteratura per la valutazione delle performance per sistemi di fake news detection: PolitiFact: 322 istanze per il training e 200 per il testing; GossipCop: sotto-campionato per includere 400 istanze di training e 100 di testing.

Questa fase sperimentale ha carattere esplorativo e preliminare, e mira a fornire indicazioni utili per l’ottimizzazione futura del framework Dkfind, in vi-

21. Y. Liu, J. Zhu, X. Liu, H. Tang, Y. Zhang, K. Zhang, X. Zhou, E. Chen, *Detect, investigate, judge and determine: A knowledge-guided framework for few-shot fake news detection*, arXiv preprint arXiv: 2407.08952, 2024, versione 5.

22. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, *Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media*, in *Big data*, 2020, vol. 8, n. 3, pp. 171-188.

sta di una sua applicazione su larga scala in ambienti informativi complessi e multilingue.

Per ciascun modello, le performance sono state valutate utilizzando tre motori di ricerca: Google, DuckDuckGo e Brave. Tutti gli esperimenti sono stati condotti con approccio few-shot utilizzando 8 esempi per classe (8 fake e 8 real). I risultati sono riportati in *Tabella 2* e *Tabella 3*.

Tab. 2 – Performance dei modelli utilizzati suddivise per dataset e motore di ricerca

Dataset	Modello	Google	DuckDuckGo	Brave
PolitiFact	Amazon-Nova-Pro	0.9050	0.9100	0.9100
	Llama3-3-70b	0.8800	0.8650	0.8750
	Mistral-8x7b	0.7950	0.8250	0.8300
	Claude-3-Haiku	0.8500	0.8450	0.8550
	Claude-3-5-Sonnet	0.9150	0.9100	0.9150
	Gpt-Oss-20b	0.6900	0.7150	0.7000
	Qwen3-32b	0.8800	0.8800	0.8800
GossipCop	Amazon-Nova-Pro	0.7000	0.7000	0.6600
	Llama3-3-70b	0.6200	0.6400	0.6300
	Mistral-8x7b	0.5700	0.5400	0.6000
	Claude-3-Haiku	0.6100	0.5800	0.6100
	Claude-3-5-Sonnet	0.6800	0.6900	0.6900
	Gpt-Oss-20b	0.4500	0.5100	0.5200
	Qwen3-32b	0.6600	0.6600	0.6600

Tab. 3 – Tempi di esecuzione dei modelli utilizzati suddivise per dataset e motore di ricerca

Dataset	Modello	Google (s)	DuckDuckGo (s)	Brave (s)
PolitiFact	Amazon-Nova-Pro	10884.16	11304.76	12384.76
	Llama3-3-70b	8993.03	12269.88	10056.51
	Mistral-8x7b	8378.85	11706.81	11750.18
	Claude-3-Haiku	9244.83	12198.58	12497.84
	Claude-3-5-Sonnet	8561.20	9049.28	8956.51
	Gpt-Oss-20b	20590.10	24059.88	21850.58
	Qwen3-32b	3884.86	5552.70	5091.45
GossipCop	Amazon-Nova-Pro	8720.68	12353.92	12343.93
	Llama3-3-70b	7399.55	9766.59	9843.04
	Mistral-8x7b	8518.80	10845.40	11623.77
	Claude-3-Haiku	7868.34	10039.19	9921.89
	Claude-3-5-Sonnet	12576.27	9603.43	6249.69
	Gpt-Oss-20b	14211.63	20511.45	21453.75
	Qwen3-32b	3935.60	5058.08	4489.21

L'analisi empirica condotta sul framework Dkfn ha evidenziato una serie di elementi chiave che meritano una riflessione approfondita in relazione agli obiettivi etici e normativi del progetto. In particolare, il confronto tra modelli linguistici e motori di ricerca ha messo in luce il delicato equilibrio tra accuratezza, efficienza computazionale e affidabilità del sistema, aprendo a considerazioni più ampie sul ruolo dell'intelligenza artificiale nella gestione della disinformazione.

- Bilanciamento tra accuratezza e sostenibilità operativa: I risultati mostrano che alcuni modelli, come Qwen3-32B e Claude-3.5-Sonnet, riescono a coniugare buone performance in termini di accuratezza con tempi di esecuzione contenuti. Questo aspetto è particolarmente rilevante in ottica di scalabilità e applicabilità in contesti reali, dove le risorse computazionali possono essere limitate.
- Trasparenza e spiegabilità come requisiti fondamentali: La variabilità nei risultati tra modelli e motori di ricerca sottolinea l'importanza di garantire trasparenza nei processi decisionali automatizzati. In un contesto in cui le decisioni algoritmiche possono influenzare l'accesso all'informazione e la percezione pubblica, è essenziale che gli utenti possano comprendere le motivazioni alla base della classificazione di una notizia. Meccanismi di spiegabilità e accountability devono essere integrati nel sistema, non solo per favorire la fiducia, ma anche per rispondere alle esigenze normative in materia di responsabilità algoritmica.
- Bias e rischio di censura automatizzata: Un ulteriore elemento critico riguarda il rischio di censura involontaria e di rafforzamento di bias preesistenti. I modelli linguistici, se non adeguatamente controllati, possono penalizzare contenuti minoritari o culturalmente specifici, compromettendo la pluralità informativa. La selezione del motore di ricerca e la composizione del dataset di addestramento giocano un ruolo determinante in questo processo. È quindi necessario prevedere forme di supervisione umana e audit periodici per garantire equità e inclusività.
- Governance algoritmica e prospettive regolatorie: Infine, i risultati offrono spunti concreti per la definizione di politiche pubbliche e linee guida regolatorie. Come evidenziato dal Parlamento Europeo, la moderazione automatizzata deve bilanciare la protezione da contenuti dannosi con la promozione di un dibattito libero e pluralistico.

6. Conclusioni

La crescente sofisticazione delle fake news, alimentata dall'evoluzione dei modelli generativi e dalla pervasività delle tecnologie digitali, impone lo sviluppo di strumenti automatizzati, scalabili e affidabili per la loro rilevazione. In questo lavoro, il framework DkfnD ha dimostrato (in forma preliminare) di essere una soluzione promettente, capace di integrare modelli linguistici avanzati con fonti esterne di conoscenza, come i motori di ricerca, per arricchire l'input informativo e migliorare la classificazione.

Attraverso una valutazione empirica condotta su due dataset reali (PolitiFact e GossipCop), e utilizzando sette modelli linguistici in combinazione con tre motori di ricerca (Google, DuckDuckGo, Brave), è stato possibile analizzare il compromesso tra accuratezza e efficienza computazionale. I risultati hanno evidenziato differenze significative tra le architetture dei modelli e tra le fonti informative utilizzate, con alcune configurazioni (es. Qwen3-32B e Claude-3.5-Sonnet) che si sono distinte per il buon bilanciamento tra prestazioni e tempi di esecuzione.

Tuttavia, per garantire la solidità e la generalizzabilità dei risultati, è fondamentale estendere l'analisi a una più ampia varietà di framework e contesti applicativi. L'inclusione di ulteriori modelli, dataset e fonti di conoscenza permetterebbe di validare le osservazioni emerse, ridurre il rischio di overfitting metodologico e rafforzare la robustezza delle soluzioni proposte.

In prospettiva, l'analisi dovrebbe orientarsi verso l'integrazione di scenari multimodali (testo, immagini, audio), l'utilizzo di knowledge graphs per migliorare la spiegabilità e la coerenza semantica e l'ottimizzazione per ambienti a risorse limitate, al fine di garantire accessibilità e sostenibilità. Questi sviluppi sono essenziali per costruire sistemi di rilevazione della disinformazione che siano non solo tecnicamente efficaci, ma anche eticamente responsabili e conformi alle normative emergenti in materia di trasparenza, equità e tutela dei diritti fondamentali.

MISURE EUROPEE E NAZIONALI CONTRO LA DISINFORMAZIONE STRATEGICA ONLINE. CONTRIBUTO A PARTIRE DAL CASO *STORM-1516*

*Domenico Bruno**

SOMMARIO: 1. Articolazione e obiettivi della ricerca – 2. Misure europee: Digital service act e AI Act – 3. Misure nazionali: quadro normativo incerto e ruolo delle Autorità indipendenti – 4. Caso studio: l'operazione *Storm-1516* – 5. Conclusioni: rendimento delle misure e ipotesi per una regolazione euro-nazionale.

1. Articolazione e obiettivi della ricerca

Il presente contributo analizza le misure europee e nazionali per il contrasto alla disinformazione strategica *online*. In particolare: la ricerca intende considerare, alla luce del caso *Storm-1516*¹, il rendimento più recente di tali misure e confrontarsi con il tentativo di elaborare, anche in prospettiva multilivello, ulteriori strategie per prevenire le minacce alla sicurezza nazionale recate dalle ingerenze straniere nel dibattito pubblico interno².

L'oggetto indagato, a ben vedere, è soltanto in parte nuovo: è risalente l'impiego, soprattutto da parte dei servizi di *intelligence*³, della disinformazione strategica per condizionare l'opinione pubblica di uno Stato nemico e indurlo ad assumere determinate decisioni⁴. Tuttavia, è nell'esperienza recente,

* Assegnista di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Napoli Federico II.

1. Si tratta di una complessa campagna di disinformazione strategica messa in campo da *hackers* collegati all'*intelligence* russa, realizzata con il principale obiettivo di screditare il Governo ucraino agli occhi dell'opinione pubblica occidentale e di rafforzare movimenti, associazioni e partiti filorussi in Europa. Al riguardo, per approfondire, si v. *infra*, § 3.

2. Sul punto si v. C. Pinelli, *Disinformazione, comunità virtuali e democrazia: un inquadramento costituzionale*, in *Diritto pubblico*, n. 1, 2022, pp. 173 ss.

3. Sul punto, per un'analisi più ampia, si v. G. Salvi, *Intelligence e potere*, in *Enc. Dir. Tematici V*, Giuffrè, 2023, pp. 247 ss.

4. Si pensi, solo per citare un caso emblematico, alle operazioni di disinformazione promosse

in cui i *social network* sono tra i principali canali di comunicazione⁵, che tali operazioni hanno conosciuto sviluppi significativi⁶. Il *web*, infatti, consente, attraverso i sistemi di profilazione e *microtargeting*, di diffondere notizie su larga scala con enorme rapidità e di influenzare milioni di persone in brevissimo tempo⁷. Un rischio, questo, ancor più amplificato dai *software* di intelligenza artificiale generativa che permettono, attraverso le tecniche di *deepfake*, di creare in pochi minuti contenuti multimediali di qualsiasi tipo⁸: senza particolari competenze, ad esempio, per indebolire la credibilità di un esponente politico, è possibile realizzare un video in cui pronunci dichiarazioni mai effettivamente rese.

Tali circostanze, soprattutto in contesti segnati da bassi livelli di alfabetizzazione mediatica o digitale come quello italiano⁹, possono condizionare il dibattito pubblico, palesando seri rischi per l'integrità dell'ordinamento democratico¹⁰.

se dal Ministro della propaganda Joseph Goebbels del regime autoritario Adolf Hitler. Sulle implicazioni della propaganda nel circuito del rapporto tra autorità e libertà si v. F. Lanchester, *Propaganda elettorale*, in *Enc. Dir.*, Giuffrè, vol. XXXVII, 1988, pp. 126 ss.

5. Riflette su tali profili C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, n. 3, 2023, pp. 543 ss. Ulteriori approfondimenti possono essere tratti da G.E. Vigevani, *Potere politico e mezzi di comunicazione*, in *Rivista AIC*, n. 2, 2025, p. 236.

6. Meria considerare in data 17 novembre 2025 il Presidente della Repubblica ha convocato il Consiglio Supremo di Difesa proprio per affrontare «le minacce ibride con riferimento anche alla dimensione cognitiva e alle possibili ripercussioni sulla sicurezza dell'Unione europea e dell'Italia». Per una analisi sugli esiti del Consiglio si v. il comunicato stampa rilasciato al termine della riunione, disponibile su: www.quirinale.it/elementi/145031.

7. Per ulteriori approfondimenti si v. A. Sterpa, *La difesa delle democrazie dalle minacce ibride. Teoria della comunità, dell'autorità, del potere e migrazioni: elementi per un'analisi strategica della "guerra tipeida"*, Editoriale Scientifica, 2025, spec. pp. 41 ss.; R. Bifulco, *Riverberi costituzionali del Metaverso*, in *MediaLaws*, n. 3, 2023, pp. 41 ss.

8. Peraltro, la maggior parte dei *social media* induce gli utenti a interagire in ambienti digitali omogenei (*echo chambers*) in cui non è presente alcuna forma di contraddittorio, ma vengono costantemente riproposti contenuti simili, riducendo sensibilmente le possibilità di confronto critico (cfr. M. Mezzanotte, *Fake news, deepfake e sovranità digitale nei periodi bellici*, in *federalismi.it*, n. 33, 2022, p. 45).

9. Si v. in tal senso il rapporto Autorità per le Garanzie nelle Comunicazioni, *I fabbisogni di alfabetizzazione mediatica e digitale*, Roma, 3 luglio 2025. Sul rapporto tra digitalizzazione e cittadinanza, si v. F. Pizzolato, *Digitalizzazione e cittadinanza: la figura dell'e-citizen*, in *Istituzioni del federalismo*, n. 3, 2024, p. 627. Da ultimo, si v. altresì G. Vasino, *Lotta alla disinformazione, garanzie costituzionali e tutela dei diritti fondamentali nel panorama normativo dell'Unione europea*, in *federalismi.it*, n. 15, 2025, pp. 97 ss.

10. L'episodio più eclatante in tal senso è rappresentato dall'ultima campagna elettorale per le elezioni presidenziali in Romania. In ragione delle interferenze russe nel dibattito

La sicurezza nazionale¹¹, dunque, non dipende più solo dalla capacità dello Stato di presidiare i propri confini¹², ma anche dalla possibilità di proteggersi da minacce ibride¹³: minacce, queste, che contemplan la combinazione di modalità tradizionali (come attacchi militari o economici¹⁴) e non convenzionali (come attacchi informatici, campagne di disinformazione *online*), e che hanno

pubblico la Corte costituzionale, la sentenza n. 32 del 2024, ha annullato l'esito della procedura elettorale. Sul punto, per un'analisi più ampia, si v. A. Lucarelli, *Nuovi mezzi di comunicazione, assetti imprenditoriali e soggettività politiche*, in *Rivista AIC*, n. 3, 2025, pp. 133 ss.; F. Rosa, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025, pp. 53 ss.; S. Sassi, A. Sterpa, *La Corte costituzionale della Romania difende la democrazia liberale dalla disinformazione. Prime note sulla sentenza n. 32 del 6 dicembre 2024*, in *federalismi.it*, n. 4, 2025, pp. 162 ss. Sul pericolo recato da tali minacce si v. anche la relazione annuale al Parlamento del Sistema di Informazione per la sicurezza della Repubblica della Presidenza del Consiglio dei Ministri: Presidenza del Consiglio dei Ministri, Sistema di Informazione per la Sicurezza della Repubblica, *Relazione annuale 2025 sulla politica dell'informazione e della sicurezza*, disponibile su: www.sicurezza-nazionale.gov.it/data/cms/posts/986/attachments/aae81005-6c29-487b-85cc-b0a310f8c9be/download?view=true.

11. Per un approfondimento sulla categoria della sicurezza si v. G.S. Stegher, *Sicurezza*, in T. Groppi (a cura di), *Lessico In-costituzionale*, Editoriale Scientifica, 2025, pp. 479 ss. Diffusamente sul tema si v. P. Piciacchia, *The proliferation of global inscurities and the transformation of the security concept among complex interconnection of responsibilities and impacts of the rule of law*, in Id. (a cura di), *Security, Constitutionalism and Legal Systems*, Giappichelli, 2025, spec. pp. 5 ss.

12. In tali casi, come noto, l'ordinamento consente l'attivazione dello Stato di guerra e il trasferimento al Governo dei poteri necessari per farvi fronte. Per un'analisi più ampia sul riassetto della forma di governo in caso di guerra si v. U. Ronga, *L'emergenza in Parlamento. Attraverso e oltre la pandemia*, Editoriale Scientifica, 52 ss.; più in generale, sulle trasformazioni del ruolo delle Camere, si v. F. Fabrizzi, *Il Parlamento nella "democrazia decidente". Il ruolo delle Camere oggi*, il Mulino, 2024, pp. 25 ss.

13. Approfondisce il tema C. Sbailò, *Guerre ibride: quali risposte possibili?*, in *DPCE online*, n. 63, 2024, pp. 437 ss. Sul punto ulteriori approfondimenti possono essere tratti da S. Monaghan, *Countering Hybrid Warfare*, in *Institute for National Strategic Security*, n. 8, 2019, pp. 86 ss.; inoltre, si v. anche G. Morgese, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, in M. Messina (a cura di), *Strengthening the European Union through the European Citizenship and the Rule of Law*, Editoriale scientifica, 2024, p. 100. Di recente il tema è stato considerato da G. De Vergottini, *Una rilettura del concetto di sicurezza nell'era digitale e della emergenza normalizzata*, in *Rivista AIC*, n. 4, 2019, p. 69. Per un approfondimento sulla categoria della guerra tradizionale si v. M. Cavino, *Il governo della guerra*, in *Quaderni costituzionali*, n. 4, 2022, pp. 753 ss.

14. Per una approfondita analisi riguardo alle differenze tra le guerre tradizionali e le guerre ibride si v. M.G. Losano, *Guerre ibride, omicidi mirati, droni: conflitti senza frontiere e senza diritto*, in L. Forni, T. Vettor (a cura di), *Sicurezza e libertà in tempi di terrorismo globale*, Giappichelli, 2018, p. 21. In senso più ampio, quanto alle ricadute su diritti e libertà, si v. F. Dal Canto, P. Consorti, S. Panizza (a cura di), *Le libertà spaventate. Contributo al dibattito sulle libertà in tempi di crisi*, Pisa University Press, 2016, pp. 10 ss.

assunto una centralità sempre maggiore nel dibattito politico, istituzionale¹⁵ e scientifico¹⁶.

Muovendo da tali premesse, la ricerca sviluppa tre principali nuclei tematici. Il primo ricostruisce le misure europee (a partire da *Digital Services Act* e *AI Act*¹⁷) e nazionali (*Strategia Nazionale di cybersicurezza*) contro la disinformazione *online*. Il secondo, poi, considera il rendimento di tali misure alla prova dell'operazione di *deepfake Storm-1516*. Il terzo, infine, sviluppa alcune ipotesi di implementazione delle misure di contrasto alla disinformazione strategica.

2. Misure europee: Digital service act e AI Act

Le misure di contrasto alla disinformazione strategica *online* hanno trovato maggiore campo nell'agenda europea¹⁸ soprattutto a partire dal 2014, dopo l'annessione della Crimea da parte della Russia¹⁹.

15. Sembra utile segnalare che in data 18 novembre 2025 il Presidente della Repubblica Sergio Mattarella ha convocato una riunione del Consiglio Supremo di Difesa proprio per discutere sulle strategie per affrontare la manipolazione informativa nello spazio digitale e le minacce ibride alla sicurezza nazionale. Per un approfondimento sul concetto di minaccia ibrida si rinvia a uno dei primi studi sul tema, condotto dalla dottrina statunitense: F. Hoffman, *Conflict in the 21st century. The rise of hybrid wars*, Potomac Institute for Policy Studies, 2007.

16. Cfr. A. Vedaschi, *Il dinamismo del concetto di guerra: una sfida per il diritto*, in *Meridiana*, n. 110, 2024, p. 156.

17. Il tema è considerato, di recente, da L. Califano, *La strategia normativa dell'Unione europea per un nuovo ordine digitale*, in *federalismi.it*, n. 15, 2025, spec. 14 ss. Ulteriori approfondimenti, con specifico riferimento alla complessa questione delle libertà nello spazio digitale si v. L. Durst, *La libertà individuale nel contesto digitale. Profili costituzionali*, in M. Belletti, A. Buratti, L. Durst, P. Mezzanotte, M.G. Rodomonte (a cura di), *Diritti e Costituzione all'alba del terzo millennio. "Liber Amicorum" per Roberto Nania*, Giappichelli, 2025, pp. 291 ss.

18. L'elenco delle misure adottate è riportato da G. Morgese, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, cit., p. 100; inoltre, per ulteriori apprendimenti sulle politiche di sicurezza europea, si v. le considerazioni di E. Gianfrancesco, *La politica estera e di sicurezza europea prima e dopo il trattato di Lisbona*, in *Forum di Quaderni costituzionali*, 2011, pp. 1 ss.

19. In quella circostanza all'aggressione armata fecero seguito numerose azioni "ibride", messe in campo nei confronti dell'Unione europea e degli Stati membri per destrutturare la coesione interna, orientare l'opinione pubblica e influenzare le scelte di politica estera verso una linea più favorevole agli interessi del Cremlino. Uno dei principali "campi di battaglia" delle azioni ibride è stato proprio l'Italia: dopo il 2014-2015, sono state diffuse attraverso il *web* numerose *fake news* sugli effetti per l'economia italiana delle sanzioni comminate dall'UE alla Russia allo scopo di «sfruttare tendenze pregresse proprie del panorama dell'opinione pubblica italiana per indebolire la posizione politica di Roma», alimentare divisioni interne sulla politica estera e, di conseguenza, indebolire la complessiva strategia europea per arginare lo sconfinamento russo (cfr. M. Freyrie, *Propaganda, il "quarto potere" diffuso di Putin in Italia*, in *ISPI*, 10 maggio 2022).

Le principali misure sono previste dal *Digital Service Act*²⁰ e dall'*AI Act*²¹, cui si aggiunge il *Codice sulla disinformazione*, sottoscritto, nel febbraio 2025, da 41 dei maggiori operatori del *web* (tra cui Facebook, Instagram, TikTok, LinkedIn e Google Search).

Quanto al Dsa – tra i vari ambiti di regolazione – prevede meccanismi di controllo per contrastare i «rischi sistemici» della comunicazione via *web*²². Secondo questa prospettiva, piattaforme e motori di ricerca sono chiamati a rimuovere, su segnalazione di utenti o autorità competenti²³, i contenuti che possano pregiudicare il dibattito politico interno, le procedure elettorali²⁴ o che possano recare danno alla sicurezza nazionale e alla salute individuale e collettiva²⁵. Altresì – ed è il meccanismo “incentivante” della clausola del *buon samaritano* – si prevede l’esonero dalla responsabilità civile per i gestori delle piattaforme che dimostrino di aver rimosso tempestivamente contenuti illeciti su base volontaria²⁶.

20. Regolamento UE 2022/2065.

21. Regolamento UE 1689/2024. A tali misure si aggiungono quelle previste dal Regolamento (UE) 2024/1083 del Parlamento europeo e del Consiglio, dell’11 aprile 2024, che istituisce un quadro comune per i servizi di media nell’ambito del mercato interno e che modifica la direttiva 2010/13/UE (regolamento europeo sulla libertà dei media) – *Media freedom Act*. Sul punto, per un’analisi più ampia, anche con specifico riferimento alle misure in materia contrasto alla disinformazione si v. D. Borges, *Media ownership transparency and the European Media Freedom Act: how did the EU get there?*, in *Rivista italiana di informatica e diritto*, 2024, p. 259 ss.

22. Sul punto si v. C. Hassan, C. Pinelli, *Disinformazione e democrazia. Populismo, rete e regolazione*, Marsilio, 2022, pp. 55 ss.

23. Cfr. Art. 34 Regolamento UE 2022/2065 (*Digital Service Act*). Per rafforzare il meccanismo di controllo il Dsa – all’art. 22 – prevede i cosiddetti *trusted flaggers*, ossia soggetti con comprovata esperienza che godono di un canale prioritario per segnalare contenuti illegali. Dopo aver ricevuto la segnalazione dai *trusted flagger* le piattaforme sono obbligate a dar seguito alla segnalazione, in tempi di risposta più rapidi rispetto a quelli ordinari (Cfr. G. Contissa, F. Galli, *La governance della “disinformazione aumentata” tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025, p. 136). Specificamente sul tema si v. anche il recente A. Ruffo, *I limiti di Kpi nel contrasto alla disinformazione online e i rischi per lo Stato*, in *federalismi.it*, n. 3, 2025, pp. 547 ss.

24. Cfr. G. Di Cosimo, *Campagne elettorali nel sistema digitale e discorsi d’odio*, in *federalismi.it*, n. 26, 2025, pp. 26 ss.

25. Poi, ulteriori obblighi di prevenzione sono previsti per prevenire forme di violenza di genere e nei confronti dei minori in rete. Sul punto si v. l’art. 34 Regolamento UE 2022/2065 (*Digital Service Act*).

26. Ricostruisce la disciplina sulla responsabilità dei gestori delle piattaforme A. Vicinanza, *La responsabilità delle piattaforme digitali nei confronti dei contenuti illegali: dal caso Telegram al Digital Service Act*, in *Quaderni AISDUE*, n. 1, 2025, pp. 1 ss. Su questo piano, viene in rilievo una delle principali novità del Dsa: la «clausola del buon samaritano» prevista dall’art. 7 Regolamento UE 2022/2065 (*Digital Service Act*). Si tratta di un meccanismo incentivante, che permette ai prestatori di servizi *online* di non perdere automaticamente il beneficio della

Il mancato rispetto degli obblighi di controllo comporta severe sanzioni, parametricate sulla base del fatturato del gestore del servizio²⁷.

In questo quadro, il Dsa promuove l'adozione di Codici di condotta da parte dei principali operatori del mercato. Tra questi, viene in rilievo il *Codice di buone pratiche sulla disinformazione*²⁸, che prevede 44 impegni per garantire la qualità dell'informazione e 128 misure nei seguenti ambiti: demonetizzazione degli introiti pubblicitari per gli utenti che conducono operazioni di disinformazione attraverso i servizi di *hosting*; trasparenza della pubblicità politica; contrasto alla manipolazione cognitiva; responsabilizzazione degli utenti; potenziamento dei meccanismi di *fact checking*²⁹. Ciò avviene attraverso l'impegno dei firmatari a introdurre sistemi di etichettatura chiara degli annunci politici (recanti sponsor, importi spesi e periodo di diffusione, nonché a creare biblioteche pubblicitarie pubbliche e consultabili, che consentano un controllo diffuso sulla provenienza e sulle modalità di circolazione dei messaggi politici), a contrastare i comportamenti manipolatori che favoriscono la diffusione della disinformazione (la creazione di account falsi, l'uso di *bot* automatizzati, l'amplificazione artificiale dei contenuti, l'impersonificazione digitale e la produzione di *deepfake* malevoli), a concordare una definizione condivisa delle pratiche manipolative vietate e ad aggiornare periodicamente l'elenco delle tattiche, tecniche e procedure (Ttp) utilizzate da chi diffonde contenuti illeciti.

L'*AI Act*, in continuità, comprende ulteriori misure per prevenire disinformazione e manipolazione cognitiva³⁰. Com'è noto, esso costituisce la prima forma di regolazione europea *risk-based*³¹ recante norme sull'utilizzo dell'*IA*

irresponsabilità civile nel caso in cui rimuovano, senza alcuna segnalazione, contenuti illeciti a esito di «indagini volontarie di propria iniziativa» o «altre misure volte a individuare, identificare e rimuovere contenuti illegali o a disabilitare l'accesso agli stessi» (cfr. G. Caggiano, *La proposta di Digital Service Act per la regolazione dei servizi e delle piattaforme online nel Diritto dell'Unione europea*, in *AISDUE*, n. 1, 2021, p. 5).

27. Le sanzioni economiche nei confronti dei prestatori dei servizi online possono arrivare al 6% del fatturato annuo globale della piattaforma, cui può aggiungersi una ulteriore penalità del 5% del fatturato globale giornaliero. Percentuali, queste, di significativo rilievo: basti pensare che nell'anno 2024 il fatturato globale annuo di Google è stato di circa 350 miliardi di dollari e che il fatturato giornaliero è di circa un miliardo di dollari. Su tale profilo si v. F. Zorzi Giustiniani, *L'Unione europea e regolamentazione del digitale: il Digital Services Package e il Codice di buone pratiche sulla disinformazione*, in *Nomos*, n. 2, 2022, pp. 2 ss.

28. Sottoscritto nel 2025 da 42 firmatari, tra cui Facebook, Instagram, LinkedIn, Bin, TikTok, YouTube, Google search.

29. Sul punto, si v. G. Lasorella, *La "par condicio" televisiva revisited: affinamenti e innovazioni in un contesto in trasformazione*, in *Nomos*, n. 2, 2025, pp. 25 ss.

30. Si tratta del Regolamento UE 2024/1689.

31. Il Regolamento considera quattro differenti categorie di rischio di utilizzo dei sistemi dell'intelligenza artificiale per salute, sicurezza e diritti fondamentali dell'individuo. Per ciascu-

secondo i principi di precauzione e proporzionalità³². Tra le principali misure: la conoscibilità di sistemi progettati per interagire direttamente con le persone fisiche, senza alcuna intermediazione umana³³; la trasparenza dei materiali video o audio creati con l'intelligenza artificiale, per evidenziare che non siano reali ma *deep fake*; il divieto di commercializzare sistemi di IA che possano, attraverso tecniche subliminali, ingannare o manipolare una persona o un gruppo di persone inducendole a comportamenti che non avrebbero autonomamente compiuto³⁴. Si tratta, in quest'ultimo caso, di sistemi che potrebbero essere impiegati in molteplici ambiti, comprese le procedure elettorali, i *referendum* e la formazione del consenso politico (ciascuno oggetto di specifica tutela) ma il cui rischio per salute, sicurezza e diritti fondamentali è considerato *eccessivo* (*high risk*) e, pertanto, se ne preclude l'utilizzo³⁵.

3. Misure nazionali: quadro normativo incerto e ruolo delle Autorità indipendenti

Diversamente dalla dimensione sovranazionale, sul piano nazionale non sono state ancora adottate strategie organiche e mirate per fronteggiare il fenomeno della disinformazione strategica *online*. Il quadro normativo, infatti, è frammentato – in parte, ancora in corso di ulteriore definizione in sede attuativa – e si registrano interventi solo parziali, prevalentemente riconducibili all'attività di alcune Autorità indipendenti.

I principali interventi sono previsti nell'ambito di misure a tutela della *cybersicurezza* e alla difesa dello spazio digitale³⁶. Tra questi, in particolare, ven-

na di tali categorie, secondo una logica orientata dal principio di precauzione e da quello di proporzionalità, sono previste diverse forme di controllo da parte dell'Unione e, nei casi di maggiore pericolo, obblighi per fornitori dei sistemi e utilizzatori.

32. Alcune considerazioni su questo approccio sono svolte da G. Guzzetta, *La svolta epocale dell'intelligenza artificiale*, in *Rivista della Corte dei Conti*, n. 2, 2024, p. 1 ss.

33. Sistemi che siano costruiti in modo da rendere pienamente consapevole la persona di star interagendo con un sistema informatico: cfr. art. 50, par. 1 del Codice.

34. Sul punto, F. Lagioia, G. Sartor, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *federalismi.it*, 11, 2020, p. 102; e, più in generale, M. Betzu, G. Demuro, *I big data e i rischi per la democrazia rappresentativa*, in *MediaLaws*, n. 6, 2020, p. 221.

35. Sull'impatto delle Ict nelle campagne elettorali cfr. L. Cavalli, *L'impatto delle Ict sulle campagne elettorali*, in *Rivista del Gruppo di Pisa*, n. 2, 2025, pp. 93 ss.

36. Sulla regolazione dello "spazio digitale", cfr. M.C. Girardi, *Libertà e limiti della comunicazione nello spazio pubblico digitale*, in *federalismi.it*, n. 17, 2024, 150, in cui si rileva che «oggi la vita di relazione è contraddistinta da almeno due ordini di fattori, la virtualità – nel senso che nelle relazioni è sempre più marginale la dimensione fisica – e la rapidità con cui maturano i giudizi sui soggetti e, naturalmente, sulla loro reputazione».

gono in rilievo il Dpcm del 17 febbraio 2017 (cosiddetto decreto Gentiloni), attraverso cui sono state adottate linee direttive in attuazione della legge 3 agosto 2007, n. 124 recante norme su *Sistema di informazione per la sicurezza della Repubblica e nuova disciplina del segreto*; il decreto legislativo 18 maggio 2018, n. 65 che ha previsto interventi per implementare le misure di sicurezza per *Operatori di Servizi Essenziali e Fornitori di Servizi digitali*; e il decreto-legge 21 settembre 2019, n. 105, che ha istituito il «perimetro di sicurezza nazionale cibernetica» e previsto norme per proteggere i sistemi informatici delle amministrazioni pubbliche, degli enti e operatori nazionali, pubblici e privati, da minacce esterne³⁷. Da ultimo, in continuità, viene in rilievo la recente approvazione della legge 23 settembre 2025, n. 132 recante “*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*” che, in attuazione della disciplina europea, prevede misure per prevenire l’impiego dell’IA nel sistema dell’informazione e nelle operazioni di *deepfake*, anche attraverso un articolato sistema di *governance*.

In questo quadro, i principali interventi di contrasto alla disinformazione *online* sono stati compiuti, come anticipato, dall’*Autorità Garante per le Comunicazioni* e dal *Garante dei dati personali*.

In particolare, l’Agcom ha messo in campo alcune iniziative e ha promosso varie indagini conoscitive – da ultimo nel 2024 – sull’operato delle piattaforme digitali e sul sistema dell’informazione *online*³⁸. Poi, attraverso l’istituzione del *Tavolo tecnico per la garanzia del pluralismo e della correttezza dell’informazione sulle piattaforme* ha attivato forme di dialogo con i principali prestatori dei servizi in rete per concordare forme di autoregolazione orientate a prevenire la manipolazione *online*. Tra le principali attività del Tavolo assumono specifico interesse le linee guida per la parità di accesso alle piattaforme *online* durante la campagna elettorale 2018³⁹. Ancora, l’Autorità ha avviato attività di vigilanza sui principali *social network*, anche attraverso strumenti di

37. Per una ricostruzione del quadro normativo si v. C. Lobascio, *La disinformazione come strumento di “guerra ibrida”: strategie di risposta*, in S. Sassi, A. Sterpa (a cura di), *Minacce ibride alla sicurezza nazionale. Disinformazione e migrazioni*, cit., pp. 122 ss.

38. Sul punto, per un’analisi più ampia, si v. M. Giannelli, *Poteri dell’Agcom e uso degli algoritmi, tra innovazioni tecnologiche ed evoluzioni del quadro regolamentare*, in *Osservatorio sulle fonti*, n. 2, 2022, p. 867.

39. Il Tavolo è stato istituito dall’Agcom con delibera 423/17/CONS, dal titolo Istituzione di un tavolo tecnico per la garanzia del pluralismo e della correttezza dell’informazione sulle piattaforme digitali, disponibile su: www.agcom.it/sites/default/files/migration/delibera/Delibera%20423-17-CONS.pdf. L’iniziativa è promossa dall’Autorità «al fine di favorire e promuovere l’adozione condivisa di modalità appropriate di intervento nel contrasto ai fenomeni di disinformazione e lesione del pluralismo informativo online» (art. 1, co. 1, delibera).

machine learning per riconoscere eventuali forme di manipolazione dell'opinione pubblica⁴⁰.

Il *Garante per la protezione dei dati personali*, in continuità, ha adottato varie azioni per affrontare le campagne di disinformazione. Da un lato, ha sanzionato le piattaforme *online* che impiegano tali tecniche di manipolazione con la limitazione provvisoria dei trattamenti di dati personali⁴¹; dall'altro, ha avviato anche una campagna di sensibilizzazione e informazione pubblica⁴².

Oltre alle attività delle Autorità amministrative indipendenti si segnala l'istituzione, avvenuta con il decreto-legge 14 giugno 2021, n. 21, dell'*Agenzia per la Cybersicurezza nazionale*. L'Agenzia – introdotta nell'ambito delle azioni previste al Pnrr per garantire la tutela della sicurezza nazionale nello spazio digitale⁴³ – tra i suoi principali compiti ha quello di adottare la *Strategia Nazionale di cybersicurezza*⁴⁴. Attualmente, la *Strategia nazionale 2022-2026* include tra le sue sfide «la disinformazione *online* nel più ampio contesto della c.d. minaccia ibrida». Una sfida che viene considerata come uno dei principali presupposti per rendere pienamente efficace il processo di transizione digitale innescato dal Pnrr⁴⁵. Per questo motivo, l'Agenzia sottolinea l'esigenza di una «azione di coordinamento nazionale, coerente con le iniziative adottate a livello europeo e in sinergia con i Paesi *like minded*» al fine di «prevenire e contrastare la disinformazione online, che sfruttando le caratteristiche del dominio cibernetico, mira a condizionare/influenzare processi politici, economici e sociali del Paese»⁴⁶.

40. Cfr. M. Giannelli, *Poteri dell'Agcom e uso degli algoritmi, tra innovazioni tecnologiche ed evoluzioni del quadro regolamentare*, cit., p. 867; G. Di Cosimo, *Campagne elettorali nel sistema digitale e discorsi d'odio*, cit., p. 26 ss.

41. Cfr. M. Rubechi, *La transizione verso il nuovo sistema delle fonti europee di protezione dei dati*, in L. Califano, C. Colapietro (a cura di), *Innovazione tecnologica e valore della persona: il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Editoriale Scientifica, 2017, pp. 367 ss.

42. In questo ambito è stato pubblicato un *vademecum* sui modi per riconoscere le operazioni di *deepfake* e sulle modalità di segnalazione del contenuto illecito al Garante.

43. Cfr. U. Ronga, *Indirizzo politico e negoziazione interistituzionale nel Pnrr*, in L. Cassetti, F. Fabrizzi, A. Morrone, F. Savastano, A. Sterpa (a cura di), *Scritti in memoria di Beniamino Caravita di Toritto*, Editoriale Scientifica, 2024, pp. 3149 ss.

44. Sul punto, M. Matassa, *Una strategia nazionale a difesa del cyberspazio*, in *Persona e Amministrazione*, n. 1, 2022, p. 626.

45. Cfr. U. Ronga, *La sostenibilità ambientale nella transizione ecologica e digitale del Pnrr. Un primo bilancio*, in *Nomos*, n. 1, 2023, p. 2 ss.

46. I virgolettati sono tratti dal Piano di Implementazione della Strategia nazionale di cybersicurezza, disponibile su: www.acn.gov.it/portale/documents/20119/87708/Acn_Implementazione.pdf.

4. Caso studio: l'operazione *Storm-1516*

Le misure europee e nazionali considerate sono state messe alla prova dalla recente operazione *Storm-1516*⁴⁷: una articolata campagna di manipolazione strategica messa in campo da *hackers* collegati all'*intelligence* russa, realizzata con il principale obiettivo di screditare il Governo ucraino agli occhi dell'opinione pubblica occidentale e di rafforzare movimenti, associazioni e partiti filorussi in Europa⁴⁸.

Dal punto di vista operativo, la campagna di disinformazione si è basata su una logica ibrida e parametrata sulla base di obiettivi differenti: linguaggio, tono e tipo di contenuti sono stati generati tenendo conto del contesto nazionale dello Stato *target* e della sensibilità del pubblico di riferimento⁴⁹. In Europa, ad esempio, sono stati prevalentemente condivisi contenuti su temi come immigrazione, sicurezza, terrorismo e crisi economica⁵⁰; negli Stati Uniti, invece, sono stati diffusi per lo più contenuti per alimentare la sfiducia verso le istituzioni federali e i grandi partiti.

Secondo tali modalità, l'operazione ha previsto l'impiego di due principali flussi narrativi: in alcuni casi sono stati diffusi, anche attraverso l'impiego delle tecniche di *deepfake*, false dichiarazioni da parte di esponenti governativi; in altri casi, invece, semplici contenuti da parte di utenti, rilanciati attraverso una fitta di rete di account tra loro collegati. Flussi non episodici, ma azioni sincronizzate, che hanno mostrato non solo una profonda capacità di analisi del contesto e di programmazione, ma che hanno rivelato altresì un'accurata attenzione ai "tempi" della comunicazione. Le operazioni, infatti, sono intervenute nei momenti di maggiore esposizione mediatica – campagne elettorali, crisi internazionali, grandi eventi sportivi o simbolici – per ottenere la massima visibilità e sfruttare la vulnerabilità emotiva dell'opinione pubblica.

47. Così F. Giaquinta, *Storm-1516: quando la disinformazione minaccia la democrazia*, in *Agendadigitale.eu*, 23 maggio 2025.

48. Così il Rapporto *Analysis of the Russian information manipulation set Storm-1516*, redatto dal *Secretariat general de la defense et de la securité nationale*.

49. Le avanzatissime tecnologie impiegate per realizzare tali contenuti rivale un ingente stanziamento di risorse a sostegno di tali operazioni: d'altronde, secondo recenti stime, la Russia spende circa 2 miliardi di dollari l'anno in operazioni per operazioni di propaganda in occidente, attraverso media, piattaforme *web*, movimenti politici, partiti, associazioni (cfr. R. Galullo, A. Mincuzzi, *Fake news, Putin spende 2 miliardi di dollari all'anno*, in *Il Sole 24 ore*, 14 maggio 2024).

50. L'Intelligence austriaca, ad esempio, ha dichiarato di avere identificato una importante campagna di disinformazione messa in campo dalla Russia per promuovere estremismi contro l'Ucraina; le risultanze dell'indagine sono disponibili su: www.reuters.com/world/europe/austria-uncovers-russian-disinformation-campaign-security-service-says-2025-03-24/.

La strategia di diffusione di tali contenuti si è articolata in quattro fasi principali: creazione del contenuto; diffusione sul *web* (*seeding*); trasmissione del contenuto a piattaforme *web* e *media* stranieri (*narrative laundering*); amplificazione e diffusione su larga scala della notizia⁵¹.

Più nel merito, le fasi di creazione e diffusione del contenuto hanno previsto l'impiego modalità diverse. Esso è stato solitamente pubblicato da *burner accounts*: profili, cioè, non associati ad alcun individuo e creati con il solo scopo di pubblicare un messaggio o un video per poi essere eliminati e rendere più complesso il tracciamento dei canali di diffusione delle informazioni⁵². In tal senso, oltre ai profili falsi, non associati a individui reali, in alcuni casi la trasmissione di contenuti è avvenuta attraverso il coinvolgimento di *account* acquistati o noleggiati da intermediari del *web*: *account* di persone reali, già dotati di una rete di contatti, *followers* e interazioni autentiche, che hanno giocato un ruolo chiave nella diffusione di informazioni false. La parvenza di normalità di tali *account* ha facilitato la circolazione dei contenuti: un messaggio pubblicato da un profilo che sembra umano, con uno "storico" di interazioni con contatti reali, ha molta più probabilità di essere percepito come "affidabile" dagli utenti del *web* e, dunque, di essere eventualmente ricondiviso rispetto a quello veicolato da *account* anonimi o appena creati⁵³.

Infine, una ulteriore modalità di diffusione dei contenuti orientati alla manipolazione cognitiva è avvenuta attraverso *account* riconducibili alla rete *CopyCop*: un *network* coordinato di siti *web*, che hanno impiegato l'intelligenza artificiale per diffondere notizie false e mettere in campo operazioni di disinformazione strategica in Europa. In questo sistema sono state create riproduzioni fraudolente dei siti *web* di autorevoli quotidiani europei, come *Le Monde*, *The Guardian*, *Bild*, *La Repubblica*, *Ansa*⁵⁴. Attraverso tali canali sono

51. Approfonditamente sul punto si v. il rapporto del *Secrétariat général del la défense et de la sécurité nationale* del Primo ministro francese dal titolo *Analysis of the Russian information manipulation set Storm-1516*.

52. Si v. il rapporto del *Secrétariat général del la défense et de la sécurité nationale* del Primo ministro francese dal titolo *Analysis of the Russian information manipulation set Storm-1516*.

53. Si tratta di strategie non recenti ma che, da ultimo e grazie all'impiego dell'IA, diventano sempre più permeanti: cfr. M. Mezzanotte, *Fake news nelle campagne elettorali digitali. Vecchi rimedi o nuove regole?*, in *federalismi.it*, n. 24, 2018, p. 1. Tra gli *account* "reali", è il caso dei contenuti diffusi da Adrien Bocquet: un ex militare francese esiliato in Russia, che attraverso il suo *account* X accusava Leonid Volkov – uno dei principali oppositori del Presidente della Russia – di aver venduto rifugiati ucraini a schiavisti europei. Descrive questo episodio A. Zitelli, *Storm-1516, l'operazione di disinformazione filorusa che minaccia il dibattito pubblico*, in *Facta.news*, 12 maggio 2025.

54. Cfr. Così F. Giaquinta, *Storm-1516: quando la disinformazione minaccia la democrazia*, in *Agendadigitale.eu*, 23 maggio 2025.

stati diffusi contenuti falsi *online*, facendo leva sulla credibilità dei quotidiani contraffatti e sull'inconsapevole affidamento degli utenti meno capaci di individuare le difformità tra il sito web falso e quello originale⁵⁵.

A esito della pubblicazione in rete, l'operazione ha previsto una fase di «lavaggio narrativo»⁵⁶: un processo orientato a occultare l'origine di una informazione attraverso la sua ripubblicazione su profili social più affidabili. A tale scopo, sono stati coinvolti *media* e piattaforme di rilievo locale complici, soprattutto africani (come il nigeriano *Dailypost*) e del Medio Oriente, nonché *broadcaster* e *influencer* filorussi. Questa fase – orientata a presentare la storia come proveniente da fonti locali o internazionalmente riconosciute – è stata propedeutica alla successiva amplificazione del contenuto, attraverso *bot Telegram*, gruppi *Facebook* e reti filorusse.

A livello europeo, le azioni di disinformazione strategica connesse all'operazione *Storm-1516* hanno prevalentemente – ma non solo – riguardato Germania e Francia. In particolare, secondo un recente *report*, sono stati individuati 102 portali *web* falsi attraverso cui sono state diffusi, anche attraverso l'impiego delle tecniche di *deepfake*, falsi messaggi di politici e contenuti per condizionare l'esito del voto⁵⁷. Allo stesso modo, in Francia sono circolate informazioni false in 38.877 *post* sui *social network*: informazioni che hanno avuto una diffusione notevole, che hanno registrato 55,8 milioni di visualizzazioni⁵⁸.

Anche l'Italia è stata bersaglio «di un'ampia campagna globale di manipolazione informativa e ingerenza volta a polarizzare l'opinione pubblica occidentale»⁵⁹ a favore, tra gli altri, della Russia. Il principale obiettivo è stato indebolire la posizione del Paese nel G7 e nell'Unione europea, promuovendo campagne di disinformazione sulla decisione di uscire dall'accordo commerciale della Via della Seta e sul conflitto israelo-palestinese. In questo quadro, i servizi di *intelligence* hanno individuato almeno 75 azioni di propaganda ostile

55. *Ibidem*.

56. L'espressione è tratta dal citato rapporto del Segretariato francese per la difesa della sicurezza nazionale (p. 16).

57. Tra i casi di notizie falsi più eclatanti vengono in rilievo un video in cui si riprende una lussuosa villa a Los Angeles dal valore di novanta milioni di dollari che viene falsamente ricondotta al Cancelliere tedesco Scholz; notizie in cui si accusa il segretario del partito dei verdi tedesco di violenza sessuale e in cui si afferma che il Presidente della Commissione parlamentare di difesa della Germania sia un agente russo.

58. Sul punto si v. N. Huet, M. Sadeghi, C. Labbe, *Un'operazione di propaganda russa colpisce la Francia con scandali inventati dall'intelligenza artificiale, attirando 55 milioni di visualizzazioni sui social network*, in *News Guard*, 17 aprile 2025.

59. Così Presidenza del Consiglio dei Ministri, Sistema di Informazione per la Sicurezza della Repubblica, *Relazione annuale 2025 sulla politica dell'informazione per la sicurezza*, Roma, 2025, p. 30.

orientate a favorire esponenti politici filorussi o a rafforzare il prestigio della Russia nell'opinione pubblica⁶⁰.

L'operazione Storm-1516, dunque, ha rivelato alcune falle nel sistema di tutela europeo e nazionale di difesa dalle minacce ibride. Infatti, nonostante il complesso coordinamento che collega le istituzioni europee nel contrasto alla disinformazione strategica⁶¹, tali campagne sono riuscite – comunque – a produrre un impatto non trascurabile nel sistema informativo. Peraltro, a una complessiva inefficienza di tale sistema di difesa, si aggiunge una certa inerzia dei prestatori dei servizi in rete nel prevenire le minacce ibride, limitandosi alla rimozione dei contenuti segnalati da utenti e autorità competenti, senza un effettivo approccio proattivo.

In altre parole, il caso *Storm-1516* mostra, forse in modo emblematico, come i sistemi democratici europei siano ancora vulnerabili rispetto alle operazioni di disinformazione strategica⁶². La capacità di attori esterni di sfruttare, in modo sofisticato e coordinato, le asimmetrie tecnologiche, cognitive e normative dello spazio digitale europeo evidenzia l'urgenza di ripensare le politiche di sicurezza informativa⁶³ e, in senso più ampio, di tutela dello spazio pubblico digitale⁶⁴.

5. Conclusioni: rendimento delle misure e ipotesi per una regolazione euro-nazionale

Lo studio condotto consente di svolgere alcune considerazioni conclusive sul rendimento delle misure europee e nazionali per il contrasto alla disinformazione strategica; e di ragionare, in chiave propositiva, su alcune ipotesi di intervento.

Al riguardo, la vicenda *Storm-1516* conferma un dato di fondo: la imprevedibilità e la incontrollabilità della circolazione dei contenuti pubblicati *online* (veri o falsi che poi si rivelino). Essi, infatti, si inseriscono in un sistema labirin-

60. Sul punto si v. il rapporto del gruppo parlamentare Europeisti MAIE Centro democratico, *La "guerra ibrida" della Russia*, luglio 2025, disponibile su: europeisti.org/linee-guida/.

61. Si v. a tal proposito G. Morgese, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, cit., p. 100.

62. Cfr. P. Passaglia, *Fake news e fake democracy: una convergenza da scongiurare*, in *federalismi.it*, n. 11, 2020, p. 126.

63. Sul punto, M. Galimberti, *La disinformazione: profili regolatori e policy*, in *Diritto Penale Contemporaneo*, n. 4, 2021, p. 251.

64. Per ulteriori apprendimenti si v. A. Mazziotti, *Fake news, fake people e società della (dis)informazione: riflessioni su democrazia, informazione e libertà fondamentali al tempo dei social network*, in *I diritti dell'uomo*, n. 1, 2017, p. 57.

tico e circolano – fino a diventare eventualmente “virali” – tra una molteplicità di piattaforme, *account* interconnessi e siti *web*. E questo è un fattore decisivo – e condizionante – per assumere una specifica scelta di campo su mezzi e modalità di contrasto del fenomeno, a livello europeo e nazionale: sia in chiave di *prevenzione*, con soluzioni che valorizzino il controllo *ex ante* (rispetto al momento della pubblicazione dei contenuti); sia in chiave di *repressione*, con interventi *ex post* che intercettino e impediscano la circolazione dei contenuti orientati alla manipolazione dell’opinione pubblica.

Il sistema di controllo europeo basato sulle segnalazioni degli utenti e delle autorità competenti definito dal *Digital service Act* e ripreso dall’*AI Act* – secondo uno schema di controllo *ex post* – da questo punto di vista, mostra alcuni limiti. D’altronde, nonostante siano previsti alcuni meccanismi incentivanti sia nel *Digital service Act* sia nel *Codice di buone pratiche sulla disinformazione*, si registra una certa inerzia – quando non vera e propria resistenza – delle piattaforme in sede applicativa delle relative misure⁶⁵. Peraltro, risulta particolarmente preoccupante la decisione di piattaforme e organizzazioni di rilievo mondiale come *Microsoft*, *Google* e *X* di rinunciare alle attività di *fact checking*, segnando un significativo arretramento rispetto agli impegni assunti con la sottoscrizione del *Codice*⁶⁶.

Ulteriori criticità, poi, si registrano anche a livello nazionale, a partire dall’assenza di politiche specifiche in materia di disinformazione strategica, nonostante il fenomeno delle ingerenze straniere nel sistema informativo abbia assunto una portata non più trascurabile – come evidenziano i rapporti dei servizi di *intelligence*. Peraltro, il sistema di autorità, agenzie, organismi di vigilanza sulla diffusione delle informazioni in rete è privo di un coordinamento organico.

Se, dunque, il rendimento di queste soluzioni (comprese le modalità di *governance*) si è rivelato solo parziale, occorre ragionare su ipotesi di controllo preventivo da parte delle piattaforme, almeno nei casi in cui il contenuto sia pubblicato da *account* “sospetti” – ma si tratterebbe, in tal caso, di individuare criteri specifici – e non verificati; e tuttavia, occorrerebbe farlo in una chiave euro-nazionale, compiendo un difficile bilanciamento che comprenda esigenze di sicurezza (nazionale) e la garanzia della libertà di espressione (in questo caso, *online*).

65. Per una analisi sulle forme di autoregolamentazione dei principali *social network* si v. P. Bonini, *L’autoregolamentazione dei principali Social Network*, in *federalismi.it*, n. 11, 2020, pp. 265 ss.

66. Sul punto si v. M. Borgobello, *Fact checking addio, anche Google contro la UE: ecco la posta in gioco*, in *Agenda Digitale*, 20 gennaio 2025.

A livello europeo⁶⁷ potrebbe essere valutata l'ipotesi di rafforzare i meccanismi di autenticazione in sede di registrazione alle piattaforme *web* attraverso i sistemi di identità digitale. Tuttavia, con ogni probabilità, tale ipotesi incontrerebbe una netta opposizione da parte degli operatori del *web*⁶⁸. I maggiori profitti dei *services providers*, infatti, derivano dalle campagne pubblicitarie: in altre parole, maggiore è il traffico di utenti, maggiore è l'introito. Il modello di *business* del mercato digitale, dunque, è strettamente connesso alla presenza di utenti in rete: non è necessario che tali utenti corrispondano effettivamente e persone reali; è sufficiente che essi operino nel *web*, generando dati utili per la profilazione da impiegare nelle sponsorizzazioni. Collegare l'accesso ai sistemi di identità digitale equivarrebbe – sotto questo profilo – a un drastico ridimensionamento della popolazione virtuale, con inevitabili ricadute sui ricavi delle pubblicità e delle sponsorizzazioni⁶⁹. In questo quadro, dunque, l'Unione europea è chiamata a trovare un punto di equilibrio che sappia fare sintesi tra l'interesse pubblico a uno spazio *web* trasparente, in cui sia possibile imputare a soggetti reali la diffusione di notizie false in rete; e gli interessi privati (dei *services providers*) – che tendono a preservare uno spazio opaco – all'interno del quale è garantita anche la possibilità di creare – senza alcun limite – identità digitali anonime o non associate a un individuo preciso⁷⁰.

Tuttavia, l'intervento europeo non è – e non deve – essere slegato da soluzioni adeguate a livello nazionale. A tale riguardo, è stato di recente depositato in Senato un disegno di legge per istituire l'*Agenzia sulla disinformazione e la sicurezza cognitiva* presso la Presidenza del Consiglio dei Ministri⁷¹. Tuttavia, più che istituire nuove strutture, si potrebbe ragionare sull'ipotesi di rafforzare i meccanismi di cooperazione tra autorità già esistenti, in una logica di leale collaborazione interistituzionale.

In un quadro così disorganico, l'approvazione della legge 23 settembre 2025, n. 132 recante disposizioni e delega al Governo in materia di intelligen-

67. Una regolazione di questo tipo – per essere efficace – dovrebbe avere una dimensione (almeno) europea. Una disciplina nazionale in materia di autenticazione digitale risulterebbe, pertanto, strutturalmente inadeguata, poiché finirebbe per produrre un effetto asimmetrico, con il rischio di consentire l'accesso delle informazioni da contesti normativi meno rigorosi.

68. Riflette su tali profili M. Betzu, *I baroni del digitale*, Editoriale Scientifica, 2022, spec. pp. 65 ss.

69. Riflette su tali profili C. Bergonzini, "Prova a prendermi". *Ecosistema digitale e consapevolezza degli utenti: uno spazio per la regolazione nazionale?*, Giappichelli, 2022, spec. pp. 98 ss.

70. Anche in questo ambito, dunque, viene in rilievo il tema della negoziazione tra decisore politico e rappresentanti di interessi. A tal proposito, tra gli altri, si v. U. Ronga, *La legislazione negoziata. Autonomia e regolazione nei processi di decisione politica*, Editoriale Scientifica, 2018, spec. pp. 56 ss.

71. Cfr. Senato della repubblica, *Disegno di legge n. 1990*, XIX Legislatura, 28 marzo 2024.

za artificiale può costituire un risultato di interessante (ancorché parziale): nel merito, specie perché prevede norme per la tutela delle procedure elettorali, della libertà del dibattito democratico e per la trasparenza del sistema informativo; e nel metodo, per il legame tra disciplina europea e nazionale⁷². Indubbiamente la strada è ancora lunga e il processo di implementazione degli strumenti di difesa è ancora in divenire. Tuttavia, l'operazione *Storm-1516* ha rivelato l'impatto – non solo potenziale – delle campagne di manipolazione dell'opinione pubblica sui processi elettorali, soprattutto in contesti segnati da bassi livelli di alfabetizzazione digitale e mediatica. Per questo, è necessario ragionare su interventi di sistema, che operino in una prospettiva multilivello e comprendano strategie regolatorie che garantiscano il bilanciamento tra esigenze di sicurezza, libertà di espressione *online* e diritto all'informazione. È una questione cruciale, specie al tempo della c.d. post-verità⁷³, e dove l'alterazione delle informazioni può assurgere a strumento per attacchi *ibridi*. In gioco – oggi più che mai – vi sono le sorti, nonché la tenuta, delle democrazie.

72. Su questi profili di metodo, si v. U. Ronga, *La delega legislativa. Recente rendimento del modello*, Editoriale scientifica, 2020, pp. 120 ss.

73. Cfr. M. Di Fiore, T. Portaluri, A. Saltelli, *Post-verità e scienza*, in *federalismi.it*, n. 11, 2020, pp. 111 ss.

LARGE LANGUAGE MODELS: APPLICAZIONI E CRITICITÀ NEL DIBATTITO PUBBLICO-INFORMATIVO. SPUNTI DI RIFLESSIONE INFORMATICO-GIURIDICI

*Casimiro Coniglione**

SOMMARIO: 1. Introduzione: la Società algoritmica e la sua opacità – 2. Addestramento e funzionalità dei *Large Language Models*. Brevi cenni sulle architetture *transformers* – 3. Due approcci con i Llm nel *fact-checking* – 4. Criticità dei Llm: *hallucinations* e *bias* – 5. Riflessioni conclusive.

1. Introduzione: la Società algoritmica e la sua opacità

Com'è noto, la Società dell'informazione e della conoscenza si è rapidamente evoluta nella Società algoritmica, dove gli algoritmi, robot e agenti intelligenti effettuano – in parziale o totale autonomia – i processi decisionali. Alla centralità dell'informazione si affianca quella degli algoritmi¹ che, eseguiti integralmente dai *software*, elaborano le informazioni stesse, plasmando l'assetto dell'attuale società².

È opportuno precisare che l'espressione "Società algoritmica" indica le attività di utenti, aziende e istituzioni mediate dal potere sociale degli algoritmi. Attualizzando le intuizioni di Michel Foucault (1926-1984) si discute di "governamentalità algoritmica", vale a dire quella specifica arte di governo che assicura la governabilità del *démos* grazie agli algoritmi³.

* Assegnista di ricerca in Filosofia del diritto, Università degli Studi di Modena e Reggio Emilia.

1. Effettivamente, citando le primissime parole di un'opera di Pedro Domingos, «Viviamo nell'era degli algoritmi» (P. Domingos, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, trad. it., Bollati Boringhieri, 2016, p. 1).

2. G. Fioriglio, *La Società algoritmica fra opacità e spiegabilità*, in *Ars Interpretandi*, n. 1, 2021, pp. 53-67.

3. A. Rouvroy, T. Berns, *Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'inviduation par la relation?*, in *Réseaux. Communication. Technologie*.

In questo contesto, il simbolo della Società algoritmica è indubbiamente rappresentato dall'Intelligenza Artificiale⁴ (IA), la quale assicura rapidità, efficienza ed efficacia⁵ nelle prestazioni.

Da una parte, sembra difficile negare all'IA una funzione di ottimizzazione nelle attività⁶; dall'altra parte, essa racchiude nuove e assai rilevanti questioni sia per la filosofia del diritto sia per l'informatica giuridica⁷.

La peculiarità di questa società è la sua immanente opacità, la totale inaccessibilità⁸ della conoscenza del procedimento con cui gli algoritmi di apprendimento elaborano *input* per restituire *output*⁹. Prendendo come punto di rife-

Société, n. 1, 2013, pp. 163-196. Sul concetto classico di governamentalità: M. Foucault, *Del governo dei viventi. Corso al College de France (1978-1979)*, trad. it., Feltrinelli, 2014; Id., *Sicurezza, territorio, popolazione. Corso al College de France (1978-1979)*, trad. it., Feltrinelli, 2015; Id., *Nascita della biopolitica. Corso al College de France (1978-1979)*, trad. it., Feltrinelli, 2017. Per una rilettura filosofico-giuridica della natura del potere nella proposta filosofica di Foucault si veda M. Brigaglia, *Potere. Una rilettura di Michel Foucault*, Editoriale Scientifica, 2019.

4. Sull'IA, *ex multis*, W. Barfield, U. Pagallo, *Advanced Introduction to Law and Artificial Intelligence*, Elgar, 2020; K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, trad. it., il Mulino, 2021; G. Sartor, *L'intelligenza artificiale e il diritto*, Giappichelli, 2022; A.C. Amato Mangiameli, *IA e diritto. In luogo di una introduzione*, in *Journal of Ethics and Legal Technologies*, n. 2, 2023, pp. 81-93; O. Pollicino, F. Donati, G. Finocchiaro, F. Paolucci (a cura di), *La disciplina dell'intelligenza artificiale*, Giuffrè, 2025.

5. F. Faini, *Intelligenza artificiale, diritto e pubblica amministrazione*, in A. D'Aloia (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, FrancoAngeli, 2021, pp. 385-413. Su tali principi, cfr. G. Tuzet, *Effettività, efficacia, efficienza*, in *Materiali per una storia della cultura giuridica*, n. 1, 2016, pp. 207-224.

6. A titolo esemplificativo: A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità*, Mondadori, 2020.

7. Tra le moltissime riflessioni che l'IA pone al diritto, si mettono in rilievo quelle elaborate da Gianfrancesco Zanetti, che argomenta come la mancata conoscenza dell'uso delle nuove tecnologie e dell'IA è da considerarsi una nuova forma di analfabetismo, da cui trarrebbe origine una nuova forma di vulnerabilità nei confronti di tutti coloro che sono rimasti esclusi – sia per scelta sia per pregresse condizioni di vulnerabilità – dalla fruizione. G. Zanetti, *L'impatto delle tecnologie: il ruolo dell'Università dinanzi a nuove sfide*, in *Crònica Jurídica Hispalense*, n. 22, 2024, pp. 369-378.

8. Sul complesso rapporto fra segretezza, potere e democrazia si vedano le osservazioni di M. Ricciardi, *La rilevanza etica del segreto*, in A. Artosi, G. Bongiovanni, S. Vida (a cura di), *Problemi dell'attuazione e della produzione normativa*, Gedit Edizioni, 2001, pp. 425-437; M. Catanzariti, *Segreto e potere. I limiti della democrazia*, Giappichelli, 2014.

9. Riguardo al fenomeno dell'opacità algoritmica la lettura scientifica è assai vasta. Ciò nonostante, è interessante osservare come sin dagli anni Duemila si ponesse questa problematica. Infatti, Lucas D. Intron e Helen Nissenbaum chiedevano «*As a first step we would demand full and truthful disclosure of the underlying rules (or algorithms) governing indexing, searching and prioritizing, stated in a way that is a meaningful to the majority of Web users. Obviously, this might help spammers. However, we would argue that the impact of these unethical practices would be se-*

rimento le argomentazioni di Niklas Luhmann (1927-1998), si può sostenere che si è in presenza di un sistema autopoietico¹⁰, un universo complesso ed ermeticamente chiuso dall'interno in cui non ci sarebbe la possibilità della sua conoscenza dall'esterno¹¹.

Fra l'altro, l'opacità algoritmica ha significative ricadute giuridico-etiche sulle attività dei consociati, in quanto possono amplificarsi discriminazioni e stereotipi, e ciò a causa di potenziali *bias* nei dati di addestramento¹².

Esposte queste brevissime ma necessarie riflessioni introduttive sulla Società algoritmica e la sua immanente opacità, il contributo metterà a fuoco i *Large Language Models* (Llm): specifici modelli di apprendimento automatico statistico-neurale basati su architetture *transformers*. Dopo aver discusso del loro addestramento, si approfondirà l'utilizzo di questi strumenti nel *fact-checking* e le diverse criticità di quest'ultimi nel dibattito pubblico-informativo.

Nelle conclusioni si forniranno degli spunti di riflessione per una *governance* trasparente, responsabile, nonché sicura di siffatte tecnologie linguistico-neurali.

verely dampened if both seekers and those wishing to be found were aware of the particular biased inherent in any given search engine» (L.D. Introna, H. Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, in *The Information Society*, n. 16, 2000, p. 176). Cfr. anche K. Yeung, M. Lodge (eds.), *Algorithmic Regulation*, Oxford University Press, 2019; M. Martoni, *Spiegare le pieghe. Alcune osservazioni su trasparenza e spiegabilità*, in *L'Ircocervo*, n. 1, 2024, pp. 212-227.

10. Il riferimento è, ovviamente, all'opera di N. Luhmann, *Sistemi sociali. Fondamenti di una teoria generale* (1988), trad. it., il Mulino, 1990.

11. Cfr. D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana* (2023), trad. it., Einaudi, 2024, pp. 132-138.

12. Emblematiche, al riguardo, le discussioni sulla giustizia predittiva dove – utilizzando gli algoritmi e l'IA – si tenta di prevedere gli esiti dei procedimenti giudiziari. Ciò nondimeno, sono diverse le criticità fra cui la violazione della presunzione d'innocenza, la deumanizzazione del giudizio e la poca affidabilità dei modelli. Senza pretesa di esaustività, si vedano i lavori di G. Sartor, F. Lagiòia, *Il sistema Compas: algoritmi, previsioni, iniquità*, in U. Ruffolo (a cura di), *XXVI lezioni di Diritto dell'Intelligenza Artificiale*, Giappichelli, 2021, pp. 226-244; D. Roccaro, *Predizioni normative. Giustizia e rivoluzione digitale*, Mimesis, 2023, pp. 95-102; L. Ercole, *Contro la "giustizia predittiva". Per una lettura conservativa del principio di certezza del diritto*, Giappichelli, 2024. Mauro Barberis, invece, preferisce chiamare quest'ultima "giustizia digitale" e non giustizia predittiva: M. Barberis, *Homo Sapiens Vs. Machinal Sapiens? Nove opinioni urticanti*, in V. Mastroiacovo (a cura di), *Giocare con altri dadi. Giustizia e predittività dell'algoritmo*, Giappichelli, 2024, pp. 57-80. Per l'impatto dell'IA nelle classiche professioni giuridiche: P. Moro, *Diritto ibrido. Metodo del giurista e tecnologie emergenti*, Mondadori, 2025.

2. Addestramento e funzionalità dei *Large Language Models*. Brevi cenni sulle architetture *transformers*

Oggi, l'IA non si limita alla mera esecuzione delle attività, ma apprende¹³ e genera contenuti grazie alla grandissima quantità di dati con la quale viene addestrata, al punto tale che l'IA generativa ha ontologicamente una dipendenza dai dati¹⁴.

Una peculiare forma di IA generativa è rappresentata dai *Large Language Models*¹⁵ (Llm), essi generano *output* per rispondere a *prompt* (richieste) degli utenti e nel far ciò si avvalgono sia di reti neurali sia di architetture *transformers*¹⁶, sostituendo così i modelli ricorrenti¹⁷ quale il Rnn (*Recurrent Neural Network*) o il Lstm (*Long Short-Term Memory*). Siffatta tecnologia di IA, come recentemente messo in luce da Federico Oliveri, scandisce il passaggio «da una tecnica-agente a una tecnica agente-linguistico o, se si vuole usare una categoria ancora più ampia, a una tecnica agente-generativa»¹⁸.

Tanto premesso, appare doveroso esporre l'addestramento dei Llm, indirizzando in seguito la riflessione sulla singolarità delle architetture *transformers*.

Generalmente tali modelli vengono addestrati con la tecnica del *masking*, ossia la rimozione di un *token* (che corrisponde pressappoco a una piccola unità di testo come una parola o una sua parte) da una frase¹⁹, domandando (*prompt*)²⁰

13. Da ultimo: L. Avitabile, *Genealogia del diritto. Una riflessione su testualità giuridica e IA*, in *Calumet – Intercultural law and Humanities review*, n. 24, 2025, p. 83-92, in part. p. 84.

14. Cfr. M. Farina, *Degenerazione e rischio creativo dell'Intelligenza Artificiale "forte": forme di prevenzione e tutela complementare*, in *L'Ircocervo*, n. 1, 2024, pp. 122-138, in part. p. 126; A. Palladino, *Il "riposo del vero". La dissimulazione artificiale per la tutela del diritto d'autore*, in *Cyberspazio e diritto. Rivista internazionale di Informatica giuridica*, n. 1, 2025, pp. 121-137, in part. p. 124.

15. Per una puntuale disamina: S. Sapienza, *Decisioni algoritmiche e diritto*, Giuffrè, 2024, pp. 22-26.

16. Imprescindibile, per le architetture *transformers*, il rimando alle lucide analisi di: A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, *Attention Is All You Need*, in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 5998-6008.

17. Cfr. O. Campesato, *Large Language Models for Developers. A Prompt-based Exploration of Llm*, Mercury Learning and Information, 2024.

18. F. Oliveri, *Machina Mundi. Per una regolazione democratica dei poteri digitali*, Mucchi Editore, 2025, p. 46.

19. Per i dettagli tecnici di questa modalità d'addestramento dei Llm: A. Fernando, S. Ranathunga, *Linguistic entity masking to improve cross-lingual representation of multilingual language models for low-resource languages*, in *Knowledge and Information Systems*, n. 9, 2025, pp. 1-42.

20. Sul *prompt* si consulti F. Galli, *L'analisi automatica delle decisioni giudiziali*, Giappichelli, 2025, pp. 80-81.

al sistema quale sarà il *token* con le probabilità più alte a occupare la posizione. Ripetendo questo processo molte volte su un ampissimo insieme di frasi, il sistema imparerà a correlare in sequenza i *token*. Tra le altre cose, questo meccanismo sta alla base di ciò che spesso viene interpretato come ragionamento e/o conoscenza²¹.

Con l'apprendimento per rinforzo (*Reinforcement Learning from Human Feedback*), si raccolgono *feedback* e si domanda di valutare – o classificare – le istanze comportamentali degli agenti. Lo scopo è quello di assicurare un modello che possa raccogliere gli *input* delle sequenze di testo e restituire *output* con qualità e pertinenza, riproducendo le preferenze degli utenti²²; altresì, è possibile addestrare ulteriormente il modello attraverso il *few-shot learning*, con la somministrazione di limitati esempi tratti dai *prompt*, dal caricamento, nonché col *fine-tuning* sui *dataset*²³.

Nella generazione delle risposte i Llm analizzano il *prompt* specifico e lo collocano in una finestra di contesto, che agisce come se fosse una memoria a breve termine. I modelli di larghe dimensioni linguistiche si avvalgono di queste informazioni per predire i lemmi e focalizzarsi su schemi grammaticali, scrittura e/o parti precedenti.

Il procedimento inizia con la scomposizione dei *prompt* in singoli *token*, elaborandoli in diversi strati di neuroni artificiali, ciascuno dei quali estrae caratteristiche sempre più complesse del testo: più parole verranno generate, maggiori saranno le probabilità con cui il sistema amplierà la finestra di contesto²⁴.

21. Del resto, l'obiettivo del gruppo di ricerca sull'IA del 1955 di John McCarthy (1927-2011) era quello di uno studio che «dovrà procedere sulla base dell'ipotesi secondo cui ogni aspetto dell'apprendimento, o qualunque altra caratteristica dell'intelligenza, può essere in linea di principio descritto con tale precisione da poter essere simulato da una macchina. Si tenterà di comprendere come rendere le macchine capaci di usare il linguaggio, formare astrazioni e concetti, risolvere tipi di problemi oggi riservati all'essere umano e migliorare se stesse. Riteniamo che un progresso significativo possa essere raggiunto in uno o più di questi ambiti se un gruppo accuratamente selezionato di scienziati vi lavorerà insieme per un'estate» (J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in *AI Magazine*, vol. 27, n. 4, 2006, p. 12). Cfr. J. McCarthy, P. Hayes, *Some philosophical problems from the standpoint of artificial intelligence*, in *Machine Intelligence*, n. 4, 1969, pp. 463-502.

22. G. Contissa, G. Sartor, *Large Language Models in the Justice Domain*, in G. Lasagni, G. Contissa, M. Caianello (eds.), *Facilitating Judicial Cooperation in the EU. A Computable Approach to Mutual Recognition in Criminal Matters*, Brill-Nijhoff, 2025, pp. 56-58.

23. È una tecnica di apprendimento automatico, che consente di prendere un modello addestrato in precedenza per adattarlo a compiti specifici. Cfr., fra i contributi più recenti, L. Wang, S. Chen, S. Pan, R. Cai, S. Yang, F. Yang, *Parameter-efficient fine-tuning in large language models: a survey of methodologies*, in *Artificial Intelligence Review*, n. 58, 2025, pp. 1-64.

24. G. Contissa, G. Sartor, *Large Language Models in Justice Domain*, in G. Lasagni, G.

Pertanto, com'è agevole constatare, è congenita una funzione predittiva in questi sistemi.

In virtù del fatto che i Llm vengono addestrati su quantità sempre più vaste di dati – e da ciò ha origine quella dipendenza dai dati segnalata nell'*incipit* del paragrafo – non solo vengono raccolte le singole probabilità delle parole, ma anche l'intero contesto delle frasi.

Quanto osservato non sarebbe possibile senza il fondamentale apporto delle architetture *transformers*, modelli di reti neurali²⁵ che trovano relazioni in sequenze di parole o in altri dati usando "l'*Attention*". Questa tecnica focalizza le parti più rilevanti della sequenza, facilitando l'attività di apprendimento e "comprensione" del significato nei testi di dimensioni assai importanti²⁶.

Peraltro, l'introduzione di interfacce conversazionali rapide e immediate fa sì che i Llm vengano adoperati quotidianamente dagli utenti: ChatGpt, Copilot, Claude, Gemini e le risposte elaborate dai sistemi di IA nei motori di ricerca (si pensi a Google) costituiscono esempi paradigmatici di siffatta tecnologia linguistico-neurale.

Alla luce di ciò, vanno valutati i benefici dei Llm nel dibattito pubblico-informativo, specialmente nella rilevazione delle *fake news*.

3. Due approcci con i Llm nel *fact-checking*

Sul *web* circolano un numero indeterminato di contenuti, per cui la "collaborazione" fra operatori, utenti e nuove tecnologie è necessaria, posto che sarebbe assai faticoso, se non del tutto impossibile, riuscire a scrutinare tutti i contenuti veri da quelli falsi.

In questo frangente, i sistemi di IA hanno il pregio di velocizzare/ottimizzare il *fact-checking*, recuperando e incrociando i dati con i contenuti sospetti; oltretutto, procedono tanto alla rilevazione quanto all'identificazione delle fonti da cui hanno origine i contenuti disinformativi. Ciò avviene con l'analisi dei dati testuali da parte dell'IA: l'algoritmo di apprendimento non solo rice-

Contissa, M. Caianello (eds.), *Facilitating Judicial Cooperation in the EU. A Computable Approach*, cit., p. 59.

25. «Le reti neurali sono sistemi informatici che consistono di nodi (i cosiddetti neuroni) collegati da link ai quali sono assegnati pesi numerici [...] la ricerca sulle reti neurali è ispirata all'idea che l'intelligenza possa ottenersi riproducendo l'hardware cerebrale (i neuroni) piuttosto che riproducendo i processi del pensiero cosciente (il ragionamento)» (G. Sartor, *L'intelligenza artificiale e il diritto*, cit., p. 56).

26. N. Koenigstein, *Transformers in Action*, Manning Publications, 2025, pp. 7-12.

ve esempi pratici, bensì – apprendendo – accerta l’attendibilità/inattendibilità dei contenuti²⁷.

Un ruolo cruciale al contrasto della disinformazione è giocato dai Llm, in quanto possiedono un importante potenziale nella rilevazione delle *fake news*. Tante ed eterogenee²⁸ sono fra loro le metodologie ideate per questi sistemi nell’individuazione dei contenuti falsi. Tuttavia, in questa sede ci si occuperà degli approcci più noti e discussi dalla letteratura scientifica: il *Fine-Grained Feedback with Reinforcement Retrieval* (Ffrr) e il *Hierarchical Step-by-Step* (Hiss).

Il primo procedimento scompone singole affermazioni in più domande mirate formulate con il *prompting*, e perlustra gli aspetti fattuali del contenuto informativo. Tale procedimento recupera le prove documentali più attinenti, usando le risposte per supportare la verifica conclusiva della veridicità della *claim* da parte del Llm²⁹. Siffatto approccio si è dimostrato assai accurato in termini di copertura delle prove, perché supera i metodi *baseline*.

Il secondo aiuta il modello a segmentare le affermazioni in forme più granulari e le verifica con sessioni di *question answering*. Le sfide affrontate nel *fact-checking* con questa tecnica sono due: l’omissione dei dettagli più importanti e le *hallucinations* (allucinazioni) provocate dal modello stesso.

Più specificatamente, i Llm segmentano la *claim* per esplicitarne tutti i punti, controllano che ogni sotto-affermazione sia attendibile e riscontrano – formulando e rispondendo a domande mirate – l’attendibilità del contenuto con il vaglio di fonti esterne attendibili. Fatto ciò, ossia dopo aver esaminato tutti gli elementi, il sistema elabora una valutazione complessiva e classifica la *claim* originale³⁰.

Il merito di questo secondo metodo sta nell’esplicabilità³¹ con cui i Llm esaminano il contenuto informativo sospetto.

Ciò nonostante, entrambe le metodologie hanno dei limiti tecnici: il Ffrr è

27. Cfr. G. Contissa, F. Galli, *La governance della “disinformazione aumentata” tra Digital Service Act e AI Act*, in *federalismi.it*, n. 15, 2025, pp. 126-165, in part. pp. 129-131.

28. Sulle diverse e innovative metodologie impiegate dai Llm nella rilevazione delle *fake news*: E. Papageorgiu, C. Chronis, I. Varlamis, Y. Himeur, *A Survey on the Use of Large Language Models (Llm) in Fake News*, in *Future Internet*, vol. 16, n. 8, 2024, pp. 1-29, in part. pp. 15-20.

29. Cfr. M. Wadhwa, X. Zhao, J.J. Li, G. Durrett, *Learning to Refine with Fine-Grained Natural Language Feedback*, in *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 12281-12308.

30. X. Zhang, W. Gao, *Towards Llm Fact-Verification on News Claims with a Hierarchical Step-by-Step Prompting Method*, in *Proceedings of the 13th International Joint Conference of Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, Association for Computational Linguistics*, 2023, pp. 996-1011.

31. Sulla esplicabilità dei Llm si veda il contributo di M. Aspromonte, G. Contissa, F. Galli,

assai oneroso in termini sia computazionali sia di dipendenza dalle fonti esterne; il Hiss dipende fortemente dalla qualità delle informazioni recuperate e può non essere trasparente nell'elaborazione dell'*output*, ed è anche a forte rischio di *bias* nei dati di addestramento.

Gli stessi Llm hanno diverse criticità, possono contribuire all'inquinamento del dibattito pubblico-informativo e alla riproduzione/amplificazione di discriminazioni e/o stereotipi a causa di eventuali *bias*.

4. Criticità dei Llm: *hallucinations* e *bias*

I Llm possono essere accostati al Giano bifronte, la nota divinità romana dai due volti, perché possiedono enormi potenzialità nel *fact-checking*, ma possono anche fuorviare gli utenti “manipolando” e “inquinando” il dibattito pubblico-informativo.

D'altronde, va ricordato che quest'ultimi sono una categoria di IA generativa in grado di produrre contenuti inesistenti, contendendosi la creatività umana³². Più nel dettaglio, nel fenomeno delle *hallucinations* i modelli linguistici di larghe dimensioni elaborano risposte ben scritte e razionali, ma sono false e prive di qualsivoglia riscontro oggettivo.

È ragionevole portare avanti l'idea che i Llm non sono in grado di stabilire l'accuratezza delle loro rappresentazioni, né di passare dal livello linguistico a quello metalinguistico per qualificare gli *output* né hanno una comprensione della realtà, bensì generano risposte che si basano su schemi statistici e sui dati di addestramento, con la conseguenza che generano l'*output* (il contenuto) riflettendone solamente la riproducibilità automatica³³.

Non solo. I *bias* – che possono essere già presenti nei dati di addestramento del modello, nell'addestramento in sé e nell'interazione (*prompting*) – causano una distorsione sistematica e la produzione di risposte non neutrali, parziali ovvero alterate rispetto alla realtà³⁴, amplificando le discriminazioni nei riguardi dei gruppi sociali più vulnerabili.

A. Loreggia, *Beyond Fact-Checking: A Scalable Domain-Agnostic, and Explainable System for Automated Fake News Detention*, in *Proceedings of ROMCIR*, 2025, pp. 70-78.

32. F. Reggio, *Tecnologie per l'umano. Spunti di restorative ethics per una tecnologia digitale “human-centered”*, in Id. (a cura di), *Honeste vivere. Percorsi filosofici per l'etica pubblica*, FrancoAngeli, 2025, p. 295.

33. Cfr. L. Paseri, M. Durante, *Examining epistemological challenges of large language models in law*, in *Cambridge Forum on AI: Law and Governance*, n. 1, 2025, pp. 1-15, in part. pp. 3-4.

34. S. Pietropaoli, *La parola del diritto e il verso della macchina. Per una critica della “decisione algoritmica”*, in *Mechane*, n. 8, 2024, pp. 49-59, in part. p. 51.

A questo proposito, possono svolgersi ulteriori riflessioni sull'opacità algoritmica e sul ruolo degli utenti.

In primo luogo, i Llm sono opachi: i modelli vengono sviluppati e gestiti da attori privati che non rendono pubblici i *dataset*, le metodologie di filtraggio dei dati, i criteri di moderazione, nonché gli algoritmi di *fine-tuning*. Ciò induce alla conclusione che dati e informazioni appartengono a coloro che, effettivamente, hanno la capacità di sostenere economicamente l'acquisto, il coordinamento e l'analisi di queste tecnologie³⁵.

In secondo luogo, le *hallucinations* plasmano le condizioni affinché gli utenti abbiano una derealizzazione della realtà³⁶, un esonero cognitivo (*Entlastung*) interpretato come un'agevolazione che accorda l'illusoria possibilità di superare i dubbi senza dilapidare eccessivamente le proprie energie³⁷. Del resto, l'inesauribile disponibilità di contenuti informativi nel *web*, spesso incoerenti fra di loro, inasprisce la vulnerabilità cognitiva³⁸ degli utenti, i quali inizieranno a fidarsi ciecamente degli *output* dell'IA per inerzia o per agevolare la funzione critica e di discernimento nel vaglio dei contenuti informativi.

Sergio Cotta (1920-2007), già negli anni Sessanta del secolo scorso, aveva intuito come

la straordinaria proliferazione dell'artificiale tende a far diventare l'uomo un “*agente di secondo grado*”, che anima e dirige macchine agenti direttamente; [...] il vivere in un mondo di artefatti influisce in maniera rilevante sulla psicologia dell'uomo; [...] l'artefatto oggi non è più soltanto il prodotto di un processo di modificazione della forma del naturale, bensì di una modificazione della sua sostanza³⁹.

5. Riflessioni conclusive

Il percorso argomentativo seguito in questo contributo – dopo aver introdotto gli aspetti teorici della Società algoritmica – ha messo a fuoco gli aspetti positivi e le criticità dei Llm nel dibattito pubblico-informativo. L'immanente

35. L. Avitabile, *Diritti umani e IA*, in A. Lalli (a cura di), *La regolazione pubblica delle tecnologie digitali e dell'intelligenza artificiale*, Giappichelli, 2024, pp. 1-14.

36. G. Vattimo, *La società trasparente* (1989), Garzanti, 2011, pp. 82-100.

37. Cfr. T. Casadei, *Istituzioni e algoritmi. Tra strategie funzionali ed “effetti collaterali”*, in U. Salanitro (a cura di), *SMART. La persona e l'infosfera*, Pacini Giuridica, 2022, pp. 239-260.

38. Per questa tematica sia concesso il rinvio a C. Coniglione, *Falsità delle informazioni e comunicazione in rete: le fake news*, in T. Casadei, V. Barone, B. Rossi (a cura di), *Giovani in rete. Guida per un uso consapevole delle tecnologie*, Giappichelli, 2025, pp. 101-112, in part. pp. 106-108.

39. S. Cotta, *Il giurista e la società in trasformazione*, in *Iustitia*, vol. XIX, 1966, p. 286.

opacità dei Llm e il ruolo “acritico” degli utenti nel recepimento delle informazioni suggeriscono la possibilità di dare spunti di riflessione informatico-giuridici sui Llm e filosofico-giuridici sul comportamento degli utenti.

Seguendo la prospettiva squisitamente informatico-giuridica, per comprendere la dimensione normativa connessa all’uso dei *Large Language Models* nel *fact-checking* e il contrasto alla disinformazione è fondamentale soffermarsi, seppur brevemente e nelle sue linee essenziali, sui regolamenti europei: il *Digital Service Act* (Dsa – 2022/2065) e l’*Artificial Intelligence Act* (AI Act – 2024/1689).

Com’è noto, il Dsa ha introdotto obblighi sia per le *Very Large Platforms* (Vlop) sia per i *Very Large Online Search Engine* (Vlose), imponendo la valutazione e la mitigazione dei rischi sistemici, la predisposizione di procedure trasparenti e la collaborazione con istituzioni indipendenti per il *fact-checking* (art. 33 Dsa). Siffatti vincoli si rendono necessari in un contesto dove l’IA generativa pare incidere significativamente sulla formazione dell’opinione pubblica, minacciata sempre di più da pratiche discorsive conflittuali⁴⁰ nel *web* in cui si perpetua una forma estrema di individualismo narcisistico⁴¹.

L’*AI Act*, dal canto suo, introduce specifici requisiti per i sistemi di IA e, in modo particolare, per quelli generativi. Fra gli obblighi assumono un particolare rilievo: l’etichettatura dei contenuti generali; la trasparenza sui dati e le modalità di addestramento; l’implementazione di misure idonee volte a prevenire e mitigare l’uso abusivo dei sistemi, inclusa la diffusione automatizzata di *fake news*.

Teoricamente i Llm fanno parte della categoria a rischio limitato, di conseguenza sono previsti alcuni obblighi di trasparenza e informazione (art. 52 *AI Act*). Ciò nondimeno, se usati con modalità che rischiano di pregiudicare i diritti fondamentali (come nel caso emblematico del dibattito pubblico-informativo), rientrano sicuramente nella categoria ad alto rischio. Sussistono, in questa eventualità, maggiori obblighi per i fornitori, dal momento che devono garantire la supervisione umana, correttezza e trasparenza, spiegabilità, interoperabilità, giustificabilità e riproducibilità (art. 15 *AI Act*). Le qualità rappresentate faciliterebbero la comprensibilità degli utenti nella conoscenza del funzionamento di un determinato modello di IA⁴².

40. Cfr. P. Savarese, *Appunti per una logica dei diritti umani*, Aracne, 2006, pp. 9 e 183.

41. E. Pulcini, *La cura del mondo. Paura e responsabilità nell’età globale*, Bollati Boringhieri, 2009, p. 12.

42. G. Ziccardi, *Una lettura dell’Artificial Intelligence Act: norme, etica, adempimenti, attuazione*, in F. Basile, M. Biasi, L. Camaldo, G. Caneschi, B. Fragasso, D. Milani (a cura di), *Intelligenza artificiale. Diritto, giustizia ed etica*, Giappichelli, 2025, pp. 22-24. Su questi prin-

Diventa fondamentale la comprensione dell'elaborazione degli *output* dell'IA da parte degli utenti, così costoro potrebbero sviluppare – in una prospettiva squisitamente informatico-giuridica – la funzione critica e di giudizio nei confronti dei contenuti informativi presenti nella rete, in modo particolare per quelli realizzati dall'IA generativa.

Un ruolo decisivo nei sistemi di IA generativa (e quindi nei Llm), se rientranti nell'ambito ad alto rischio, è assolto dalla *cybersecurity*: per evitare, o perlomeno attenuare, i *bias* è necessario stabilire rigorosi requisiti di robustezza e sicurezza nei sistemi di IA (art. 15 *AI Act*), posto che l'impiego di dati d'addestramento avvelenati realizza significative distorsioni nei Llm con inevitabili pregiudizi negli *output*.

In quest'ultima prospettiva, fra gli attacchi a cui i Llm sono più vulnerabili c'è il *data poisoning* (avvelenamento dei dati d'addestramento), l'*adversarial attack* (attacchi antagonisti), il *model extraction* (estrazione del modello) e il *membership inference*⁴³.

Promuovere la cultura della *Cybersecurity*, come argomentato da Giuseppe Corasaniti, esprime la «definizione (sintetica quanto concreta) di nuove forme di protezione possibili di quanto abbiamo più caro, dei nostri interessi, dei dati che ci sono indispensabili per vivere, delle risorse che approntiamo, delle condizioni stesse alla base di ogni forma di convivenza sociale»⁴⁴.

L'ultimo spunto di riflessione è giusto dedicarlo agli utenti, poiché essi sono più a contatto con le *hallucinations* dei Llm.

Oltre la prospettiva informatico-giuridica, una buona pratica potrebbe essere quella di incentivare la consapevolezza⁴⁵ che i Llm non sono affatto in

cipi anche L. Palazzani, *Etica della regolazione dell'intelligenza artificiale*, in *Rivista di filosofia del diritto*, n. 1, 2025, pp. 9-20, in part. pp. 18-15.

43. Per la puntuale descrizione di questi attacchi e dei suoi effetti nei sistemi informativi e di IA si rinvia alle riflessioni di G. Fioriglio, *Intelligenza artificiale e cibersicurezza: profili informatico-giuridici, fra vulnerabilità delle macchine e delle persone, in i-lex. Rivista di Scienze Giuridiche, Scienze Cognitive e Intelligenza artificiale*, vol. 18, n. 2, 2025, pp. 1-16, in part. pp. 7-8.

44. G. Corasaniti, *Sicurezza informatica e intelligenza artificiale. Rischio e resilienza nello spazio giuridico europeo*, Giappichelli, 2025, p. 4.

45. Come giustamente messo in evidenza da Lucia Corso, l'IA è «capace di condurre a un modello politico sociale caratterizzato da due fenomeni: i margini di trasgressione della regola vengono ridotti al minimo, e le persone sono via via esonerate dal gravoso compito di interpretare le norme giuridiche (e forse anche di formularle) [...]. Non c'è spazio per riflettere sugli effetti sull'umanità di un meccanismo normativo che via via esoneri i destinatari della regola dal dire la loro (ricordiamo il banchetto di Aristotele); e su un'umanità progressivamente esentata dal farsi domande sui valori fondanti della propria comunità politica quando anche l'interpretazione costituzionale verrà demandata al Chief Justice Robot» (L. Corso, *Legge dell'algoritmo e rule of law. Riflessioni preliminari*, in *PasSaggi Costituzionali*, n. 1, 2024, p. 213).

grado di comprendere le loro rappresentazioni e la realtà effettiva⁴⁶, quindi gli utenti dovrebbero coltivare la capacità di ricercare la molteplicità delle fonti informative con spirito critico.

La chiave di lettura che si sta tentando di rappresentare è quella di un dialogo costruttivo fra la filosofia del diritto, l'informatica giuridica, le scienze dell'informazione e le neuroscienze cognitive, per fissare principi di *governance* efficaci ed efficienti in stretta osservanza al rispetto dei valori etici e dei principi giuridici fondamentali⁴⁷. Le connessioni con questi campi della conoscenza dà la possibilità di restare ancorati «all'esperienza comune e a esigenze sociali fondamentali e consente di non perdere il contatto con la realtà»⁴⁸.

Questa metodologia – che segue volutamente un approccio che va dalla teoria alla prassi – può tutelare gli utenti da nuove forme sistemiche di “obbedienza cieca”⁴⁹, scongiurando in questo modo il rischio di un'eccessiva fiducia e di un'incondizionata accettazione degli *output* dell'IA generativa⁵⁰.

La ricerca delle informazioni, il confronto e il dibattito degli utenti è – e dovrà continuare a essere – la manifestazione della volontà, della consapevolezza e della comprensione del contesto⁵¹ per il fatto che, per dirlo *à la* Angelo Ermanno Cammarata (1899-1971), sarebbe «filosoficamente assurdo parlare di una coercizione della volontà, sia codesta coercizione “materiale” sia “psicologica”, perché l'essenza della volontà è la libertà»⁵².

Ragionare diversamente avrà come risultato quello dell'accettazione passiva degli *output* dei *Large Language Models*, che si traduce in una disponibilità a subire le azioni altrui⁵³ con conseguenziali forme di polarizzazione del dibat-

46. M. Cossutta, *Questioni sull'informatica giuridica*, Giappichelli, 2003, pp. 209-210.

47. Cfr. M. Farina, *Degenerazione e rischio creativo dell'Intelligenza Artificiale “forte”: forme di prevenzione e tutela complementare*, cit., pp. 137-138.

48. F. D'Agostino, *Jus quia justum. Lezioni di filosofia del diritto e della religione*, Giappichelli, 2012, p. 5.

49. Doveroso il richiamo alla monumentale opera di H. Arendt, *La banalità del male. Eichmann a Gerusalemme* (1963), trad. it., Feltrinelli, 2023.

50. Cfr. G. Fioriglio, *La disobbedienza civile digitale nella Società algoritmica*, in *Calumet – Intercultural Law and Humanities Review*, n. 23, 2025, pp. 73-92, in part. pp. 88-89.

51. Per l'inquadramento filosofico-giuridico della tematica sono ancora attuali le argomentazioni di A. Catania, *Decisione e norma* (1979), a cura di V. Giordano, F. Mancuso, Castelvocchi, 2023.

52. A.E. Cammarata, *Scritti sul formalismo giuridico*, a cura di F. Modugno, L. Pace, Roma Tre-Press, 2023, p. 74. Oltre a ciò, per un'ottimale analisi della proposta filosofico-giuridica di Cammarata, si vedano le riflessioni di T. Serra, *A. Ermanno Cammarata. La critica gnoseologica della giurisprudenza*, Edizioni Scientifiche Italiane, 1998.

53. Cfr. B. Pastore, *Soggettività giuridica e vulnerabilità*, in O. Giolo, B. Pastore (a cura di), *Vulnerabilità. Analisi multidisciplinare di un concetto*, Aracne, 2018, pp. 127-145.

tito pubblico-informativo, la trasformazione del *Logos* in forme algoritmicamente sussumibili⁵⁴, la deresponsabilizzazione collettiva nonché forme di “servitù volontaria”, per meglio dire un decisivo indebolimento della libertà d’agire⁵⁵ del *dēmos* nell’attuale Società algoritmica.

54. Martin Heidegger (1889-1976) comprese molto bene che «Con il dominio incondizionato della tecnica moderna si accresce la potenza – la pretesa come pure la prestazione – del linguaggio [...] Se l’informazione, nel senso di dominio in una tecnica che determini tutto quanto, viene ritenuta la forma suprema del linguaggio, a motivo della sua univocità, della sua sicurezza e della sua velocità nella comunicazione di notizie e di comandi, da ciò risulterà pure una concezione corrispondente dell’essere umano e della vita umana» (M. Heidegger, *Linguaggio tramandato e linguaggio tecnico* (1989), trad. it., Ets, 1997, p. 52).

55. Importanti, in quest’ultimo senso, le argomentazioni di T. Menissier, *La servitù volontaria nella società algoritmica*, in *Filosofia politica*, 1, 2022, pp. 85-100.

THE STATE OF THE ART LEGAL AND TECHNICAL OF AUTOMATED FAKE NEWS MODERATION SYSTEMS

*Camilla Scarpellino**

SOMMARIO: 1. Introduction to disinformation and legal framework – 2. The new medium of mass disinformation – 3. Disinformation governance under the Digital Services Act – 4. Conclusion.

The Internet has exponentially expanded the quantity and diversity of accessible content, allowing anyone with a computer to instantly and freely access information shared across the network. Yet this abundance has generated a paradox: a constantly informed population that is, nonetheless, increasingly susceptible to deception. Unlike traditional media, which are bound by obligations of accuracy and completeness, online platforms are not subject to equivalent regulatory duties. Indeed, the Digital Services Act (Dsa) reaffirmed the absence of a general monitoring obligation for hosting providers with respect to user-generated content.

The rise of generative artificial intelligence has further intensified this challenge by enabling the rapid and inexpensive production of false information and its simultaneous dissemination across multiple websites, often accompanied by fabricated supporting sources. The exponential growth of this phenomenon is partly attributable to online platforms and their business models, which foster the creation of echo chambers where repetition amplifies the visibility of falsehoods.

In response, the European Union has promoted a series of soft-law instruments aimed at enhancing algorithmic transparency in moderation systems adopted by major platforms such as Facebook, Instagram, X, TikTok, Amazon, and Google services. This article examines the Code of Practice on Disinformation, identifies the primary actors involved in its dissemination, and evaluates the governance mechanisms implemented by platforms to counter

* Assegnista di ricerca in Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

this phenomenon and their performance. These mechanisms generally rely on a hybrid model combining automated content screening with subsequent human or institutional interventions, including referral to third-party assessors or direct removal of the content by platforms and search engines.

1. Introduction to disinformation and legal framework

One of the central functions, and dysfunctions, of the World Wide Web lies in the vast and uncontrolled daily upload of information. Since anyone can create a website and publish content, the internet has revolutionized communication by enabling messages to reach both targeted individuals and global audiences. Indeed, the main function of Internet is to provide every kind of information freely and from everywhere there is a connection. Conversely, the absence of quality control over online content facilitates the rapid spread of false information. This phenomenon reached the scale of an «infodemic»¹ during the Covid-19 health crisis, fueled by conspiracy theories and fake news concerning the origins and treatment of the disease.

The current governance of the internet largely rests on twenty-six words of the “Good Samaritan”² clause in Section 230 of the Communications Decency Act, which shields Internet Service Providers (Isp) from liability for third-party content published on their platforms. The purpose of this liability exemption was «to promote the continued development of the Internet and other interactive computer services»³. A similar approach was adopted in the European Union, first through Directive 2000/31/EC and more recently in the Digital Service Act (Dsa), both of which exempt Isp from a general obligation to monitor user-generated content. However, the regulation also imposes a duty to act on these actors when they receive orders from competent authorities or notices from users or trusted flaggers, thereby ensuring that providers are deemed to have knowledge or awareness of the illegal content.

The Recitals of the Dsa place disinformation within the broad category of illegal content, yet the Act does not establish specific measures to address the spread of fake news. This omission may be attributed both to the difficulty of defining and delineating the phenomenon and to the legislative choice

1. References of the world “infodemic” available at: accademiadellacrusca.it/it/parole-nuove/infodemia/19506.

2. J. Kosseff, *The Twenty – Six Words That Created the Internet*, Cornell University Press, 2019, p. 9.

3. Communications Decency Act, 47 U.S.C. §230.

to craft the Dsa as a general framework applicable to all digital services. Although the Dsa introduces enhanced obligations concerning advertising transparency, transparency reporting, risk assessment, and mitigation, it does not prescribe a set of practical or detailed measures to be implemented. This gap may be addressed through codes of conduct, which operate as «steering instruments»⁴ capable of transforming the Dsa into a hybrid legal framework, where stakeholders develop codes of conduct with commitments to specific risk-mitigation measures, accompanied by regular reporting obligations on both the measures adopted and their outcomes.

One of the most significant gaps of Dsa concerns disinformation and misinformation, which remain largely unaddressed by both national and EU legislation, despite constituting a pressing threat to democratic debate and electoral processes. To bridge this gap, voluntary signatories elaborated and adopted the Code of Practice on Disinformation, now incorporated into the Dsa framework. The Code defines the notion of «Disinformation» excluding from the case «misleading advertising, reporting errors, satire and parody, or clearly identified partisan news and commentary, and is without prejudice to binding legal obligations, self-regulatory advertising codes, and standards regarding misleading advertising»⁵. Furthermore, it establishes concrete measures to enhance transparency, particularly in the context of political and issue advertising. Signatories are required to label such content, whether published on homepages or disseminated through messaging services, to enable users to recognize it and to ensure disclosure of the sponsor's identity. In addition, signatories must verify the accuracy of this information and implement proportionate and appropriate identity verification systems for sponsors and for advertising service providers acting on their behalf. Finally, they should also maintain repositories of political or issue advertising and allow users and researchers to access those data.

2. The new medium of mass disinformation

Social media platforms (Smp) have evolved into primary spaces for social interaction. However, platforms such as Facebook, Twitter, Snapchat, and Instagram are increasingly used to watch short videos or consume news,

4. R. Griffin, *Codes of Conduct in the Digital Services Act*, in *Technology and Regulation*, 2024, p. 169.

5. Cite note n. 5 of the *Code of Conduct on Disinformation*, available at: digitalstrategy.ec.europa.eu/en/library/code-conduct-disinformation.

functioning more as entertainment or information services than as venues for interpersonal exchange. Users' expectations of these platforms are therefore shifting from viewing them as spaces for social interaction to perceiving them as content providers or intermediaries⁶. This perceived "socialness" of Smp is shaped not only by their functions but also by design features such as homepage structure and content-sharing policies; for instance, which content are allowed to be shared, pictures, videos and/or texts. However, the perceived nature of Smp does not change its legal qualification and its liability regime.

At the same time, social media platforms do not merely facilitate interactions among users through messaging services; they also shape the visibility and sequencing of the content users encounter. Recommendation algorithms curate information flows by prioritizing material that aligns with a user's preferences, the tastes of their friends, or even the behaviors of strangers exhibiting similar patterns of engagement. In doing so, platforms actively construct personalized content environments rather than serving as neutral intermediaries. In fact, Facebook's algorithm was programmed to favor the visibility of content with more likes and visualization, to select specifically all contents that keep users engaged in their spaces. Algorithms tend to prioritize content that invites active social engagement, particularly posts generating a high volume of comments. However, such content often revolves around polarizing topics, including political discourse or misleading narratives.

This dynamic helps explain why false news spreads up to six times faster than verified information, deceptive or sensational content is deliberately crafted to provoke heightened emotional responses, such as outrage or relief, thereby amplifying its diffusion across networks. On the other hand, new tools available to content creators have become increasingly sophisticated. From large language models (Llm) to AI systems capable of generating realistic images and videos, these technologies enhance the production of misleading yet seemingly credible content. Moreover, many social media accounts are now partially or fully automated, programmed to reshare posts containing specific keywords, follow like-minded accounts, or reply to comments to simulate engagement. Such bots constitute powerful instruments for amplifying the reach and visibility of low-credibility information⁷.

The other dimension of Smp's automated management systems is content moderation, which is intended to remove or demote material that violates

6. L. Rhee, J.B. Bayer, D.S. Lee, O. Kuru., *Social by Definition: How Users Define Social Platforms and Why It Matters*, in *Telematics and Informatics*, 2021, p. 59.

7. H. Sun, *Regulating Algorithmic Disinformation*, in *Columbia Journal of Law & the Arts*, 2023, p. 374.

the platforms' Terms of Service (ToS) or community standards. This filtering process now extends to illegal content, defined as information, products, or services deemed unlawful under EU or national law, effectively delegating to private entities the responsibility of addressing legal infringements and balancing freedom of expression against the societal values threatened by such content. Notably, false speech does not enjoy constitutional protection in Italy, whereas in the United States it remains encompassed within the broad category of free expression safeguarded by the First Amendment⁸.

In this context, Smp have incorporated provisions on misinformation into their community standards, defining it as misleading content disseminated without intent to deceive. For example, Meta's demotion policy targets disinformation only when it can influence electoral processes or inciting physical violence, and when it detects inauthentic behavior of fake accounts. The company explicitly relies on independent third-party experts to evaluate both the accuracy of content and the potential risks it poses to public or private safety⁹. By contrast, X applies a narrower definition of illicit or misleading content, removing only false information concerning the procedures or requirements necessary to participate in civic processes¹⁰. The platform also targets inauthentic material, including synthetic or manipulated media, as well as coordinated inauthentic activity designed to influence user engagement, such as the use of multiple accounts interacting on identical or closely related topics to inflate or manipulate the prominence of content and/or accounts¹¹.

Similarly, TikTok «want[s] you to feel confident that what you're watching is reliable, original, and shared by real people»¹². Thus, to evaluate the accuracy of content, the platform collaborates with independent fact-checkers, local experts and authorities, and refers to a database of verified claims¹³. Community standards include an example list of misinformation

8. Si v. C. Brannon, *False Speech and the First Amendment: Constitutional Limits on Regulating Misinformation*, accessed on 1th October 2025, [www.congress.gov/crs-product/IF12180 - %3A~%3Atext=The Supreme Court has said,without violating the First Amendment%20said,without%20violating%20the%20First%20Amendment](http://www.congress.gov/crs-product/IF12180-%3A~%3Atext=The%20Supreme%20Court%20has%20said%20without%20violating%20the%20First%20Amendment).

9. Cite Meta's Transparency Service, *Misinformation*, in *Community Standards*, accessed on 4th October 2025, transparency.meta.com/it-it/policies/community-standards/misinformation/.

10. Cite. X Help Center, *Civic integrity policy*, in *the X Rules*, accessed in April 2025, available at: help.x.com/en/rules-and-policies/election-integrity-policy.

11. Cite. X Help Center, *Authenticity*, in *the X Rules*, accessed in April 2025, available at: help.x.com/en/rules-and-policies/authenticity.

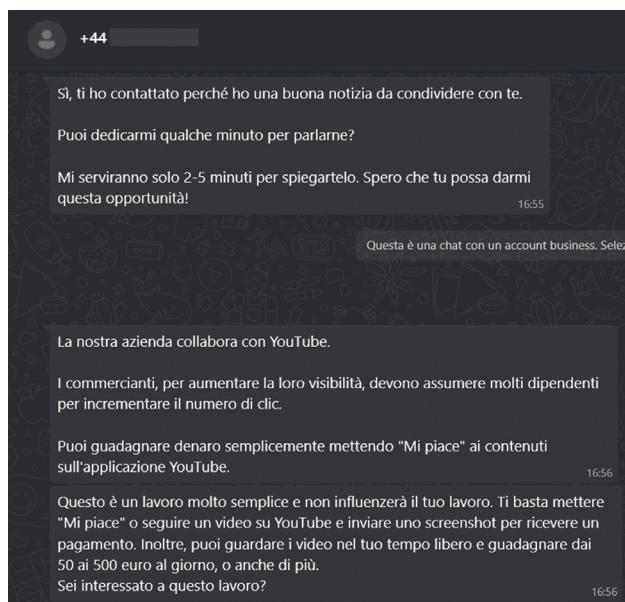
12. Cite TikTok, *Community Guidelines*, accessed on 4th October 2025, available at: www.tiktok.com/community-guidelines/en.

13. TikTok, *Centro Trasparenza*, accessed on 4th October 2025, available at: www.tiktok.com/transparency/it-it/combating-misinformation#GFCP/?lang=en.

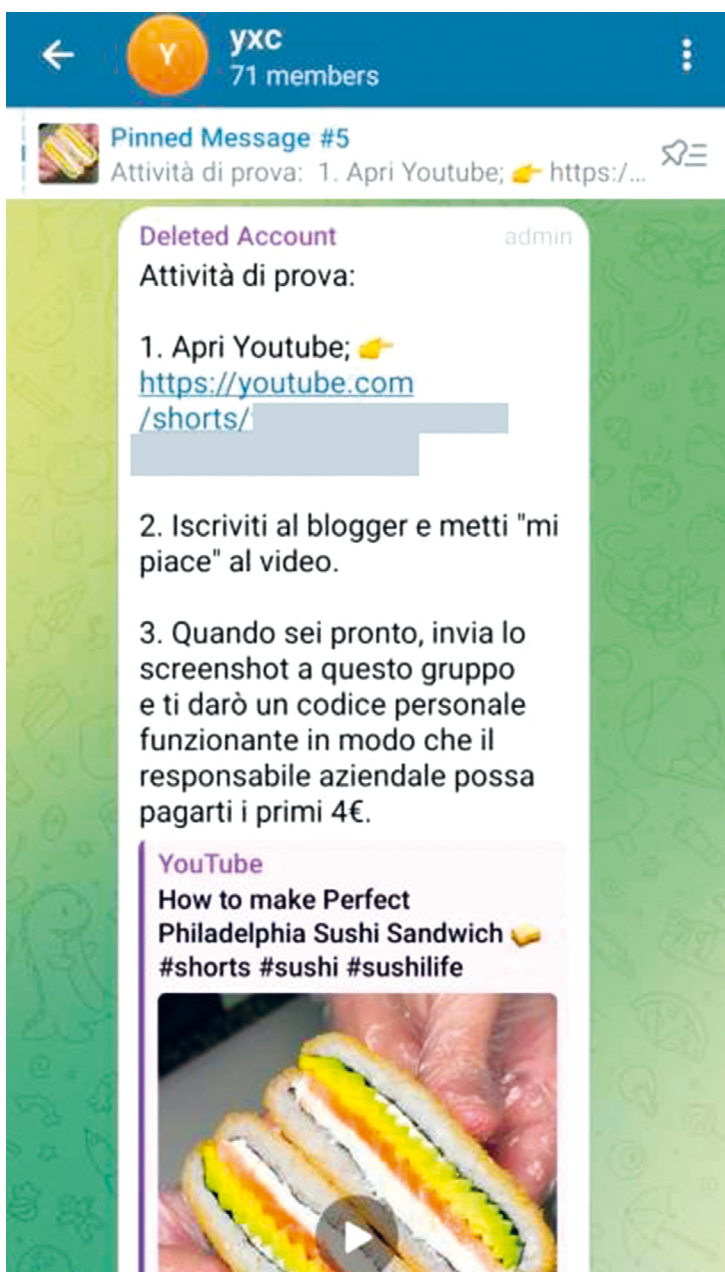
able to cause significant harm to individuals or society, it encompasses health-related falsehoods, conspiracy theories, manipulated media, distorted use of authoritative sources, and unverified statements concerning emergencies. TikTok, for its part, prohibits political advertising except in limited, regulated circumstances, and imposes strict requirements on accounts associated with public officials or political entities. The platform also mandates clear labeling of realistic AI-generated content to prevent confusion with authentic material. Furthermore, it bans practices intended to artificially amplify engagement or manipulate recommendation systems, such as mass account creation, bulk commercial posting, or deceptive techniques to increase visibility.

Disinformation also extends to e-commerce platforms such as Amazon, where false statements often take the form of manipulated product or seller reviews aimed at artificially inflating or deflating ratings. Amazon's policies are designed to safeguard the integrity of consumer feedback¹⁴; however, reviews are frequently fabricated, purchased, or coerced through deceptive practices, including scams and private solicitations such as those illustrated in *Fig. 1*.

Fig. 1 – Private chats in which users are persuaded to post likes and comments in exchange for payment



14. Amazon, *Community Guidelines*, accessed in April 2025, available at: www.amazon.com/gp/help/customer/display.html?nodeId=GLHXEX85MENUE4XF.



YouTube's misinformation policy largely mirrors that of X, prohibiting the publication of videos containing false information about electoral processes, as well as manipulated or misattributed content. Beyond the general

category of harmful material, the platform specifically targets misleading or deceptive information related to elections and medical issues that contradict guidance issued by local health authorities¹⁵. YouTube's strategy for addressing disinformation relies primarily on removing content that clearly violates its Community Guidelines, while content deemed "borderline" is not promoted through recommendation systems. The platform also prioritizes information from authoritative sources and rewards creators that consistently produce reliable material. However, as happened for Meta in U.S. and X worldwide, YouTube does not engage external fact-checking organizations in its verification processes¹⁶. Meta and Google have declared to stop verifying with fact-checkers networks content released in their spaces, thus it is plausible that trend pushed the EU Commission have integrated the Code of Conduct on Disinformation within the Dsa legal framework, to bind those platforms to moderation and verifying activity.

The comparative analysis of major Smp's Community Standards identifies two main models of platform-level disinformation governance. Meta and TikTok adopt content-based frameworks that define disinformation through evolving illustrative categories, while

X focuses on detecting the automated and coordinated activities behind its dissemination, such as automated bots and spam systems. In contrast, Amazon and Google regulate disinformation indirectly through advertising and commercial integrity policies, with Google

additionally restricting synthetic media and political misrepresentation. These differing strategies illustrate divergent institutional understandings of platforms' responsibility for information integrity.

The absence of a general duty to monitor content posted by third parties was initially justified by the limited material control that social media platforms could exert over the vast volume of content shared daily. However, the development of recommendation and moderation systems has fundamentally altered this rationale. These technologies actively shape and personalize each user's homepage, predominantly displaying information that reinforces individual or community biases, thereby transforming platforms from passive hosts into active curators of content. This selective shaping of information may amount to a form of editorial activity over users' homepages and feeds.

15. YouTube Centro Assistenza, *Politiche di disinformazione*, accessed in October 2025, available at: support.google.com/youtube/answer/10834785?hl=en#zippy=%2Cmisattributed-content.

16. L. Veronese, *Google rifiuta le regole UE sul fact-checking*, in *Il Sole 24 Ore*, 16th January 2025, www.ilsole24ore.com/art/google-rifiuta-regole-ue-fact-checking-AGMO5QLC.

Unlike traditional media services, which distribute identical content to all viewers, Smp present each user with a distinct set of topics and narratives. Such personalization renders them less akin to entertainment providers or spaces for social interaction, and more comparable to matching or profiling services, that curate content based on individual behavioral patterns¹⁷.

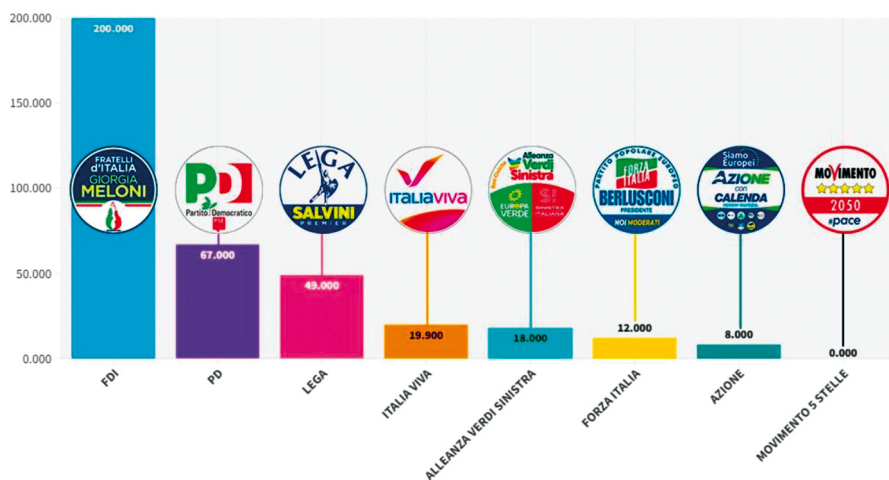
This unidirectional flow of communication can convey misleading messages very effectively, as such content is directed toward users who have already expressed interest in similar material, and are therefore more inclined to believe and reshare it. A similar dynamic applies to generative AI systems, such as ChatGpt or Google's AI Overview, which can inadvertently serve as channel for misinformation. In the absence of systematic verification mechanisms, the summaries provided by the search engine may report inaccurate or misleading information. Moreover, these summaries often fail to adequately disclose their sources, making it increasingly difficult even for the most conscientious users to assess the reliability or credibility of the underlying material.

3. Disinformation governance under the Digital Services Act

Very Large Online Platforms (Vlop) cover a dominant position within the digital information and communications ecosystem. To ensure that this market remains open to competition and new entrants, hence the EU Commission introduced the Digital Markets Act (Dma). Furthermore, these platforms hold a structurally stronger position in their contractual relationships with users, who must accept ToS and Community Standards that define the criteria and rules guiding algorithmic operations. As a result, social media platforms have evolved into de facto public forums for discussion, given that nearly everyone maintains one or more accounts. With their capacity to reach a global audience far beyond the limits of traditional media, the internet has effectively become the central space for contemporary public discourse. Politicians now disseminate their statements primarily through Smp rather than traditional news outlets, conducting much of their electoral campaigning through smartphone screens and investing heavily in sponsored content on these platforms.

17. U. Ruffolo, *AI Act: gli effetti settoriali – Artificial Intelligence Act e AI generativa*, in *Giurisprudenza Italiana*, 2025, p. 438; Y.R. Du, *Personalization, Echo Chambers, News Literacy, and Algorithmic Literacy: A Qualitative Study of AI-Powered News App Users*, in *Journal of Broadcasting & Electronic Media*, pp. 246-273, available at: www.tandfonline.com/doi/full/10.1080/08838151.2023.2182787.

Fig. 2 – Data concerning the financing of paid advertisements in the thirty-day period prior to the European parliamentary elections among Italian political parties¹⁸



Given the role of Smp as spaces for public debate, both the Dsa and the Dma seek to enhance transparency and accountability in their governance. Article 15 of the Dsa, in particular, requires platforms to publish annual transparency reports detailing their content moderation activities, including notices received from authorities and trusted flaggers, as well as removals carried out on the providers' own initiative. These reports must also indicate the tools employed to detect and assess the content in question, together with the legal or policy basis on which such determinations were made. Providers of online platforms are also required to report the number of out-of-court dispute settlements reached and account suspensions imposed on users. In addition to these reporting obligations, platforms must inform users about the advertisements displayed on their homepages, disclosing the identity of the sponsor and the parameters used to determine the recipients of such advertisements. They are further obliged to explain the main parameters underlying their recommender systems and how these influence the prioritization or visibility of content.

Meta Platforms Inc. has established a comprehensive set of governance

18. P. Decrestina, *Quanto spende la politica per i social? Manifesti addio, i partiti scelgono i post: dai 200 mila euro di FdI ai 65mila del Pd*, in *Corriere della Sera*, 14 June 2024, available at: www.corriere.it/politica/24_giugno_14/manifesti-addio-ora-i-partiti-scelgono-i-post-dai-200-mila-euro-di-fdi-ai-65mila-del-pd-ecco-quanto-la-politica-ha-speso-per-i-social-in-campagna.shtml.

instruments, including its Community Standards, Advertising Standards, Commerce Policies, and Monetization Policies. Automated detection systems are deployed to identify, restrict, or remove content and accounts that breach these rules, while content assessed as posing a higher risk of harm is referred to human reviewers for further evaluation. The criteria used to identify potential violations include posting frequency, the speed of user interactions, and response patterns, signals often associated with automated or bot-driven behavior. In addition, Meta's systems monitor Url that match or resemble those previously linked to removed material, thereby reinforcing the detection of recurrent policy breaches. Finally, Meta also employs artificial intelligence (AI) to enhance the accuracy and efficiency of its moderation tools, whose Automation Overturn Rate amount to 4,1% for Instagram and Facebook combined¹⁹.

X's Transparency Report describes a comparable hybrid approach, combining heuristic methods with machine learning techniques such as natural language processing, image recognition, and related models. These systems are trained on a dataset of illicit content previously identified, enabling them to detect similar material proactively. Decisions regarding enforcement actions against specific posts or accounts may subsequently be referred to human reviewers, depending on the historical accuracy and reliability of the model's output. A heuristic approach implies that AI systems are employed to identify behavioral patterns typically associated with abusive or manipulative activity, using extensive datasets and labeled examples for training. Before deployment, the algorithms undergo performance testing and validation, which involves evaluating «a statistically significant sample and conducting item-by-item human review»²⁰ to ensure reliability and minimize false positives. Notably, in assessing the accuracy of its moderation systems, X reports that their effectiveness is demonstrated by the number of content removal complaints received and the proportion of cases in which moderation decisions were subsequently reversed. During the reporting period, the platform recorded 205.146 complaints out of 55.331.835 enforcement actions initiated on its own initiative, with a median response time of 1,49 hours. Of all the complaints received, 26.930 were overturned, meaning that 13,12% of automated decisions were incorrect²¹.

19. Meta, *Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Facebook*, 29 August 2025, available at: transparency.meta.com/reports/regulatory-transparency-reports/.

20. Cite XIUC Transparency, *Dsa Transparency Report – April 2025*, this report covers the content moderation activities of X's international entity Twitter International Unlimited Company (TIUC) under the Digital Services Act (Dsa), during the date range 1 October, 2024 to 31 March, 2025, available at: transparency.x.com/en/reports/dsa-transparency-report.

21. XIUC Transparency, *Dsa Transparency Report – April 2025*.

TikTok adopts a slightly different approach, it also combines automated and human moderation systems, however its automated tools operate at a pre-publication stage, during which uploaded content remains visible only to the creator. Moderation algorithm removes or suspends posts or accounts just when a clear policy violation is detected. However, when a degree of interpretive discretion is required, the content is labeled and forwarded to human reviewers for further assessment. Moderation algorithms can analyze images, videos, audio, and text to identify symbols or words classified as harmful or resembling previously detected abusive material. In addition, Llm and multimodal Llm are employed to enhance the detection of more subtle violations, such as misinformative or context-dependent content. Teams of experts in safety and local market dynamics are responsible for updating detection rules and training

both algorithms and human reviewers to recognize new forms of illicit content and abusive behavior. According to the TikTok Transparency Report, only 0.8% of decisions were overruled²².

YouTube represents another example of Smp, a video-sharing platform, whose initial content screening is conducted through automated systems continuously refined by moderator feedback. The vast majority of content on YouTube is reviewed automatically, while human moderators are engaged «to apply a more nuanced approach to content assessment»²³; however, the platform does not clarify the specific circumstances in which human intervention occurs. The accuracy of YouTube's moderation algorithms is measured through the Violative View Rate (Vvr), an estimate of illicit videos based on a representative sample of total views. According to the most recent data, the Vvr was 0,14% of all views, corresponding to 11.408.261 videos removed. Of these, only 3,1% (66.096 videos) were taken down for disinformation, while the majority (76%) were removed for spam, scams, or fraudulent activity.

Ultimately, Amazon also engages in moderation activities targeting a specific form of misinformation disseminated through customer reviews that «intentionally mislead customers by providing information that is not impartial, authentic, or relevant to the product or service»²⁴. To address the

22. TikTok Transparency Center, *TikTok's Dsa Transparency Report*, the Report contains information the reporting period included between January to June 2025, available at: www.tiktok.com/transparency/en/dsa-transparency.

23. Cite Google Inc, *EU Digital Services Act (EU Dsa) Biannual Vlose/Vlop*, 28 August 2025, available at: digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency.

24. Amazon, *Digital Services Act (Dsa): Amazon EU Store Transparency Report*, April 2025, the report contains information the reporting period of July 2024 to December 2024, available at: trustworthysshopping.aboutamazon.com/digital-services-act-dsa-amazon-eu-store-transparency-report.

scalability challenges posed by the vast number of products, accounts, and reviews hosted on the platform, Amazon employs automated tools capable of detecting and acting upon violations with a high degree of certainty, allowing for immediate enforcement without human intervention. With respect to the accuracy of its moderation systems and following the same evaluative logic adopted by other platforms, Amazon reported that out of 393.712 complaints received, 272.242 resulted in reversals, indicating an error rate of approximately 69.14%. Nevertheless, in the section addressing the accuracy of its moderation model, the platform asserted that it has «a high confidence rate that our automated tools operate as intended and minimize mistakes»²⁵. However, it is evident that such statements, when unaccompanied by clear indicators of accuracy or quantifiable error rates, fail to meet the transparency requirements set out in Article 15 of the Dsa²⁶.

4. Conclusion

The phenomenon of fake news has expanded exponentially in both scale and impact, posing serious threats to fundamental societal values such as the integrity of electoral processes, public health, and the strategy to the climate crisis. Yet, within the European regulatory framework, no specific provisions establish accountability for the harm caused by disinformation, nor do they delineate corresponding liability or compensation mechanisms²⁷. As a result, existing measures to counter disinformation remain largely superficial and ineffective, as no concrete action is taken to address, for example, the inaccuracy or opacity of moderation algorithms employed by platforms such as X and Amazon, which exhibited the poorest performance in terms of accuracy. On the one hand, sanctions under the Dsa are imposed when platforms repeatedly infringe the obligations set out in the Regulation.

On the other hand, the financial penalties, amounting to up to six percent of the infringer's annual global turnover, are not directed toward compensating affected users or communities.

25. *Ibidem*.

26. Article 15, lett. e, of the Digital Service Act include among mandatory information also «any use made of automated means for the purpose of content moderation, including a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error of the automated means used in fulfilling those purposes, and any safeguards applied».

27. N. Brutti, *Il blocco del social network X in Brasile: Corte Suprema versus Elon Musk*, in *La Nuova Giurisprudenza Civile Commentata*, 2025, p. 133.

Just as Smp are not held economically accountable for the inaccuracy of their moderation systems, they also remain insufficiently accountable from an oversight perspective. Under the Dsa, platforms are required to issue Statements of Reasons (SoRs) specifying the factual basis, legal grounds, and tools employed in restricting or removing content or accounts. In practice, however, these SoRs often lack clarity and detail, providing only fragmented references to the information mandated by the Regulation²⁸. As a result, an average user is unable to obtain a concrete and comprehensible explanation for the estriction imposed, undermining the principle of transparency that the Dsa seeks to ensure.

Fig. 3 – An example of a Statement of Reason in the Transparency Data Base²⁹

Statement of reason details: 6d39517b-abf8-42fd-9113-bd48692069b0

Platform name	Facebook
Received	2025-05-28 10:37:19 UTC
Visibility restriction	Removal of content
Facts and circumstances relied on in taking the decision	This content was determined to violate terms of service
Ground for Decision	Content incompatible with terms and conditions
Reference to contractual ground	This was a violation of "sections 3.2" of our terms of service
Explanation of why the content is considered as incompatible on that ground	This was DECISION_VISIBILITY_CONTENT_REMOVED because it violated DECISION_GROUND_INCOMPATIBLE_CONTENT.
Is the content considered as illegal?	N/A
Territorial scope of the decision	European Economic Area
Content Type	Image
When the content was posted or uploaded	2025-05-27
Category	Scope of platform service
Information source	Own voluntary initiative
Was the content detected/identified using automated means?	No
Was the decision taken using other automated means?	Not Automated
Application date of the decision	2025-05-27

28. M. Aspromonte, A. Ferraris, F. Galli, and G. Contissa., *Llm to the Rescue: Explaining Dsa Statements of Reason with Platform’s Terms of Services*, in *Proceedings of the Natural Legal Language Processing Workshop 2024*, edited by N. Aletras, I. Chalkidis, L. Barrett, C. Goanță, D. Preoțiuc-Pietro, G. Spanakis, 2024, Association for Computational Linguistics, pp. 205-207 available at aclanthology.org/2024.nllp-1.17/; A.T. Omowale, M.J. Owojuyigbe, F.D. Adeyemo, *New Media Versus Traditional Media: Power Dynamics And The Struggle For Credibility*, in *African Journal of Social and Behavioural Sciences*, 2025, p. 1420-1426, available at: journals.aphriapub.com/index.php/AJSBS/article/view/3159.

29. Figure extracted from the *Dsa Transparency Database*, accessed in October 2025, transparency.dsa.ec.europa.eu/statement.

In light of the societal values endangered by disinformation, the private interests that shape moderation algorithms, and the persistent lack of accountability among social media platforms, the oversight phase of the compliance evaluation process should not rely primarily on users' complaints. Instead, it ought to be entrusted to competent public authorities, such as the Digital Services Coordinator or the relevant body responsible for safeguarding the affected public interest, to ensure impartial supervision and effective enforcement. Authorities should be able to explore the logic, datasets and source code of moderation algorithms to review them with the contribute of governance experts, technicians, and Smp's users and representative bodies pursuing general or collective interests.

AFFRONTARE LA DISINFORMAZIONE ONLINE: IL SISTEMA DI MITIGAZIONE DEI RISCHI NEL *DIGITAL SERVICES ACT*

Sara Gallone*

SOMMARIO: 1. Introduzione: il fenomeno vecchio e nuovo della disinformazione – 2. Illegal vs legal but harmful – 3. La disinformazione come rischio sistemico – 4. Gli articoli 34 e 35 del Dsa: un'analisi al sistema di mitigazione dei rischi – 5. Conclusioni.

1. Introduzione: il fenomeno vecchio e nuovo della disinformazione

Per “disinformazione”, secondo la definizione fornita dall'*High Level Expert Group on Fake News and Online Disinformation*¹ e recepita dall'Unione europea nel preambolo al Codice di buone pratiche², si intende quell'insieme di contenuti deliberatamente creati e diffusi con l'intento di ingannare, per scopo di lucro o per arrecare un pregiudizio. Tuttavia, la diffusione di notizie false e manipolate non costituisce un fenomeno esclusivo dell'epoca contemporanea, ma si colloca in una più ampia tradizione storica di utilizzo dell'informazione quale strumento di potere e di influenza sulla società. Ne deriva che il dibattito sull'opportunità e sulle modalità di intervento regolatorio non è una questione nuova, ma un motivo ricorrente che accompagna l'evoluzione dei mezzi di comunicazione e delle relazioni sociali. Ma «se la tecnica della disinformazione è largamente attestata nella storia, il panorama è oggi cambiato in relazione agli strumenti che la supportano»³. Internet e i social network,

* Dottoranda di ricerca in Diritto dell'Unione europea, Alma Mater Studiorum – Università di Bologna.

1. High Level Expert Group on Fake News and Online Disinformation, *Report to the European Commission on a multi-dimensional approach to disinformation*, 2018, disponibile su: digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation.

2. European Commission, *EU code of practice on disinformation (Preamble)*, 2018, disponibile su: digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation.

3. P. Passaglia, *Fake news e fake democracy: Una convergenza da scongiurare*, in *federalismi.it*, n. 11, 2020, p. 128.

pur costituendo strumenti straordinari di accesso all'informazione e di partecipazione civica, offrono alla disinformazione possibilità di diffusione senza precedenti, rendendola più pericolosa per la democrazia e il dibattito pubblico rispetto al passato. La pervasività dei contenuti generati dagli utenti consente, infatti, a informazioni false o distorte di circolare rapidamente, attraverso semplici condivisioni o riproposizioni da parte e tra una platea potenzialmente illimitata di destinatari. La profilazione basata sui dati personali, spesso forniti inconsapevolmente dagli utenti, rende possibile una mirata segmentazione dei contenuti, veicolando informazioni specifiche a soggetti ritenuti, in base ai dati raccolti, più sensibili e predisposti alla loro ricezione. Inoltre, la creazione delle cosiddette *filter bubbles*, bolle informative generate dalla selezione algoritmica dei contenuti in funzione delle inclinazioni dell'utente, desunte dai suoi comportamenti online, produce un isolamento intellettuale che riduce lo spazio per la dialettica e favorisce la polarizzazione sociale, creando un contesto particolarmente propizio alla diffusione e al radicamento della disinformazione, anche di carattere sovversivo.

Ed è proprio alla luce di tali dinamiche che, in particolare a livello europeo, si è avvertita l'esigenza di elaborare strumenti capaci di contrastare il fenomeno della disinformazione, garantendo al contempo la tutela dei diritti fondamentali e, *in primis*, della libertà di espressione. La dimensione sovranazionale e transfrontaliera della diffusione dei contenuti tramite le piattaforme digitali spiega, del resto, perché l'intervento normativo sia avvenuto non su base nazionale, ma europea⁴.

In questo contesto, un ruolo centrale è svolto dal Regolamento sui servizi digitali, ovvero dal *Digital Services Act* (Dsa), entrato in vigore nel novembre 2022 e pienamente applicabile da febbraio 2024. Il Dsa si propone di «garantire un ambiente online sicuro, prevedibile e affidabile, in cui i diritti fondamentali sanciti dalla Carta siano efficacemente tutelati e l'innovazione sia agevolata, contrastando la diffusione di contenuti illegali online e i rischi per la società che la diffusione della disinformazione o di altri contenuti può generare»⁵.

La disinformazione viene così espressamente riconosciuta quale possibile rischio per la società, poiché compromette la fiducia nelle istituzioni e altera il confronto pubblico. Ne deriva la necessità di adottare misure di mitigazione.

Muovendo da tali premesse, il presente contributo si propone di riflettere

4. La base giuridica a fondamento di tale intervento normativo è individuata nell'art. 114 Tfeue nell'ambito della disciplina sul mercato interno. Si v. considerando n. 2 del Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali (*Digital Services Act*), disponibile su: eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065.

5. Ivi, considerando n. 9.

sul sistema delineato dagli articoli 34 e 35 del Regolamento, soffermandosi sulle principali questioni giuridiche che tali disposizioni sollevano: dalla distinzione tra i c.d. contenuti *illegal* e i c.d. contenuti *legal but harmful*, alla qualificazione della disinformazione come rischio sistemico, sino alle misure di mitigazione, tra cui figurano anche i codici di condotta. L'analisi intende mettere in luce le criticità interpretative e applicative di tale sistema, senza pretendere di offrire soluzioni definitive, ma piuttosto contribuendo a una maggiore consapevolezza delle difficoltà che esso pone.

2. Illegal vs legal but harmful

In primo luogo, occorre chiarire come viene inteso il fenomeno della disinformazione all'interno del Regolamento sui servizi digitali. Per scongiurare il rischio di sconfinare nella censura, vi è infatti una linea rossa⁶ da tracciare: distinguere tra i c.d. contenuti *illegal* e i c.d. contenuti *legal but harmful*. I primi sono tali in quanto violano una disposizione del diritto dell'Unione o degli Stati membri, come stabilito dall'art. 3, lett. *b*, Dsa⁷. Ne consegue che l'obiettivo del *Digital Services Act* non è quello di creare delle nuove regole in materia di contenuti⁸ bensì riaffermare il principio secondo cui «what is illegal offline, should be illegal online»⁹. La qualificazione di un contenuto come “illegale” resta, pertanto, demandata al diritto nazionale o dell'Unione.

Tuttavia, dall'art. 34 del Dsa emerge l'esistenza di un'ulteriore categoria di contenuti che, pur non essendo illegali, possono generare rischi sistemici. Tale disposizione introduce obblighi specifici per le piattaforme online e i motori di ricerca che registrano un numero medio mensile di utenti attivi nell'Unione pari o superiore a 45 milioni, imponendo loro di individuare, analizzare e valutare non solo i rischi connessi alla diffusione di contenuti illegali, ma anche quelli derivanti dalla circolazione di altri contenuti che possono avere effetti negativi, attuali o prevedibili, sull'esercizio dei diritti fondamentali nonché sull'integrità del dibattito civico, dei processi elettorali e della sicurezza pub-

6.M. Husovec, *The Digital Services Act's red line: What the Commission can and cannot do about disinformation*, in *Journal of Media Law*, vol. 16, n. 1, 2024, pp. 47-56.

7. Si v. considerando n. 12 e art. 3, lett. *b*, del Regolamento (UE) 2022/2065, cit.

8. M. Husovec, *General report: Topic II – EU digital economy: General framework (Dsa/Dma) and specialised regimes*, in K. Pacuła (ed.), *EU digital economy: General framework (Dsa/Dma) and specialised regimes*, vol. 2, Wydawnictwo Uniwersytetu Śląskiego, 2025, pp. 55-56.

9. Si v. comunicato stampa della Commissione europea, del 23 aprile 2022, *Digital Services Act: Commission welcomes political agreement on rules ensuring a safe and accountable online environment*, disponibile su: www.ec.europa.eu/commission/presscorner/detail/en/ip_22_2545.

blica. Il mancato adempimento di tali obblighi espone i fornitori a sanzioni pecuniarie che, ai sensi dell'art. 74, possono raggiungere il 6% del fatturato annuo mondiale realizzato nell'esercizio precedente¹⁰. Ciò vale nonostante, in linea di principio e con riferimento ai contenuti illegali, resti ferma l'assenza di un obbligo generale di monitoraggio¹¹ e l'esenzione di responsabilità previste dalla direttiva e-commerce 2000/31/CE¹².

Dunque, il Dsa utilizza, in contrapposizione ma anche in parallelo all'espressione "contenuti illegali", il termine contenuti «comunque dannosi»¹³, per indicare che esistono altri rischi derivanti dall'uso delle piattaforme online, in particolare quelle di grandi dimensioni, e che tali rischi devono essere identificati, valutati e adeguatamente mitigati. In tale contesto si inserisce il fenomeno della disinformazione che, ove non configuri anche un illecito (come, ad esempio, la propaganda bellica)¹⁴, è suscettibile di generare rischi sistemici, tanto in relazione alla tutela dei diritti fondamentali individuali, quanto rispetto agli effetti sul dibattito civico e sui processi elettorali¹⁵.

Tale distinzione, come rilevato dalla dottrina¹⁶, è cruciale poiché dalla qua-

10. Si v. art. 74 par. 1. del Regolamento (UE) 2022/2065 cit.

11. Ivi, art. 8 che sostituisce l'art. 15 della direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000 relativa a taluni aspetti giuridici dei servizi della società dell'informazione, in particolare il commercio elettronico, nel mercato interno ("Direttiva sul commercio elettronico"), disponibile su: eur-lex.europa.eu/legal-content/IT/ALL/?uri=celex:32000L0031.

12. Si v. artt. 4, 5 e 6 del Regolamento (UE) 2022/2065 cit. che sostituiscono gli artt. 12, 13 e 14 della direttiva 2000/31/CE cit. Tale esenzione, però, non opera quando il prestatore di servizi assume un ruolo attivo, andando oltre la mera funzione tecnica e neutrale, oppure quando, pur avendo conoscenza della presenza di contenuti illegali, omette di rimuoverli o di disabilitarne tempestivamente l'accesso.

13. Ivi, considerando nn. 5 e 68.

14. È importante osservare che il Dsa adotta un'accezione ampia del termine "contenuto illegale", come si evince dalla lettura dell'art. 3, lett. b, per cui «qualsiasi informazione che, di per sé o in relazione a un'attività [...] non è conforme al diritto dell'Unione o di qualunque Stato membro conforme con il diritto dell'Unione, indipendentemente dalla natura o dall'oggetto specifico di tale diritto». Inoltre, in alcuni Stati membri, come Malta e Cipro, la disinformazione costituisce reato quando arreca pregiudizio all'ordine pubblico o mina la fiducia nelle istituzioni. Si v. R.Ó. Fathaigh, N. Helberger, N. Appelman, *The perils of legally defining disinformation*, in *Internet Policy Review*, vol. 10, n. 4, 2021. In Italia, vi è la fattispecie contravvenzionale di cui all'art. 656 c.p. che, a tutela dell'ordine pubblico, punisce la diffusione di notizie false esagerate e tendenziose.

15. O. Pollicino, P. Dunn, & L. Violante, *Intelligenza artificiale e democrazia: Opportunità e rischi di disinformazione e discriminazione*, Bocconi University Press, 2024, p. 53.

16. M. Husovec, *The Digital Services Act's red line: What the Commission can and cannot do about disinformation*, cit., pp. 50-52. Si v. altresì M. Husovec, *General report: Topic II – EU digital economy: General framework (Dsa/Dma) and specialised regimes*, cit., p. 57; Zeno-Zencovich, *The EU regulation of speech: A critical view*, in *MediaLaws*, n. 1, 2023, pp. 175-184;

lificazione del contenuto come *illegal* o *legal but harmful*, discendono obblighi e conseguenze giuridiche differenti. Nei primi casi sono ammissibili misure *content-specific*, quali la rimozione del contenuto o la chiusura dell'account dell'utente che lo ha diffuso; nei secondi, invece, si ricorre a misure *content-neutral*, agendo ad esempio sulla trasparenza o sul rafforzamento del potere decisionale degli utenti. Il perimetro deve, come visto, essere segnato dal legislatore, nazionale o unionale, a cui solo spetta stabilire l'illiceità di un determinato contenuto. Il Dsa resta infatti «a procedural tool»¹⁷: l'idea di fondo è che, se le piattaforme moderano i contenuti generati dagli utenti, devono essere soggette a supervisione sulle modalità con cui operano¹⁸.

3. La disinformazione come rischio sistemico

Come evidenziato, la disinformazione, pur non costituendo necessariamente un illecito, può comunque assumere la dimensione di rischio sistemico. Sono i considerando 83¹⁹ e 84²⁰ ma soprattutto il considerando 104 a stabilire un collegamento in tal senso. Specialmente quest'ultimo, con riferimento ai codici di condotta delle piattaforme digitali, sottolinea la necessità di fare attenzione agli «eventuali effetti negativi dei rischi sistemici sulla società e sulla democrazia, quali la disinformazione o le attività di manipolazione e abuso o eventuali effetti avversi sui minori»²¹.

Ne consegue che la disinformazione non costituisce di per sé un rischio si-

J. Bayer, *The EU policy on disinformation: Aims and legal basis*, in *Journal of Media Law*, vol. 16, n. 1, 2024, pp. 18-27.

17. M. Husovec, *The Digital Services Act's Red Line: What the Commission Can and Cannot Do About Disinformation*, cit., p. 54.

18. J. Bayer, *The EU Policy on Disinformation: Aims and Legal Basis*, cit., p. 26.

19. Considerando n. 83 del Regolamento (UE) 2022/2065 cit.: «La quarta categoria di rischi deriva da preoccupazioni analoghe relative alla progettazione, al funzionamento o all'uso, anche mediante manipolazione, di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi, con ripercussioni negative, effettive o prevedibili, sulla tutela della salute pubblica e dei minori e gravi conseguenze negative per il benessere fisico e mentale della persona o per la violenza di genere. Tali rischi possono derivare anche da campagne di disinformazione coordinate relative alla salute pubblica o dalla progettazione di interfacce online che possono stimolare le dipendenze comportamentali dei destinatari del servizio».

20. Ivi considerando n. 84: «Nel valutare i rischi sistemici individuati nel presente regolamento, tali fornitori dovrebbero concentrarsi anche sulle informazioni che non sono illegali ma contribuiscono ai rischi sistemici individuati nel presente regolamento. Tali fornitori dovrebbero pertanto prestare particolare attenzione al modo in cui i loro servizi sono utilizzati per diffondere o amplificare contenuti fuorvianti o ingannevoli, compresa la disinformazione».

21. Ivi, considerando n. 104.

stemico, ma lo diventa qualora produca effetti dannosi per la società. Va però rilevato come il concetto di “rischio sistemico” presenti un’evidente indeterminatezza. Dalla lettera del Regolamento, infatti, non è chiaro quale sia la soglia oltre la quale un rischio diventi sistemico²² e si lascia alle stesse piattaforme digitali il compito di effettuare tale valutazione²³.

In linea generale, si può ritenere che un rischio assuma carattere sistemico quando risultano compromessi la stabilità e il regolare funzionamento del sistema di riferimento²⁴. Si tratta di un concetto già ampiamente esplorato in diverse discipline, dall’economia alla finanza, dalla sociologia alla filosofia. Anche il diritto dell’Unione europea non è estraneo a questa nozione, che costituisce infatti il fondamento di numerosi interventi regolatori, segnatamente nei settori dell’ambiente, della politica dei consumatori, della sicurezza alimentare e della tutela della salute umana, animale e vegetale²⁵.

Negli ultimi anni, inoltre, un approccio *risk-based* ha progressivamente informato anche la normativa europea in materia di protezione dei dati personali, cybersicurezza e intelligenza artificiale²⁶. Con riferimento a quest’ultimo ambito, anche il Regolamento sull’intelligenza artificiale (c.d. *AI Act*) condivide con il *Digital Services Act* un problema di fondo, ossia che pur individuando i settori potenzialmente interessati dal rischio sistemico, essi non forniscono definizioni puntuali degli eventi idonei a compromettere la stabilità e il funzionamento del sistema.

22. In questo senso, ivi, il considerando n. 79 si limita a riportare alcuni esempi, senza tuttavia individuare criteri chiari per stabilire quando un rischio possa qualificarsi come sistemico: «Nel determinare la rilevanza dei potenziali effetti e impatti negativi, i fornitori dovrebbero considerare la gravità dell’impatto potenziale e la probabilità di tutti questi rischi sistemici. Ad esempio, potrebbero valutare se il potenziale impatto negativo possa incidere su un gran numero di persone, la sua potenziale irreversibilità o la difficoltà nel porvi rimedio e ripristinare la situazione esistente prima dell’impatto potenziale».

23. Si v. A. Iliopoulou-Penot, *Governing the “Governors” of online speech: A review of the Digital Services Act’s first steps in regulating social media platforms*, in *Yearbook of European Law*, 2025, p. 14.

24. M. Loi, M. Fabbri, & A. Ferrario, *Regulating the Undefined: Addressing Systemic Risks in the Digital Services Act (with an Appendix on the AI Act): Regulating the Undefined: Addressing Systemic*, in *Philosophy & Technology*, vol. 38, n. 2, 2025, pp. 5-9.

25. Nel diritto europeo, il concetto di rischio è strettamente legato al principio di precauzione e alla gestione preventiva delle minacce potenziali per la salute, l’ambiente e il mercato interno. Questo principio è espressamente sancito in materia di politica ambientale all’art. 191 Tfe e prevede che, in caso di rischio potenziale per persone o ambiente e in assenza di consenso scientifico sufficiente, le autorità europee possano adottare misure di precauzione. Si v. art. 191 del Trattato sul funzionamento dell’Unione europea, C 202, 7 giugno 2016, disponibile su: eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A12016ME%2FTXT.

26. Sul punto, L. Califano, *La strategia normativa dell’Unione europea per un nuovo ordine digitale*, in *federalismi.it*, n. 15, 2025, p. 9.

Con riferimento al Dsa, in particolare, l'identificazione e la valutazione dei rischi è rimessa alle stesse piattaforme digitali²⁷ e «Regulators play the secondary role of overseeing whether these internal risk management systems are adequate»²⁸. Questo approccio è denominato “meta-regolazione”: “meta” perché un regolatore (macro) sovrintende un altro regolatore (micro) nella gestione del rischio; “enforced” perché, in caso di inadeguatezza delle pratiche di autoregolamentazione, il regolatore (macro) ha il potere di adottare misure coercitive²⁹. Secondo la dottrina³⁰, questo modello presenta diversi vantaggi, quali una maggiore flessibilità e una più accurata sensibilità al contesto rispetto a degli standard prefissati, rigidi e applicabili all'intero settore. Inoltre, esso consente ai regolatori di valorizzare le risorse tecniche e finanziarie delle stesse piattaforme, così come la loro conoscenza già acquisita in materia. Tuttavia, ciò comporta anche il rischio che tali poteri privati interpretino le disposizioni a proprio piacimento e secondo i propri interessi economici, erodendo così l'autorità del potere pubblico (c.d. corporate capture)³¹.

Sembra alquanto chiaro che in sede di redazione e approvazione del Dsa non si intendesse definire un approccio dettagliato e prescrittivo alla valutazione dei rischi, ma piuttosto creare un quadro di riferimento capace di accogliere obiettivi e priorità regolamentari diversi, e suscettibile di essere sviluppato e precisato nel tempo³². Va sottolineato, in via generale, che i rischi, soprattutto nel contesto delle scienze sociali, non sono concetti neutri: essi sono in larga misura influenzati dalle diverse agende politiche, dai valori e dagli interessi in gioco, e per questo risultano difficili da misurare o parametrare in modo oggettivo. Inoltre, il settore digitale è in continua evoluzione e prevedere delle regole prefissate a monte comporta una rapida obsolescenza della stessa normativa.

27. Il considerando n. 79 del Regolamento (UE) 2022/2065 cit cerca di dare delle indicazioni senza tuttavia risolvere: «Nel determinare la rilevanza dei potenziali effetti e impatti negativi, i fornitori dovrebbero considerare la gravità dell'impatto potenziale e la probabilità di tutti questi rischi sistemici. Ad esempio, potrebbero valutare se il potenziale impatto negativo possa incidere su un gran numero di persone, la sua potenziale irreversibilità o la difficoltà nel porvi rimedio e ripristinare la situazione esistente prima dell'impatto potenziale».

28. R. Griffin, *Governing platforms through corporate risk management: the politics of systemic risk in the Digital Services Act*, in *European Law Open*, vol. 4, n. 2, 2025, p. 9.

29. N. Zingales, *The Dsa as a paradigm shift for online intermediaries' due diligence: Hail to meta-regulation*, in *VerfBlog*, 2 November 2022, disponibile su: verfassungsblog.de/dsa-meta-regulation/. Per un approfondimento sul concetto di *meta-regulation*, C. Coglianese, E. Mendelson, *Meta-regulation and self-regulation*, in M. Cave, R. Baldwin, M. Lodge (eds.), *The Oxford handbook on regulation*, University of Pennsylvania, 2010, pp. 1-34.

30. R. Griffin, *Governing platforms through corporate risk management: the politics of systemic risk in the Digital Services Act*, cit., p. 10.

31. Ivi, p. 21.

32. Ivi, p. 13.

Infine e più pragmaticamente, spesso la formulazione ampia della norma è il risultato della necessità di raggiungere un compromesso politico, nell'equo contemperamento dei diversi interessi coinvolti³³.

Ciononostante, a parere di chi scrive, pur condividendo alcune critiche (v. *supra*), non va trascurato che il Dsa non conferisce alle piattaforme digitali completa libertà d'azione. Al contrario, contribuisce a creare un ecosistema in cui diversi attori, dalle autorità nazionali alla società civile e alla componente accademica, collaborano tra loro per «garantire un ambiente online sicuro, prevedibile e affidabile»³⁴. Inoltre, le piattaforme, tramite le proprie condizioni di servizio, erano già impegnate in pratiche di moderazione dei contenuti; il Dsa interviene ora per regolamentarle e per fornire strumenti chiari a tutela degli utenti. Ad esempio, con riferimento ai prestatori di servizi di memorizzazione di informazione³⁵, tra cui vi rientrano le piattaforme online³⁶, è previsto l'obbligo di predisporre meccanismi di segnalazione e azione, di facile accesso e utilizzo, che agevolino la notifica della presenza di contenuti illegali. Il prestatore può poi decidere se condividere o meno tale valutazione e, in caso positivo, procedere alla rimozione o disabilitazione dell'accesso³⁷, motivando l'azione intrapresa e garantendo la possibilità di presentare reclami avverso la medesima³⁸. La predisposizione di un sistema interno di gestione dei reclami è nello specifico previsto dall'art. 20 per quei prestatori di servizi di memorizzazione di informazione denominati «piattaforme online», ossia servizi di memorizzazione di informazioni che, su richiesta di un destinatario, memorizzano e diffondono informazioni al pubblico. Su quest'ultime ricadono altresì i costi di un eventuale ricorso agli organismi di risoluzione extragiudiziale delle controversie, ai fini della risoluzione delle dispute inerenti alle decisioni in materia di moderazione dei contenuti, compresi i reclami che non è stato possibile risolvere mediante il sistema di gestione interno³⁹.

Da ultimo, merita attenzione l'articolo 14 sui termini e condizioni d'uso che, sebbene possa apparire finalizzato a legittimare interventi *content-based* contro la disinformazione lecita, si configura piuttosto come uno strumento

33. *Ibidem*.

34. Considerando n. 9 del Regolamento (UE) 2022/2065 cit.

35. Ivi, ai sensi dell'art. 3 lett. g iii), un servizio di memorizzazione di informazioni (cosiddetto «*hosting*»), consiste nel memorizzare informazioni fornite da un destinatario del servizio su richiesta dello stesso.

36. Ivi, art. 3, lett. j, per cui una piattaforma online è un servizio di memorizzazione di informazioni che, su richiesta di un destinatario del servizio, memorizza e diffonde informazioni al pubblico.

37. Si v. ivi, considerando n. 50 e art. 16.

38. Ivi art. 17.

39. Ivi art. 21.

di *empowerment* degli utenti, imponendo maggiore trasparenza nell'adozione delle misure di moderazione dei contenuti online e la valutazione del loro impatto sui diritti fondamentali⁴⁰. Quest'ultima, del resto, deve orientare in generale l'elaborazione e l'applicazione delle misure di mitigazione dei rischi, con particolare attenzione alla tutela della libertà di espressione⁴¹.

4. Gli articoli 34 e 35 del Dsa: un'analisi al sistema di mitigazione dei rischi

Dopo aver distinto tra contenuto illegale e contenuto legale ma dannoso e aver delineato le principali difficoltà nella definizione di rischio sistemico, ci si soffermerà di seguito sul contenuto normativo degli articoli 34 e 35 del Dsa. La prima disposizione, come già scritto, richiede che i fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi individuino, analizzino e valutino con diligenza gli eventuali rischi sistemici che possono derivare dalla progettazione o dal funzionamento del loro servizio e dei suoi relativi sistemi, compresi i sistemi algoritmici, o dall'uso dei loro servizi. Salvo quanto previsto dall'art. 33 paragrafo 6⁴², tale valutazione va effettuata annualmente e deve essere specifica e proporzionata alla gravità e alla probabilità dei rischi sistemici. Tra quest'ultimi, oltre a quelli già menzionati come la diffusione di contenuti illegali e gli effetti negativi sui diritti fondamentali, sui processi elettorali e sulla sicurezza pubblica, vi rientra anche «qualsiasi effetto negativo, attuale o prevedibile, in relazione alla violenza di genere, alla protezione della salute pubblica e dei minori e alle gravi conseguenze negative per il benessere fisico e mentale della persona»⁴³. In tal senso,

40. Ivi art. 14 par. 4: «I prestatori di servizi intermediari agiscono in modo diligente, obiettivo e proporzionato nell'applicare e far rispettare le restrizioni di cui al paragrafo 1, tenendo debitamente conto dei diritti e degli interessi legittimi di tutte le parti coinvolte, compresi i diritti fondamentali dei destinatari del servizio, quali la libertà di espressione, la libertà e il pluralismo dei media, e altri diritti e libertà fondamentali sanciti dalla Carta». Si v. sul punto M. Husovec, *General Report: Topic II – EU Digital Economy: General Framework (Dsa/Dma) and Specialised Regimes*, cit., p. 57. Si v. altresì R.Ó. Fathaigh, D. Buijs, J. van Hoboken, *The Regulation of Disinformation Under the Digital Services Act*, in *Media and Communication*, vol. 13, 2025, p. 5.

41. Si v. considerando n. 90 del Regolamento (UE) 2022/2065, cit.

42. Ivi, art. 33, par. 6, co. 2: «Gli obblighi di cui alla presente sezione si applicano, o cessano di applicarsi, alle piattaforme online di dimensioni molto grandi interessate e ai motori di ricerca online di dimensioni molto grandi decorsi quattro mesi dalla notifica al fornitore interessato di cui al primo comma».

43. Ivi, art. 34, par. 1, lett. d.

le grandi piattaforme online e i grandi motori di ricerca devono valutare come diversi fattori – quali la progettazione degli algoritmi, la moderazione dei contenuti, le condizioni d’uso, la selezione pubblicitaria e le pratiche di gestione dei dati – influenzino tali rischi sistemici, inclusa la possibile manipolazione intenzionale del servizio e la rapida diffusione di contenuti illegali o non conformi alle condizioni d’uso.

Una volta individuati, analizzati e valutati i rischi sistemici ai sensi dell’art. 34, le piattaforme sono tenute ad adottare misure di mitigazione in conformità con l’art. 35, prestando particolare attenzione all’impatto che tali misure possono avere sui diritti fondamentali. Le possibili azioni individuate dal Regolamento, elencate in via esemplificativa ma non esaustiva, comprendono: modifiche alla progettazione e alle funzionalità dei servizi; revisione e adeguamento delle condizioni generali d’uso e delle relative prassi applicative; miglioramento delle procedure di moderazione dei contenuti; sperimentazione e aggiornamento dei sistemi algoritmici, inclusi i meccanismi di raccomandazione; adattamento dei sistemi pubblicitari, mediante interventi mirati a limitarne o modificarne le modalità di presentazione associate ai servizi.

Analogamente a quanto osservato per l’art. 34, anche l’art. 35 presenta un contenuto ampio e indeterminato. Esso si limita a indicare in termini generici le misure da adottare, senza precisare come quest’ultime debbano essere concretamente attuate dalle piattaforme, le quali risultano così vincolate allo svolgimento di compiti che esse stesse devono declinare, avvalendosi delle proprie conoscenze tecniche e operative. In questo senso, il *Digital Services Act* si conferma essere uno strumento che rafforza la legittimità procedurale, in un contesto in cui non è possibile prefissare standard uniformi e regole sostanziali⁴⁴.

Inoltre, tra le misure di mitigazione specificamente previste per le piattaforme online di grandi dimensioni, l’art. 45 contempla l’adozione di codici di condotta volontari, da elaborarsi con il coinvolgimento di una pluralità di portatori di interesse, incluse le organizzazioni della società civile. Il rifiuto da parte di una piattaforma di aderire a tali codici, su invito della Commissione, può essere considerato rilevante ai fini della valutazione del rispetto degli obblighi regolamentari. Resta tuttavia fermo che l’adozione e l’attuazione dei medesimi non costituiscono, di per sé, prova definitiva di conformità alla normativa⁴⁵.

44. In senso critico e sulla necessità di accompagnare le tutele procedurali con un’azione sul piano sostanziale, L. Califano, *La strategia normativa dell’Unione europea per un nuovo ordine digitale*, cit., p. 14.

45. Art. 45 in combinato disposto con il considerando n. 104 del Regolamento (UE) 2022/2065 cit.

In adempimento a tale previsione, il Codice di buone pratiche sulla disinformazione del 2022, a seguito dell'approvazione della Commissione nel febbraio 2025, è stato integrato nel quadro del Regolamento sui servizi digitali. Dal 1° luglio 2025 esso è divenuto il Codice di condotta sulla disinformazione, trasformandosi da strumento di autoregolamentazione a meccanismo di co-regolamentazione⁴⁶. Questo approccio consente di definire regole più specifiche e settoriali mediante il ricorso a strumenti più flessibili rispetto alla legislazione, chiarendo al contempo le previsioni del Dsa e tenendo conto delle esigenze dell'industria⁴⁷. Infine, la natura volontaria dei codici, combinata con il monitoraggio e la valutazione da parte della Commissione, dà vita a una forma di regolazione partecipata e, al tempo stesso, assicura rapidità di intervento in un settore caratterizzato da una continua e accelerata evoluzione tecnologica e sociale.

5. Conclusioni

In conclusione, il presente contributo, attraverso una riflessione sulla differenza tra contenuti *illegal* e contenuti *legal but harmful*, sulle questioni connesse al concetto di rischio sistemico, sul contenuto normativo degli articoli 34 e 35 e sul ruolo dei codici di condotta nel sistema di mitigazione dei rischi, ha inteso affrontare il tema del contrasto alla disinformazione nell'ordinamento dell'Unione europea. Come si è osservato (v. par. 2), il concetto di disinformazione non compare nell'articolato del *Digital Services Act*, ma è presente solo nei considerando. Ciononostante, non vi è dubbio che il Dsa rappresenti oggi uno degli strumenti più rilevanti per fronteggiare gli effetti negativi che tale fenomeno produce sulla tenuta dei sistemi democratici e sulla tutela dei valori europei. Tale centralità è confermata, del resto, sia dalle dichiarazioni della Presidente Ursula von der Leyen⁴⁸, sia dal programma politico della nuova Commissione per il periodo 2024-2029⁴⁹.

46. Si veda sul ruolo dei codici di condotta, E. Brogi, G. De Gregorio, *From the code of practice to the code of conduct? Navigating the future challenges of disinformation regulation*, in *Journal of Media Law*, vol. 16, n. 1, 2024, pp. 38-46.

47. O. Pollicino, P. Dunn, & L. Violante, *Intelligenza artificiale e democrazia opportunità e rischi di disinformazione e discriminazione*, cit., p. 52.

48. U. Von der Leyen, *Speech at the Copenhagen Democracy Congress*, 14 May 2024, disponibile su: www.youtube.com/watch?v=nd8TqAur-wE.

49. U. Von der Leyen, *Political Guidelines for the Next European Commission 2024–2029*, 18 July 2024, disponibile su: commission.europa.eu/document/download/e6cd4328-673c-4e7a-8683-f63ffb2cf648_en?filename=Political_Guidelines_2024-2029_EN.pdf. Si v. altresì M. McGrath, *Speech by Commissioner McGrath at the Foreign Information Manipulation*

Considerato che «tackling disinformation directly could not be rationally aimed at, because of the legal elusiveness of the definition of disinformation. Neither truthfulness, nor harmfulness of the content could be definitional elements»⁵⁰; l'importanza del Dsa sta nel suo approccio sistematico, nel porre le regole del "gioco". Seppure siano state sollevate critiche circa l'ampia discrezionalità riconosciuta alle piattaforme nella valutazione e nella gestione dei rischi, il Regolamento, in realtà, impone ai fornitori di servizi online specifici obblighi di diligenza e trasparenza⁵¹. Al contempo, introduce garanzie a tutela degli utenti, in particolare rispetto alla libertà di espressione, volte a prevenire abusi nell'applicazione delle misure di mitigazione. Inoltre, fa convergere le competenze tecniche delle piattaforme con la necessità di una supervisione pubblica. Ciò non esclude, tuttavia, il rischio che tale impostazione possa essere piegata a usi distorti: l'assenza di regole di contenuto non elimina la possibilità di abusi, soprattutto se le piattaforme tendono a esercitare il loro potere regolatorio in chiave autoreferenziale. Da qui l'esigenza di un'attenta vigilanza sull'applicazione del Regolamento, affidata in primo luogo alla Commissione – che rimane pur sempre un'istituzione politica – e, nei casi che le verranno sottoposti, alla Corte di Giustizia, chiamata a fornire chiarimenti interpretativi, come ad esempio in ordine alla nozione di rischio sistemico, a garanzia che la discrezionalità privata resti compatibile con la tutela dei diritti fondamentali⁵². In tale prospettiva, sarà inoltre essenziale rendere effettivo il coordinamento multilivello previsto dal Regolamento e promuovere un dialogo costante tra fornitori di servizi, Commissione, autorità indipendenti e società civile, così da garantire che le misure di mitigazione dei rischi siano efficaci e implementate nel rispetto della libertà di espressione.

and Interference (FIMI) Conference, 19 marzo 2025, disponibile su: ec.europa.eu/commission/presscorner/detail/en/speech_25_837.

50. J. Bayer, *The EU policy on disinformation: Aims and legal basis*, cit., p. 22.

51. Si v. G.E. Vigevani, *Potere politico e mezzi di comunicazione*, in *Rivista AIC*, n. 2, 2025, pp. 237-268.

52. L.S. Rossi, *Brevi osservazioni sulle recenti tendenze evolutive della giurisprudenza della Corte di giustizia dell'Unione europea sulla protezione dei dati personali*, in *Eurojus*, n. 7 (Fascicolo speciale "Mercato Unico Digitale, dati personali e diritti fondamentali"), 2024, pp. 51-65; C. Amalfitano, & F. Ferri, *Transizione digitale e dimensione costituzionale dell'Unione europea: tra principi, diritti e valori*, in R. Torino, S. Zorzetto (a cura di), *La trasformazione digitale in Europa. Diritti e principi*, Giappichelli, 2023, pp. 1-34; A. Iliopoulou-Penot, *Governing the "Governors" of online speech: A review of the Digital Services Act's first steps in regulating social media platforms*, cit., pp. 31-33.

RIFLESSIONI CONCLUSIVE

L'ATTUALITÀ DEL COSTITUZIONALISMO A GARANZIA DEI DIRITTI FONDAMENTALI NELLA SOCIETÀ DIGITALE

*Licia Califano**

1. L'applicazione della tecnologia digitale e l'uso degli algoritmi richiede un approccio metodologico trasversale che, quale punto di partenza, evidenzi la pregiudiziale necessità di comprendere il funzionamento della macchina – capire, cioè, quali siano le modalità di scelta e di elaborazione delle informazioni che l'utilizzo dell'AI produce – in relazione ai rischi di condizionamento e di manipolazione dei comportamenti del corpo sociale, fino al punto di comprometterne la riflessione critica e, in ultima analisi, l'esercizio del libero arbitrio. Una premessa in stretta relazione, peraltro, con la ricerca della definizione degli effetti sui sistemi democratici e sulla conseguente urgenza della elaborazione di una regolazione idonea a fissare la misura del potere e l'effettiva garanzia dei diritti fondamentali del singolo.

In piena coerenza con tali premesse il percorso di riflessione seguito in questo Volume ha coinvolto tanto l'analisi delle caratteristiche dell'approccio regolatorio delle nuove tecnologie che l'Unione europea ha negli ultimi anni adottato, nel tentativo di coniugare il crescente e rapido sviluppo dell'innovazione digitale con la tutela dei diritti fondamentali della persona, quanto, sul piano più strettamente metodologico, la problematizzazione della tenuta delle categorie del diritto costituzionale, già sperimentate e consolidate nell'esperienza delle democrazie europee del secondo Novecento, in un ambiente digitale privo di confini e dominato da una realtà oligopolistica (le grandi *corporations*) capace di esercitare un'influenza che supera la sfera economica presentandosi, piuttosto, come un potere sostanzialmente politico.

D'altra parte, se nel corso del XX secolo la riflessione politico costituzionale si è in gran parte concentrata nella ricerca di un punto di equilibrio fra Stato e mercato, non vi è dubbio che nel Secolo presente la riflessione si è spostata

* Professoressa ordinaria di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

sul versante del controllo che, sul piano tanto del singolo individuo quanto del corpo sociale nel suo insieme, gli algoritmi sono in grado di esercitare con effetti inediti sui diritti di libertà e di autogoverno che costituiscono il nucleo della democrazia. Ed è proprio a fronte di un modello di regolamentazione del digitale in cui ha prevalso la concentrazione del potere in capo a un oligopolio di soggetti privati che l'Unione europea ha impresso una direzione di marcia che supera l'idea che il controllo dei dati e delle informazioni, così come il conseguente fenomeno della "disinformazione online", si possa affidare alla capacità auto-correttiva del mercato delle idee, scegliendo piuttosto la strada delle forme della regolazione.

Un *corpus* normativo che la ricerca ha evidenziato essere orientato alla *governance* e alla trasparenza: un regime a "responsabilità limitata" delle grandi piattaforme che, ispirato al valore del principio personalista e alla garanzia dei diritti fondamentali, sconta ancora il limite di soluzioni troppo circoscritte alla cd. procedimentalizzazione del rischio. Ciò nondimeno emerge la consapevolezza dei rischi e la conseguente reazione dei pubblici poteri che dalla *self-regulation* muovono nella direzione della regolazione e dalla metafora del libero mercato alla costruzione di un bilanciamento fra gli interessi pubblici al corretto funzionamento del sistema democratico e alla garanzia dei diritti costituzionalmente sanciti e le libertà economiche delle imprese.

Difficile, oggi, parlare di un inquadramento costituzionale delle piattaforme digitali, ma le osservazioni svolte nel corso della ricerca consentono qualche spunto conclusivo.

Un percorso di riflessione, va anzitutto sottolineato, che nel suo sviluppo ha seguito come prevalenti profili tematici:

- l'attualità delle categorie tradizionali del costituzionalismo liberal-democratico, chiamato ad affrontare una dimensione digitale priva di confini e, forse ancor più, caratterizzata da un diversificato o, forse meglio, capovolto rapporto pubblico/privato;
- gli effetti dirompenti del rapporto "disinformazione"/AI progressivamente applicata anche all'ambito della comunicazione politica. Una cornice, quest'ultima, nella quale il tema delle *fake-news* ripropone in termini moderni il rapporto tra informazione e verità, errore e finzione, inganno e menzogna.

Terreno complesso che nel Volume si affronta coniugando crisi della verità, crisi del pluralismo delle idee come prodotto della personalizzazione dei contenuti, nella prospettiva dell'incidenza sulla libertà del dibattito pubblico e, in definitiva, sul funzionamento della democrazia.

L'osservazione della storia costituzionale contemporanea e, con essa, la traiettoria espansiva che nei decenni del Secondo dopoguerra, con l'affermarsi di regimi politici democratici, ne ha caratterizzato l'affermazione e l'espansione, vive una fase che se non interrotta è certamente ostacolata da una serie di eventi (e di emergenze globali) che caratterizzano la contemporaneità e che hanno posto problemi nuovi e, al contempo, evidenziato possibili incrinature nel rapporto tra percorsi del costituzionalismo e prospettive della democrazia.

In che misura, chiediamoci, si può condividere l'osservazione che le democrazie del secondo Novecento, nate sull'ipotesi di una permanente capacità dei momenti costituenti e dei processi democratici di integrare il futuro nel loro programma, rischiano di perdere l'orizzonte temporale, estraneo alla tecnica, che invece ha caratterizzato tutte le epoche storiche: una "eccedenza di presente" che non consente di comprendere il mutamento del regime di storicità sottostante alle Costituzioni contemporanee.

Ancora, e per altro verso, a un approccio nostalgico del programma futuro di allora, in che misura si può opporre una lettura ricostruttiva critica capace di cogliere le dinamiche del costituzionalismo e tentare di riportare al centro della riflessione i problemi di cui lo Stato costituzionale è riuscito, fino a ora, a prendersi cura.

Altrimenti detto, la sostenibilità teorica di quei principi, di quelle idee e contenuti trova ancora oggi le condizioni che ne determinano i presupposti funzionali?

Anche su questo difficile terreno di indagine il Volume cerca di offrire al lettore una elaborazione degli strumenti concettuali che consenta di interpretare il mutamento costituzionale in ragione di una sua rinnovata attualità e, lungo questa strada, auspicabilmente, aiutare a comprendere i problemi della contemporaneità.

2. Indispensabile, in questa prospettiva, fare un passo indietro nella storia, e tornare alle radici di cui è espressione il movimento del costituzionalismo democratico.

Quali siano i contenuti concettuali del costituzionalismo (cioè attraverso quali tecniche esso garantisce l'esercizio dei diritti fondamentali) e il nesso funzionale che lo lega allo Stato moderno è certamente il punto di partenza imprescindibile di ogni riflessione sui diritti fondamentali di libertà e sui percorsi che hanno condotto alla loro progressiva positivizzazione.

Quali siano, in secondo luogo, le funzioni attuali del costituzionalismo, in ragione dell'affermarsi della dottrina del costituzionalismo multilivello, con la moltiplicazione delle sedi di riconoscimento e di protezione dei diritti che ne

consentirebbe l'assestamento della tutela al livello più alto, è la prima domanda che si lega alla progressiva erosione dei tratti costitutivi dello Stato moderno.

Certo, le moderne Costituzioni nascono (nel senso che vengono formalizzate in documenti scritti) allo scopo di fissare limiti al potere di chi comanda, definendo le condizioni e i modi di esercizio dell'autorità e, al contempo, per garantire i diritti dei singoli nei confronti del potere pubblico.

L'idea moderna di Costituzione si lega, insomma, a precisi contenuti delle leggi fondamentali e il costituzionalismo, a sua volta, non è una scienza neutrale delle leggi costitutive dello Stato, ma è presupposto e sviluppo di quelle idee fondamentali.

Saranno poi le Costituzioni del Secondo dopoguerra a intrattenere un vero e proprio rapporto costitutivo con i diritti fondamentali trovando, anzi, proprio nella tutela di questi ultimi la loro più profonda ragione d'essere. Nel costituzionalismo liberal-democratico la sovranità dello Stato diviene l'altra faccia della libertà degli individui: il momento dell'autorità e il momento della libertà si presentano così simmetricamente opposti ma, al tempo stesso, inscindibilmente legati.

Non solo con le Costituzioni contemporanee la disciplina costituzionale dei diritti fondamentali e delle libertà raggiunge un grado di diffusione da rappresentare una costante che ha influenzato la stessa architettura sistematica dei documenti costituzionali quanto, soprattutto, il rapporto costitutivo assume un significato più profondo: i diritti fondamentali diventano tali, cioè in senso giuridico, proprio in forza della disciplina costituzionale. Con le Costituzioni le libertà assurgono a rango di diritti, configurandosi come limiti all'azione del potere sovrano; insomma, di tali diritti lo Stato nazionale diviene al contempo controparte e garante.

Per altro verso, e in primo luogo, il carattere dell'autosufficienza del riconoscimento costituzionale dei diritti di libertà è costretto a confrontarsi con l'avvento dei diritti sociali che richiedono per la loro realizzazione interventi positivi rimessi al legislatore; la relazione costitutiva fra diritti fondamentali e Costituzione, cioè, si attenua nel senso che quest'ultima non è più autosufficiente, dovendo trovare il proprio sviluppo nella normativa di attuazione (profilo questo che, peraltro, non esclude affatto la produzione di effetti immediati e diretti).

In secondo luogo l'internazionalizzazione dei diritti, cioè il loro ancoraggio anche in documenti internazionali, ha comportato anzitutto, e in maniera definitiva e inequivocabile, il tramonto dell'idea che la tutela dei diritti fondamentali rientri fra gli interessi esclusivi degli Stati.

Il problema si sposta, piuttosto, nel verificare quando, e in che misura, il circuito internazionale si affianca e quando più propriamente si integra, si compe-

netra con il circuito costituzionale interno, ciò che avviene quando gli accordi sui diritti ricevono un riconoscimento espresso nei testi costituzionali venendo così in tutto o in parte costituzionalizzati.

Ma la portata innovativa di questo processo di reciproca integrazione aumenta in maniera esponenziale, se riflettiamo sul dialogo e sul ruolo progressivamente assunto dalle Corti.

Questo accade quando gli organi di giurisdizione costituzionale usano gli accordi internazionali sui diritti come coefficiente ermeneutico, tendendo cioè a leggere le rispettive Costituzioni alla luce di questi; va poi considerato il ruolo ormai indiscutibile della Corte di Lussemburgo come giudice dei diritti personali (da tempo non più solo economici) in progressivo consolidamento rispetto alla Corte di Strasburgo. Infine, ma non ultimo, il profilo del riconoscimento reciproco tra Corti europee e Corti nazionali. Insomma, i diritti nati negli Stati contemporanei come libertà dal potere nazionale, stanno diventando, grazie all'opera delle Corti, strumenti di costruzione di un potere sovranazionale. Un processo di accelerazione non sempre lineare, come del resto fisiologico, ma già ricco di effetti da segnalare con attenzione; perché sullo sfondo della dissociazione fra diritti e positività della Costituzione prende corpo l'idea che la migliore protezione dei diritti si realizza nella molteplicità dei livelli di riconoscimento e di tutela.

Se per un verso la forza di norme pattizie a carattere generale è legata al superamento della barriera della sovranità nazionale (e della lacerazione storica fra diritto interno e diritto esterno allo scopo di salvaguardare il principio dell'autorità), per l'altro la domanda a cui dare una risposta sempre di più coinvolge le funzioni e le nuove frontiere di un costituzionalismo che, nato e costruito come limite al potere assoluto, è oggi in affanno, fatica a trovare risposte a interrogativi il cui tema di fondo non è più il limite al potere dello Stato, quanto l'essenza stessa, il contenuto delle libertà individuali e della loro tutela.

In questo quadro, già di per sé complesso e in evoluzione, si innesta la crescente pervasività dell'innovazione tecnologica, la cui prima specificità è proprio il suo andare al di là dei confini nazionali e sovranazionali, ponendo al centro della riflessione un costituzionalismo che si trova a fare i conti con la crescente centralità di attori privati, come le grandi piattaforme digitali, nella tutela dei diritti e delle libertà fondamentali degli individui.

Da un lato, la rete e i suoi attori costituiscono evidentemente il contesto fondamentale per l'esercizio di vecchi e nuovi diritti fondamentali: così la libertà di manifestazione del pensiero o la libertà di associazione, che grazie alle piattaforme *social* assume oggi una dimensione pervasiva nelle formazioni sociali; così se pensiamo alla discussione sul diritto di accesso a internet, come nuovo diritto fondamentale della persona.

Dall'altro, meno evidentemente, ad attori privati è chiesta sempre più una responsabilizzazione nei confronti dei diritti e delle libertà fondamentali delle persone, che si traduce non semplicemente nel rispetto passivo dei diritti e delle libertà fondamentali, ma nella capacità proattiva di attuare misure tecniche e organizzative tali da poter ridurre il rischio di un loro effettivo godimento da parte delle persone.

In un "ecosistema digitale" privo di confini il tema della regolazione privata, riassuntivamente indicato come *governance* o normazione alternativa, ha inizialmente definito forme di partecipazione di soggetti privati a funzioni tradizionalmente prerogativa dei poteri pubblici. Le manifestazioni più note, almeno in ambito UE, potevamo individuarle nei codici di condotta, nelle linee guida, nelle buone pratiche, nei c.d. *standards* quali misure che, se riconducibili alla categoria degli atti di *soft-law*, perché formalmente non vincolanti, ciò nondimeno erano astrattamente in grado di indirizzare la condotta di amministrazioni, imprese e dei cittadini stessi.

Un approccio iniziale improntato prevalentemente alla logica dell'auto-regolazione e del liberismo "digitale" basato sull'idea di una "rete libera" sia con riguardo ai profili economicistici sia in relazione ai contenuti, anche sul modello statunitense, superato come si è detto, in Europa, da una strategia maggiormente interventista, finalizzata a costruire un sistema regolatorio ben preciso e ponendo a sua tutela un meccanismo di *law enforcement* di carattere nuovamente pubblicistico e, in qualche caso, anche fortemente accentrato.

3. Per altro verso, la domanda che attende risposta è in quali modi e forme attualizzare, per non disperdere, l'istanza liberale di limitazione del potere e, con essa, il principio personalista che si configura quale principio cardine nei sistemi costituzionali democratici. E, ancora, come ricondurre il discorso pubblico a un apparato istituzionale che sia in grado di garantire costituzionalmente le comunicazioni e mantenerle al riparo da processi di privatizzazione realizzati da soggetti (privati) in posizione di oligopolio detentori del potere nel *web*, la cui peculiarità si lega a un carattere apparentemente a-spaziale e a vocazione universale. Quale spazio residua, in questo panorama, alla potestà normativa e repressiva dello Stato se le piattaforme digitali sono, di fatto, interlocutori degli individui che esercitano in rete i diritti inviolabili di cui sono titolari.

In più saggi di questo Volume si è sottolineata la coesistenzialità fra modello democratico e difesa della dignità e libertà individuali evidenziando, peraltro, in che termini la coesistenza di un tale sistema di valori pluralistici trovi fra le sue principali declinazioni la libertà di manifestazione del pensiero. Sul punto la radicale quanto veloce trasformazione del sistema tradizionale dei sistemi di comunicazione verso una dimensione globale veicolata dalla rete, divenuta

per ampi strati della popolazione la principale, quando non esclusiva, fonte di informazione, ha cambiato la dimensione costituzionale in termini di tempo e di spazio. Una velocità di trasmissione delle notizie che produce un inevitabile impatto sociale caratterizzato, come già osservato, dall'assenza di frontiere; insomma, una diffusione globale di dati e informazioni.

Informazione e disinformazione superano i confini degli Stati nazionali e trasformano la sfera pubblica in un luogo bidimensionale nel quale l'individuo è solo e fruisce in maniera costante di informazioni prodotte da altri individui in assenza di intermediazione. Contesto cui va aggiunta la circostanza non marginale che sono i *social-network* a selezionare e ordinare le informazioni da proporre agli utenti sulla base di istruzioni proceduralizzate di un algoritmo che personalizza le informazioni sulla base degli interessi e preferenze individuali, frutto di una avvenuta precedente profilazione. In altre parole, a un confronto fra diversi punti di vista si sostituisce una rappresentazione autoreferenziale della realtà.

Accertati i margini di vulnerabilità delle società aperte (e la conseguente necessità della ricerca di nuove strategie di sicurezza nazionale e internazionale) diviene centrale provare a stabilire quali siano gli strumenti che gli Stati liberali sono in grado di opporre ai rischi che la crescente disinformazione in rete rischia di produrre.

Certamente gli oligopolisti dell'informazione esercitano poteri confrontabili con quelli esercitati dagli Stati, determinando un contesto nel quale, al contempo, il ruolo dei poteri pubblici nella regolamentazione del *cyberspazio* viene messo in discussione.

Certamente gli Stati democratici si trovano, per un verso, a dover garantire l'effettività dei diritti e delle libertà nella rete, riconoscendo un alto grado di libertà economica alle piattaforme che, dal canto loro, non sono tenute a perseguire interessi pubblici né, tantomeno, a garantire i diritti fondamentali degli individui. La conseguenza è la creazione di un ampio margine di discrezionalità dei soggetti privati nell'attuazione dei diritti e nel loro bilanciamento. Per l'altro, la previsione di un sistema di controllo dei soggetti privati non è certo esente dal rischio di censura sia pubblica che privata.

Quali rimedi?

Una prima riflessione è d'obbligo.

Conciliare il nostro passato con il nostro futuro, a partire proprio dai diritti fondamentali e dalla loro tutela, non può prescindere dal mantenimento di uno dei tratti che hanno caratterizzato i regimi democratici. L'equilibrio dinamico fra i diversi diritti, tutti fondamentali e tutti sullo stesso piano. Una premessa strutturale, questa, indispensabile nella prospettiva del pluralismo dei valori e del funzionamento della democrazia. Si vuole in altri termini sot-

tolineare che l'eventuale prevalenza di un diritto sugli altri riposa sulla fase applicativa in relazione, cioè, a casi concreti e ci conduce, evidentemente, a questioni di giustizia sostanziale. L'esperienza democratica ci insegna che è il bilanciamento di natura giurisprudenziale (e non la graduazione che sottende un diverso modo di intendere la normatività dei fatti) a confermare il primato delle Costituzioni e del diritto costituzionale. Quella rigidità che se sul piano formale trova una sponda procedimentale, sul versante sostanziale si affida alla giurisprudenza costituzionale. Una coesistenza dei diritti regolata dal bilanciamento che ha implementato nella società l'esercizio di una pluralità di diritti fondamentali individuali e collettivi. Vi è poi un ulteriore aspetto, sul quale non è possibile soffermarsi in queste brevi riflessioni finali, che coinvolge la dimensione della sostenibilità di un regime di eccezione, concepito in base a un paradigma che spiega l'eccezione in termini non riconducibili all'ordine giuridico, i cui contorni rischiano di intercettare le problematiche sottese alle nostre riflessioni.

Tanto premesso, è fuor di dubbio che la tenuta del costituzionalismo quale dottrina volta a limitare i poteri pubblici e definire spazi di autonomia garantiti per i consociati, attraversa, in primo luogo, la rivendicazione della centralità della sua dimensione prescrittiva e, con essa, l'imprescindibilità della regolazione pubblica. Un'affermazione, questa, che nel contesto digitale in cui la realtà oligopolistica esercita un'influenza tale da possedere, in sostanza, un potere che supera la mera sfera economica e si presenta come un potere sostanzialmente politico, ha posto alla ricerca il tema della possibilità di un'estensione degli obblighi costituzionalmente previsti anche nei confronti di soggetti privati, superando così la rigida separazione fra attori pubblici e privati. A un potere che nella società digitale si è progressivamente spostato dalla dimensione verticale verso quella orizzontale, il costituzionalismo, per rimanere fedele alla sua missione originaria di limitare il potere per garantire il rispetto dei diritti fondamentali, deve aprire a una riflessione che, di fronte a un fenomeno senza precedenti nella storia dell'uomo, ridefinisca gli spazi in cui rimane fondamentale l'intervento delle istituzioni democratiche al fine di garantire i diritti fondamentali degli individui in Rete. In che modo, in altri termini, creare le condizioni che impongano anche agli attori privati di confrontarsi con le norme applicabili ai soggetti pubblici.

Un quadro in cui gli Stati, pur mantenendo la propria identità, sono anzitutto chiamati a confrontarsi con la dimensione globale che la società digitale dischiude e, con essa, con una visione dei principi che identificano le democrazie liberali la cui garanzia di effettività impone la ricerca di un punto di equilibrio tra cessione di quote di sovranità a organizzazioni sovranazionali e internazionali e mantenimento della funzione di limitazione del po-

tere. Una dinamica di cooperazione fra Stati in una prospettiva valoriale, di integrazione reciproca fra principi e tutele costituzionali, sul terreno specifico della salvaguardia dei diritti fondamentali, già in atto e che vede nella dimensione europea il terreno più fertile alla reciproca implementazione. Limitare e, dunque, regolamentare il potere delle *big corporations* in rete mediante la predisposizione di un nuovo quadro normativo impone la dimensione europea, concretizzando l'idea della reciproca integrazione, di un ordine "intercostituzionale" nel quale le Costituzioni si alimentano l'una dall'altra sullo specifico terreno dei diritti fondamentali. Da qui la ricerca di un punto di equilibrio fra potere privato e garanzia dei diritti inviolabili esercitati nelle piattaforme online che il legislatore europeo ha affrontato con un approccio normativo regolamentare.

Le attuali scelte legislative europee sui mercati digitali delineano un quadro normativo (dati, contenuti, *software*) della responsabilità degli attori del digitale, aprendo orizzonti a una declinazione del costituzionalismo che aspira a governare l'innovazione digitale all'interno della tenuta delle categorie giuridiche tradizionali, rivisitate e rivitalizzate, quale strumento di democrazia. La prospettiva europea è delineare un percorso, per quanto difficile, di regolazione pubblica omogenea nell'ambito territoriale degli Stati membri che possa offrire maggiore garanzia ai diritti della persona-utente. Un insieme di misure che innovano là dove impongono, soprattutto alle piattaforme di grandi dimensioni, vincoli procedurali legati agli obblighi di trasparenza e all'*accountability*. Una regolamentazione dell'offerta dei servizi digitali che si focalizza su tre presupposti cardine: trasparenza, responsabilità e garanzia nei confronti dell'utente. Un regime a responsabilità diversificato in forza del quale i destinatari delle prescrizioni devono assicurare di aver posto in essere i rigidi adempimenti individuati dall'atto normativo affinché possano essere considerati esenti da responsabilità per i contenuti illegali che si trovino involontariamente a ospitare.

In altri termini una riorganizzazione dell'azione dei colossi digitali all'interno di un nuovo e composito regime normativo che traccia le linee di un disegno di controllo esercitato dal soggetto pubblico e dall'esigenza, dotata del medesimo peso assiologico, di un sistema che aspira comunque a conformarsi sulle basi fondanti della *rule of law*. Aspirazione che si coglie non soltanto dai ricorrenti ed espliciti riferimenti a valori nodali, dal principio della certezza del diritto a quello di eguaglianza, ma si deduce implicitamente dalle dettagliate garanzie procedurali, che ricalcano i fondamenti tipici dello Stato di diritto costituzionale: dall'obbligo di motivazione della decisione al diritto di appello di fronte a un soggetto indipendente e imparziale.

Permangono però diffuse criticità che ridimensionano in radice gli intenti garantistici dell'atto e le sue potenzialità di salvaguardia nei confronti dei diritti

ti della persona-utente. Difficile risolvere le problematiche che emergono nei meccanismi algoritmici, propri dei nuovi poteri, attraverso garanzie procedurali che non sono da sole in grado di compensare l'assenza di garanzie sostanziali comparabili a quelle proprie dei rapporti con i soggetti pubblici.

Un regime "a responsabilità condizionata" che viene poi implementato mediante l'adozione di differenti misure caratterizzate da procedure più celeri e articolate sia per l'eliminazione di elementi o prodotti illegali, sia per la presentazione di reclami e segnalazioni da parte degli utenti. In altri termini è intorno alla effettiva concretizzazione della tutela delle posizioni soggettive coinvolte nello spazio digitale che tende a modellarsi il complesso regime di responsabilità costruito su vincoli procedurali (dagli obblighi di trasparenza alle periodiche valutazioni dei rischi sistemici che condizionano l'operato degli intermediari) funzionali al raggiungimento dello scopo dell'intervento di rafforzare il contrasto all'illecito. Ora, anche a voler prescindere dalla conseguenza non voluta – ma probabile – che il rigido sistema di obblighi introdotti possa in realtà incentivare una politica volta all'eliminazione indiscriminata di elementi qualificabili come "illegali" – perché in presenza di doveri e sanzioni, lo sforzo di svolgere un'attività di moderazione scevra da errori di valutazione viene sovrastato dall'esigenza di porre in essere un blocco sistematico dei contenuti sospetti – è un altro il profilo che qui interessa sottolineare. L'intervento perseguito dal legislatore continentale di "porre le briglie" al dilagante potere delle piattaforme, mediante un'onnicomprendente articolazione delle garanzie, rimane esso stesso imbrigliato in una logica di espansione delle tutele di natura meramente procedurale.

Un approccio che se rappresenta certamente la via più percorribile per costruire un sistema di salvaguardia dei diritti fondamentali nel *cyberspazio*, si configura quale compromesso che se non accompagnato da una robusta azione sul fronte sostanziale, rischia di creare dei presidi garantistici formalmente imponenti, ma vuoti sul piano sostanziale, perché ancora esposti alla discrezionalità interpretativa e applicativa delle piattaforme libere, ancora una volta, di plasmarne il significato.

C'è però un ulteriore importante elemento di riflessione, già ricordato in queste pagine, ma che merita di essere opportunamente sottolineato: il costituzionalismo del secondo Novecento si è caratterizzato per il rafforzamento delle Corti costituzionali come difensori ultimi dei diritti fondamentali. Possiamo cioè affermare che il costituzionalismo democratico oltre a limitare il potere ha nel proprio DNA la vocazione a "costituzionalismo dei diritti". Una tradizione ormai consolidata che ha aperto la strada all'affermazione dell'efficacia interpretativa dei diritti fondamentali che, a ben vedere, si configura quale estensione orizzontale delle norme costituzionali relative ai diritti fondamentali. Una

modalità di contenimento del nuovo potere digitale che è stata seguita dalla Corte di giustizia dell'Unione europea con particolare attenzione grazie a una adeguata valorizzazione dei precedenti in tema di applicazione orizzontale dei principi generali del diritto dell'Unione o, ancora, di alcuni articoli della Carta dei diritti fondamentali dell'Unione europea.

Impossibile, nella fase che viviamo di profonda e radicale trasformazione del mondo, tracciare delle conclusioni, possiamo solo limitarci a osservare che se dalla storia costituzionalismo e democrazia traggono la forza per riconsiderare il presente e la propria prospettiva di azione, le pagine di questo Volume aspirano a inserirsi nel pensiero critico volto a proporre soluzioni funzionali a un ordinamento costituzionale chiamato a confrontarsi con le dinamiche in atto di cambiamento della società contemporanea.

AUTRICI E AUTORI

Giovanni Barozzi Reggiani, Ricercatore di Diritto costituzionale e pubblico, Università degli Studi di Sassari.

Domenico Bruno, Assegnista di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Napoli Federico II.

Licia Califano, Professoressa ordinaria di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

Massimo Cavino, Professore ordinario di Diritto costituzionale e pubblico, Università del Piemonte Orientale.

Federico Cerutti, Professore ordinario di Ingegneria industriale e dell'Informazione, Università degli Studi di Brescia.

Casimiro Coniglione, Assegnista di ricerca in Filosofia del diritto, Università degli Studi di Modena e Reggio Emilia.

Giuseppe Contissa, Professore associato di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

Lorenzo De Carlo, Dottorando di ricerca in Diritto, Scuola superiore di Studi Universitari e di Perfezionamento "Sant'Anna" di Pisa.

Allegra Dominici, Dottoranda di ricerca in Scienze giuridiche, Università degli Studi di Urbino Carlo Bo.

Massimo Durante, Professore ordinario di Filosofia del diritto, Università degli Studi di Torino.

Ludovica Durst, Assegnista di ricerca in Diritto costituzionale e pubblico, Sapienza Università di Roma.

Federica Fabrizzi, Professoressa associata di Diritto costituzionale e pubblico, Sapienza Università di Roma.

Emanuele Gabriele, Dottore di ricerca in Diritto pubblico, comparato e internazionale, Sapienza Università di Roma.

Federico Galli, Ricercatore di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

Sara Gallone, Dottoranda di ricerca in Diritto dell'Unione europea, Alma Mater Studiorum – Università di Bologna.

Rosa Iannaccone, Contrattista di ricerca in Diritto pubblico comparato, Università degli Studi di Sassari.

Andrea Loreggia, Professore associato di Ingegneria industriale e dell'informazione, Università degli Studi di Brescia.

Giulia Napoli, Dottoranda di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Milano-Bicocca.

Andrea Ongarini, Collaboratore di ricerca in Ingegneria industriale e dell'Informazione, Università degli Studi di Brescia.

Emanuela Palomba, Dottoranda di ricerca in IA – Diritto costituzionale e pubblico, Università degli Studi di Napoli Federico II.

Anna Papa, Professoressa ordinaria di Diritto costituzionale e pubblico, Università degli Studi di Napoli Parthenope.

Andrei-Mihai Pop, Dottorando di ricerca in Filosofia del diritto, Università degli Studi di Milano.

Andrea Ruffo, Dottorando di ricerca in Diritto e Impresa, Università degli Studi di Roma Unitelma Sapienza.

Giovanni Sartor, Professore ordinario di Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

Camilla Scarpellino, Assegnista di ricerca in Filosofia del diritto, Alma Mater Studiorum – Università di Bologna.

Giuliaserena Stegher, Ricercatrice di Diritto pubblico comparato, Università degli Studi di Urbino Carlo Bo.

Luca Maria Tonelli, Assegnista di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

Giulia Vasino, Ricercatrice di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

Giulio Enea Vigevani, Professore ordinario di Diritto costituzionale e pubblico, Università degli Studi di Milano-Bicocca.

a strong international commitment

The **FrancoAngeli**, **FrancoAngeli Journals** and **FrancoAngeli Series** websites now offer a completely dual language interface, in Italian and English.

Ensuring international visibility and discoverability for our authors is of crucial importance to us.



Questo LIBRO



ti è piaciuto?

Comunicaci il tuo giudizio su:

www.francoangeli.it/opinione



**VUOI RICEVERE GLI AGGIORNAMENTI
SULLE NOSTRE NOVITÀ
NELLE AREE CHE TI INTERESSANO?**



ISCRIVITI ALLE NOSTRE NEWSLETTER

SEGUICI SU:



FrancoAngeli

La passione per le conoscenze

Vi aspettiamo su:

www.francoangeli.it

per scaricare (gratuitamente) i cataloghi delle nostre pubblicazioni

DIVISI PER ARGOMENTI E CENTINAIA DI VOCI: PER FACILITARE
LE VOSTRE RICERCHE.



Management, finanza,
marketing, operations, HR

Psicologia e psicoterapia:
teorie e tecniche

Didattica, scienze
della formazione

Economia,
economia aziendale

Sociologia

Antropologia

Comunicazione e media

Medicina, sanità



Architettura, design,
arte, territorio

Informatica, ingegneria
Scienze

Filosofia, letteratura,
linguistica, storia

Politica, diritto

Psicologia, benessere,
autoaiuto

Efficacia personale

Politiche
e servizi sociali



FrancoAngeli

La passione per le conoscenze

Disinformazione digitale

La sfida posta dalla disinformazione digitale non è rinviabile: essa interpella il presente delle istituzioni democratiche e richiede uno sforzo di riflessione che coniughi analisi critica, responsabilità normativa e innovazione tecnologica. Partendo da tale premessa, il Volume offre una mappa teorica e applicativa delle tensioni che attraversano l'odierno cyberspazio, adottando una prospettiva multidisciplinare e una duplice direttrice analitica: da un lato, la riflessione sulle condizioni di compatibilità tra il contrasto alla disinformazione e la tutela dei diritti fondamentali nel quadro del costituzionalismo liberal-democratico; dall'altro, l'esigenza di tradurre tale bilanciamento in architetture tecnologiche e sistemi operativi coerenti con un paradigma human-centric.

Licia Califano è professoressa ordinaria di Diritto costituzionale presso l'Università degli Studi di Urbino Carlo Bo e Direttrice del Dipartimento di Giurisprudenza presso la medesima Università. È stata componente del Collegio del Garante per la protezione dei dati personali dal 2012 al 2020.

Federica Fabrizzi è professoressa associata di Istituzioni di diritto pubblico presso Sapienza – Università di Roma. È attualmente Vicedirettrice della Rivista di diritto pubblico italiano, comparato ed europeo *federalismi.it*.

Giovanni Sartor è professore ordinario di Filosofia del diritto e Informatica giuridica presso l'Alma Mater Università di Bologna. Nel 2019 ha ottenuto un Advanced Grant dell'ERC sul tema del *Computable Law* e nel 2025 un FIS Advanced Grant su *Generative AI for the Law*.