Ordine internazionale e diritti umani
International Legal Order and Human Rights
Ordenamiento Jurìdico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino
EDITORIALE SCIENTIFICA

LORENZO DI ANSELMO*

INTELLIGENZA ARTIFICIALE, LOTTA ALLA DISINFORMAZIONE *ONLINE* E TUTELA DELLA DEMOCRAZIA NELL'UNIONE EUROPEA ALLA LUCE DELL'*AI ACT*

SOMMARIO: 1. Considerazioni preliminari sul rapporto tra intelligenza artificiale e disinformazione. – 2. Brevi cenni sulla struttura e sulle finalità dell'*AI Act*. – 3. Le pratiche vietate ai sensi dell'art. 5: quale rilievo ai fini del contrasto alla disinformazione *online*? – 4. Il contrasto alla disinformazione nell'ambito dei sistemi di intelligenza artificiale ad alto rischio. – 5. Gli obblighi di trasparenza per i sistemi di IA a rischio minimo e le disposizioni specifiche in materia di *deep fake*. – 6. Conclusioni.

1. *Considerazioni preliminari sul rapporto tra intelligenza artificiale e disinformazione*

In un'epoca caratterizzata dalla c.d. infodemia, ossia dalla diffusione virale e spesso incontrollata delle informazioni, l'integrazione dei sistemi di intelligenza artificiale (IA) negli strumenti di comunicazione digitale appare destinata ad impattare profondamente sul dibattito pubblico. Anche laddove il ricorso a tali sistemi non celi un intento malevolo, il loro utilizzo può favorire la circolazione di notizie false o non attendibili, amplificando gli effetti dannosi della disinformazione[1]. Ciò perché, nell'epoca plasmata dalle reti virtuali, l'efficacia e la visibilità di un contenuto non dipendono dalla sua veridicità, quanto piuttosto dalla sua capacità di generare interazioni. Tale dinamica, premiando le notizie più divisive, ancorché non verificate, poiché più funzionali ad intercettare l'interesse degli utenti, tende ad accelerare il flusso di contenuti disinformativi[2]. Non a caso, i sistemi algoritmici che governano le piattaforme digitali sono progettati per raggiungere gli utenti più sensibili a specifici contenuti, con lo scopo di attirarne l'attenzione e massimizzarne il coinvolgimento. Sotto questo profilo, l'IA può essere impiegata per finalità di profilazione, permettendo, specialmente attraverso tecniche di apprendimento automatico, di elaborare i dati, le

* Borsista di ricerca *post-doc* in Diritto dell'Unione europea, Università di Roma La Sapienza.

[1] In generale sul rapporto tra intelligenza artificiale e disinformazione, cfr. N. BONTRIDDER, Y. POULLET, *The Role of Artificial Intelligence in Disinformation*, in *Data & Policy*, n. 3, 2021, p. 3 ss. Per una prospettiva più prettamente giuridica, cfr. O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2024; A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, 2025, p. 407 ss.

[2] S. VOSOUGHI, D. ROY, S. ARAL, *The Spread of True and False News Online*, in *Science*, 2018, p. 1146 ss.

abitudini e le preferenze degli utenti al fine di proporre loro contenuti personalizzati e, quindi, più attrattivi[3]. Oltre ad amplificare il fenomeno della disinformazione *online*, l'IA, attraverso i c.d. modelli generativi, può essere sfruttata per produrre intenzionalmente contenuti falsi o, comunque, manipolati in modo tale da alterare la percezione della realtà. Tali sistemi, infatti, permettono di creare testi e immagini altamente realistici, difficili da distinguere persino dagli utenti più accorti: emblematico, in tal senso, il caso dei *deep fake*, video in cui personaggi pubblici sembrano compiere azioni o pronunciare parole in realtà mai avvenute.

Benché tali fenomeni minaccino costantemente di inquinare il dibattito pubblico[4], ci sembra irrealistico immaginare di passare in rassegna tutte le notizie che circolano nella rete al fine di censurare quelle ritenute false, non solo perché la velocità di diffusione delle informazioni *online* renderebbe vana, o quantomeno parziale, siffatta operazione, ma anche perché un simile scenario presupporrebbe l'individuazione di criteri rigorosi sulla base dei quali valutare la veridicità di un contenuto, con la conseguenza di comprimere la libertà di espressione degli utenti[5], tutelata, nell'ordinamento dell'Unione, dall'art. 11 della Carta dei diritti fondamentali. In ragione di ciò, non sorprende che le iniziative intraprese al riguardo dall'Unione europea si siano concentrate sui contenuti disinformativi suscettibili di pregiudicare i valori fondamentali dell'Unione di cui all'art. 2 TUE, *in primis* i diritti umani e i principi democratici[6], specialmente allorché la circolazione di tali contenuti sia riconducibile a campagne disinformative architettate deliberatamente da attori stranieri[7].

In questi casi, infatti, i sistemi di intelligenza artificiale possono essere sfruttati per promuovere narrazioni faziose e non aderenti alla realtà, aumentandone esponenzialmente la visibilità nella sfera digitale. Se pianificate in corrispondenza di importanti appuntamenti elettorali, quando i cittadini diventano particolarmente interessati alla ricezione di messaggi politici, queste campagne possono contribuire a manipolare l'opinione pubblica, influenzando la formazione del consenso popolare e, quindi, alterando il regolare

---

[3] Sul tema, cfr. F. LAGIOIA, G. SARTOR, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *federalismi.it*, n. 11, 2020, p. 85 ss.

[4] Una preoccupazione in tal senso è espressa, non a caso, anche nella Dichiarazione europea sui diritti e i principi digitali per il decennio digitale, la quale, nel declinare la libertà di espressione secondo le peculiarità proprie dell'ambiente digitale, precisa che le piattaforme digitali, «visto il ruolo svolto dai loro servizi nel plasmare l'opinione pubblica e il dibattito pubblico […] dovrebbero attenuare i rischi derivanti dal funzionamento e dall'uso dei loro servizi, anche in relazione alle campagne di disinformazione e cattiva informazione» (Dichiarazione comune di Parlamento europeo, Consiglio e Commissione europea, *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, 23 gennaio 2023, par. 15). Per un commento generale sulla Dichiarazione, si rinvia a P. DE PASQUALE, *Verso una Carta dei diritti digitali (fondamentali) dell'Unione europea?*, in *Il Diritto dell'Unione europea*, 2022, p. 163 ss.

[5] B. BOTERO ARCILA, R. GRIFFIN, *Social media platforms and challenges for democracy, rule of law and fundamental rights*, Study requested by the LIBE Committe, Brussels, 2023, p. 77, reperibile *online*.

[6] In effetti, nell'accezione accolta dall'Unione europea (cui si farà quindi riferimento in questa sede), non è il contenuto falso o ingannevole di per sé a qualificare una notizia come disinformazione, quanto la sua predisposizione ad arrecare un "pregiudizio pubblico", dunque un danno a un interesse di carattere generale per la collettività. Rileva, al riguardo, la definizione contenuta in una comunicazione della Commissione del 2018, secondo cui per disinformazione debba intendersi «un'informazione rivelatasi falsa o fuorviante concepita, presentata e diffusa a scopo di lucro o per ingannare intenzionalmente il pubblico, e che può arrecare un pregiudizio pubblico» (Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Contrastare la disinformazione online: un approccio europeo*, 26 aprile 2018, COM(2018) 236 fin.).

[7] G. MORGESE, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, in M. MESSINA (a cura di), *Cittadinanza e stato di diritto per un'unione europea più forte*, Napoli, 2024, p. 89 ss.

svolgimento della competizione elettorale[8]. La stessa Commissione europea ha messo in guardia dal fatto che, nel contesto delle elezioni, «la creazione e l'uso diffuso di sistemi di intelligenza artificiale […] potrebbero comportare specifici rischi sistemici»[9]. Ne è un esempio quanto accaduto a fine 2024 in Romania, dove la Corte costituzionale ha annullato il primo turno delle elezioni presidenziali sostenendo che le attività di propaganda e disinformazione condotte sulla piattaforma *TikTok* grazie al sostegno di soggetti terzi riconducibili alla Russia avessero avuto l'effetto di manipolare la volontà degli elettori[10].

La consapevolezza della facilità con cui le tecnologie digitali possano essere utilizzate per destabilizzare le istituzioni democratiche non implica che l'intelligenza artificiale debba essere demonizzata, o vietata *tout court*. Se è vero, come si è visto, che i sistemi di IA possano favorire la diffusione e persino la creazione di notizie false, è altrettanto vero che tali strumenti possano contribuire anche ad arginare il fenomeno della disinformazione, o perlomeno a ridurne le conseguenze dannose[11]. Sotto questo profilo, ad esempio, i meccanismi di IA, attraverso sofisticati modelli di analisi del dato testuale, possono aiutare a rintracciare i contenuti potenzialmente falsi e ingannevoli pubblicati nella rete. Tale circostanza impone, a livello regolatorio, un'attenta ponderazione degli interessi in gioco, affinché, da un lato, non siano ostacolati lo sviluppo e la progettazione dei sistemi di IA, dall'altro, ne siano impediti eventuali utilizzi distorsivi per finalità discriminatorie, se non addirittura sovversive.

Si tratta, come appare evidente, di una sfida piuttosto complessa da affrontare, non solo perché presuppone, da parte del legislatore, un delicato bilanciamento tra esigenze contrapposte, ma soprattutto perché sconta una debolezza strutturale nel sistema dell'Unione europea, ossia l'assenza, in capo a quest'ultima, di una competenza esplicita in materia di disinformazione: una carenza che, vale la pena ricordarlo, non può essere superata per il tramite dell'art. 2 TUE, giacché tale disposizione non costituisce, come già evidenziato in

---

[8] L. CALIFANO, F. FABBRIZZI, G. SARTOR, *Information disorder e sistema democratico*, in *federalismi.it*, n. 15, 2025, p. iv ss. In generale, sul tema, cfr. A. NICITA, *Il mercato delle verità: come la disinformazione minaccia la democrazia*, Bologna, 2024.

[9] 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, par 36.

[10] La vicenda ha suscitato ampio interesse sia a livello mediatico che accademico, poiché chiama in causa due aspetti essenziali per la tenuta degli ordinamenti democratici: da un lato, l'esigenza di salvaguardare il legittimo interesse di uno Stato sovrano di impedire ingerenze straniere nei propri affari interni senza, al contempo, svuotare di significato la sovranità popolare; dall'altro, la consapevolezza della fragilità dei sistemi democratici di fronte alle minacce poste dalle tecnologie digitali. Per un'analisi del caso rumeno, cfr., *ex multis*, D. A. CĂRĂMIDARIU, A. VERTEŞ-OLTEANU, *Safeguarding democracy: constitutional insights from Romania's election annulment*, in *Diritto pubblico comparato ed europeo*, 2025, p. 139 ss.; F. ROSA, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025, p. 62 ss.; A. IANNOTTI DELLA VALLE, *Libertà di espressione e valori democratici alla prova dei social media: il DSA e un nuovo caso TikTok europeo*, in *federalismi.it*, n. 13, 2025, p. 84 ss.; B. SELEJAN-GUTAN, *The Second Round that Wasn't: Why The Romanian Constitutional Court Annulled the Presidential Elections*, in *Verfassungsblog*, 7 December 2024.

[11] Al riguardo, CONTISSA e GALLI hanno parlato di «disinformazione aumentata», intendendo, con tale espressione, «l'insieme delle trasformazioni che il fenomeno della disinformazione subisce in conseguenza dell'integrazione dell'IA nei processi di produzione, diffusione e gestione dei contenuti informativi. L'IA agisce come un potente moltiplicatore, capace sia di rendere la disinformazione più veloce, personalizzata e convincente, sia di offrire strumenti efficaci per contrastarla» (G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025, p. 131).

dottrina, una base giuridica attributiva di competenza[12]. Questi limiti, tuttavia, non hanno impedito all'Unione di intervenire nel settore *de quo* nel solco della c.d. via europea all'evoluzione tecnologica, ossia un modello di sviluppo che, pur incoraggiando la transizione digitale, non intende abdicare al substrato di valori condiviso a livello sovranazionale[13].

È in questo scenario che si collocano le iniziative normative intraprese dall'Unione europea in materia di lotta alla disinformazione[14], la quale, pur senza essere oggetto di una regolamentazione specifica, è venuta in rilievo nell'ambito di taluni strumenti legislativi adottati nel contesto del "decennio digitale europeo"[15]. Ci si riferisce, in particolar modo, al *Digital Services Act* (DSA)[16] e all'*AI Act*[17], i quali, sebbene caratterizzati da una matrice prevalentemente economica[18], mirano a garantire che lo sviluppo e la gestione di specifici servizi e prodotti digitali nel mercato interno – sia pur nella sua dimensione di "mercato unico digitale" – non comporti un'erosione dei valori fondamentali dell'Unione[19]. Dal momento che, come si è già avuto modo di notare, la disinformazione può rappresentare un fattore di rischio per la solidità dello Stato di diritto e dei processi democratici, pilastri del sistema valoriale alla base dell'ordinamento dell'Unione, non sorprende che tale aspetto abbia trovato spazio negli atti normativi menzionati.

---

[12] 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? Diritti fondamentali, dati personali e regolazione*, vol. I, Bologna, 2022, p. 127 ss.

[13] A. ADINOLFI, *Evoluzione tecnologica e tutela dei diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *I Post di AISDUE*, 14 marzo 2023, p. 321 ss.; F. FERRI, *Transizione digitale e valori fondanti dell'Unione: riflessioni sulla costituzionalizzazione dello spazio digitale europeo*, in *Il Diritto dell'Unione Europea*, 2022, p. 277 ss.

[14] In generale sul contrasto alla disinformazione *online* nell'Unione europea, cfr. 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*, 2025, p. 272 ss.; J. BAYER, *The EU policy on disinformation: aims and legal basis*, in *Journal of Media Law*, n. 1, 2024, p. 18 ss.; S. SASSI, *L'Unione Europea e la lotta alla disinformazione* online, in *federalismi.it*, n. 3, 2023, p. 183 ss.

[15] Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Bussola per il digitale 2030: il modello europeo per il decennio digitale*, 9 marzo 2021, COM(2021) 118 fin.

[16] 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, in *GUUE* L 277 del 27 ottobre 2022, p. 1 ss.

[17] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828, in *GUUE* L 1689 del 12 luglio 2024, p. 1 ss.

[18] Ne è una riprova il fatto che entrambi i regolamenti – al pari, peraltro, di tutti gli altri adottati nell'ambito del mercato unico digitale, con la sola eccezione della direttiva 2024/2831 sul miglioramento delle condizioni di lavoro mediante piattaforme digitali, fondata sull'art. 153, par. 2, lett. *b*), TFUE – si basino sull'art. 114 TFUE relativo al ravvicinamento delle normative nazionali aventi ad oggetto l'instaurazione ed il funzionamento del mercato interno, cui si aggiunge in qualche caso, *AI Act* compreso, l'art. 16 TFUE per i profili riguardanti il trattamento dei dati personali. Sull'utilizzo di tali disposizioni per l'adozione di atti normativi nel settore digitale, cfr. S. POLI, *Il rafforzamento della sovranità tecnologica europea e il problema delle basi giuridiche*, in *I* Post *di AISDUE*, 20 dicembre 2021, p. 69 ss. In generale sul ricorso all'art. 114 TFUE come base giuridica, si rinvia a T. M. MOSCHETTA, *Il ravvicinamento delle normative nazionali per il mercato interno. Riflessioni sul sistema delle fonti alla luce dell'art. 114 TFUE*, Bari, 2018.

[19] In tal senso, con riferimento all'*AI Act*, cfr. R. PALLADINO, *L'"approccio europeo" al contrasto alla disinformazione digitale e alla protezione dei valori democratici: quale contributo dell'*AI Act*?*, in *Freedom, Security & Justice: European Legal Studies*, n. 2, 2025, p. 296 ss.; F. M. MANCIOPPI, *La regolamentazione dell'intelligenza artificiale come opzione per la salvaguardia dei valori fondamentali dell'UE*, in *federalismi.it*, n. 7, 2024, p. 112 ss.

Il contrasto alla disinformazione costituisce, non a caso, un elemento centrale degli obblighi di *compliance* incombenti alle piattaforme digitali e ai motori di ricerca di dimensioni molto grandi ai sensi del DSA[20], nella convinzione che i *social network*, consentendo la rapida trasmissione dei contenuti a un numero potenzialmente enorme di utenti, possano essere utilizzati per «diffondere o amplificare contenuti fuorvianti o ingannevoli», suscettibili di produrre effetti negativi su questioni di interesse generale per la collettività, quali la salute pubblica, il pluralismo dei media e l'integrità delle consultazioni elettorali[21]. In ragione di ciò, secondo quanto sancito dall'art. 34, siffatte piattaforme sono tenute ad individuare, analizzare e valutare con diligenza i *rischi sistemici* derivanti dai loro servizi, compresi gli «eventuali effetti negativi, attuali o prevedibili, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica», apportando, laddove necessario, adeguate misure correttive, le quali possono interessare, *inter alia*, il funzionamento dei relativi sistemi algoritmici e di raccomandazione[22]. Il ruolo del DSA nel contrastare la disinformazione nella sfera digitale è già stato ampiamente discusso in dottrina[23]. In questa sede, pertanto, si intende porre l'attenzione sul solo *AI Act*[24], al fine di verificare se, e in che misura, tale regolamento contribuisca ad affrontare i rischi correlati alla disinformazione *online*[25].

2. *Brevi cenni sulla struttura e sulle finalità dell'*AI Act

Si è già osservato come l'approccio dell'Unione europea al progresso tecnologico intenda collocare gli sviluppi inerenti al completamento del mercato unico digitale entro un solido orizzonte valoriale, al fine di evitare che la progettazione, lo scambio e l'utilizzo (in

[20] Ai sensi dell'art. 33 del DSA, per piattaforme digitali e motori di ricerca di dimensioni molto grandi si intendono gli operatori con un bacino medio mensile di destinatari del servizio nell'Unione almeno pari a 45 milioni.

[21] Regolamento (UE) 2022/2065, cit., considerando n. 84.

[22] Per una panoramica di tali profili del DSA, cfr. L. D'AGOSTINO, *Disinformazione e obblighi di compliance degli operatori del mercato digitale alla luce del nuovo Digital Services Act*, in *MediaLaws*, n. 2, 2023, p. 33 ss.

[23] Cfr. E. LONGO, *Libertà di informazione e lotta alla disinformazione nel Digital Services Act*, in *Giornale di diritto amministrativo*, 2023, p. 737 ss.; A. STROWEL, J. DE MEYERE, *The Digital Services Act: transparency as an efficient tool to curb the spread of disinformation on online platforms?*, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, n. 1, 2023, p. 66 ss.; G. CAGGIANO, *Il contrasto alla disinformazione tra nuovi obblighi delle piattaforme online e tutela dei diritti fondamentali nel quadro del Digital Service Act e della co-regolamentazione*, in *Papers di diritto europeo*, 2021, p. 45 ss.

[24] L'*AI Act* è stato oggetto di numerosi studi, che ne hanno indagato i vari profili d'interesse. Per un inquadramento generale del regolamento e delle sue implicazioni più rilevanti sotto il profilo giuridico, cfr., *ex multis*, M. ALMADA, N. PETIT, *The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights*, in *Common Market Law Review*, 2025, p. 850 ss.; F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, O. POLLICINO, *La disciplina dell'intelligenza artificiale*, Milano, 2025, p. 101 ss.; F. FERRI (a cura di), *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, in *Rivista Quaderni AISDUE*, fascicolo speciale n. 2, 2024; R. PETRUSO, G. SMORTO, *Il Regolamento europeo sull'intelligenza artificiale: una prima lettura*, in *La Nuova Giurisprudenza civile commentata*, 2024, p. 989 ss.; C. GRIECO, *Intelligenza Artificiale e tutela degli utenti nel diritto dell'Unione europea*, Napoli, 2023, spec. p. 59 ss.; N. T. NIKOLINAKOS, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies – The AI Act*, Cham, 2023; G. CONTALDI, *La proposta di regolamento sull'intelligenza artificiale e la protezione di dati personali*, in G. CAGGIANO, G. CONTALDI, P. MANZINI (a cura di), *Verso una legislazione europea su mercati e servizi digitali*, Bari, 2022, p. 205 ss.

[25] Per una panoramica sul contributo dell'*AI Act* in materia di contrasto alla disinformazione, cfr. M. MESARČÍK, N. SLOSIAROVÁ, *Regulating AI for a truthful tomorrow: addressing disinformation in the EU Artificial Intelligence Act*, in *International Journal of Law and Information Technology*, vol. 33, 2025, p. 1 ss.

questo caso) dei sistemi di intelligenza artificiale possa seguire traiettorie eterogenee all'interno dei vari Paesi membri. Tale dialettica costituisce, invero, una prassi consolidata nell'azione legislativa dell'Unione, stante la consueta tendenza del legislatore europeo – riscontrabile soprattutto in materia di sicurezza dei prodotti – a far interagire le esigenze di armonizzazione del mercato interno con la protezione dei diritti fondamentali e di taluni interessi generali, quali l'ambiente e la salute pubblica, nell'ottica, da un lato, di realizzare quell'economia sociale di mercato invocata dall'art. 3, par. 3, TUE, dall'altro, di dare attuazione alle numerose clausole trasversali previste dai Trattati. Ciò detto, è chiaro che i pericoli posti dai sistemi di intelligenza artificiale, ancora parzialmente ignoti, impongano di rimodulare tale schema normativo, inquadrando la regolamentazione dell'IA secondo una prospettiva complessivamente orientata alla prevenzione e alla gestione del rischio[26].

Questa consapevolezza si è riflessa nella previsione di un articolato schema normativo, per cui le regole applicabili ai vari sistemi di intelligenza artificiale nell'ambito dell'*AI Act* non sono uniformi, bensì differiscono in ragione dei rischi che detti sistemi sono suscettibili di produrre sui diritti fondamentali e sui valori fondanti dell'Unione. Sebbene l'inclusione dei diritti fondamentali tra gli obiettivi di un atto normativo adottato dall'Unione non costituisca certo una novità, giacché il loro rispetto vincola le istituzioni europee nell'esercizio delle loro funzioni[27], l'*AI Act* inaugura una prospettiva inedita. In questo caso, infatti, lungi dal costituire – in continuità con la disciplina prevista abitualmente in materia di sicurezza dei prodotti – soltanto uno dei requisiti di cui tenere conto in fase di progettazione[28], i diritti fondamentali, nonché i valori di cui all'art. 2 TUE, configurano lo «scheletro» dell'intero *corpus* normativo[29], configurando il parametro cui associare la soglia di tollerabilità del rischio di un dato sistema di IA[30].

Abbracciando questa logica, il regolamento introduce, nei riguardi sia dei produttori che degli utilizzatori (*deployer*)[31], una serie di divieti e obblighi procedurali di intensità variabile,

---

[26] Sotto questo profilo, con riferimento all'*AI Act*, cfr. M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno*, in F. FERRI (a cura di), *op. cit.*, p. 71 ss.

[27] Come affermato dalla Corte nelle note sentenze *Ungheria* e *Polonia* sul meccanismo di condizionalità relativo allo Stato di diritto, l'art. 2 TUE, lungi dal costituire «una mera enunciazione di orientamenti o di intenti di natura politica […] contiene valori che […] fanno parte dell'identità stessa dell'Unione quale ordinamento giuridico comune, valori che sono concretizzati in principi che comportano obblighi giuridicamente vincolanti per gli Stati membri (sentenza della Corte (Seduta Plenaria) del 16 febbraio 2022, causa C-156/21, *Ungheria* c. *Parlamento e Consiglio*, ECLI:EU:C:2022:97, punto 232; sentenza della Corte (Seduta Plenaria) del 16 febbraio 2022, causa C-157/21, *Polonia* c. *Parlamento e Consiglio*, ECLI:EU:C:2022:98, punto 264.

[28] In tal senso, v. ad esempio il regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio, in *GUUE* L 117 del 5 maggio 2017, p. 1 ss., il cui considerando n. 89 recita: «Il presente regolamento rispetta i diritti fondamentali e i principi riconosciuti, in particolare, dalla Carta e, segnatamente, la dignità umana, l'integrità della persona, la protezione dei dati di carattere personale, la libertà delle arti e delle scienze, la libertà d'impresa e il diritto di proprietà».

[29] E. CIRONE, *L'*AI Act *e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in F. FERRI (a cura di), *op. cit.*, p. 61.

[30] F. FERRI, *Transizione digitale e valori fondanti dell'Unione*, cit., p. 319.

[31] Si tratta dei principali destinatari degli obblighi stabiliti dall'*AI Act*. In particolare, per fornitore si intende «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un modello di IA per finalità generali o che fa sviluppare un sistema di IA o un modello di IA per finalità generali e immette tale sistema o modello sul mercato o mette in servizio il sistema di IA con il proprio nome o marchio, a titolo oneroso o gratuito»; mentre il *deployer* è «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso

a seconda della pericolosità attesa di ciascun sistema di IA. In sostanza, i sistemi di IA sono classificati in categorie distinte sulla base del livello di rischio che li caratterizza. Tale operazione presuppone un'analisi *ex ante* degli interessi in rilievo, allo scopo di valutare se le possibili conseguenze negative di un'applicazione di intelligenza artificiale siano tali da giustificare restrizioni più o meno rigorose alla loro immissione sul mercato o al loro utilizzo[32].

Com'è facile immaginare, un impianto normativo così concepito non può tollerare lo sviluppo di sistemi di IA palesemente contrari ai diritti previsti dalla Carta e ai valori di cui all'art. 2 TUE, giacché nessuna misura correttiva adottata *ex post* sarebbe idonea a compensarne gli effetti dannosi. Nei confronti di tali sistemi, i cui rischi per la salute, la sicurezza e i diritti fondamentali sono ritenuti "inaccettabili", vige, pertanto, un divieto assoluto, sancito dall'art. 5 del regolamento 2024/1689. A un livello inferiore si situano i sistemi ad alto rischio, la cui regolazione costituisce la parte più significativa della disciplina introdotta dall'*AI Act*. Tale categoria comprende quei sistemi il cui livello di pericolosità, pur non essendo sufficiente a determinarne la proibizione, è tale da far sorgere, in capo ai fornitori, ai distributori e agli utilizzatori, specifici obblighi di *compliance*, al fine di minimizzarne i rischi attesi. Le disposizioni meno stringenti, di fatto limitate a meri obblighi di trasparenza e informazione, cui si aggiunge l'adesione volontaria a codici di condotta, si applicano, invece, ai sistemi a rischio minimo o limitato. Una disciplina a sé stante è riservata, infine, ai sistemi di IA "per scopi generali", i quali, potendo trovare applicazione per una molteplicità di compiti e funzioni, potrebbero comportare rischi sistemici, compresi, *inter alia*, «eventuali effetti negativi, effettivi o ragionevolmente prevedibili, sui processi democratici e sulla sicurezza pubblica ed economica», nonché «la diffusione di contenuti illegali, mendaci o discriminatori»[33].

Viste le finalità del presente scritto, occorre verificare in che modo tale approccio basato sul rischio possa venire in rilievo in materia di contrasto alla disinformazione posto che, come recita il considerando n. 133, «i nuovi rischi di cattiva informazione e manipolazione su vasta scala» costituiscono una delle minacce cui l'*AI Act* intende far fronte.

3. *Le pratiche vietate ai sensi dell'art. 5: quale rilievo ai fini del contrasto alla disinformazione* online*?*

Seguendo la struttura del regolamento in esame, il primo livello su cui occorre soffermarsi riguarda le pratiche vietate ai sensi dell'art. 5. Per i nostri fini rileva, nello specifico, l'ipotesi prevista dal par. 1, lett. *a*), la quale vieta «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,

in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale» (regolamento (UE) 2024/1689, cit., art. 3, par. 3 e 4).

[32] Su tali aspetti, cfr. C. SCHEPISI, *Le "dimensioni" della regolazione dell'intelligenza artificiale nella proposta di regolamento della Commissione*, in *I Post di AISDUE*, 28 marzo 2022, p. 330 ss., spec. p. 339 ss.

[33] Regolamento (UE) 2024/1689, cit., considerando n. 110.

a un'altra persona o a un gruppo di persone un danno significativo». La fattispecie sembrerebbe includere *prima facie* i sistemi di raccomandazione utilizzati dalle piattaforme digitali, nel caso in cui tali sistemi, favorendo la circolazione di contenuti disinformativi, possano indurre gli utenti in errore. Viene da chiedersi, pertanto, se siffatti sistemi rientrino nell'ambito di applicazione della disposizione in parola, dal momento che la norma non proibisce le tecniche subliminali *tout court*, ma soltanto quelle che rispondano alle caratteristiche espressamente previste.

Innanzitutto, occorre evidenziare che la disposizione distingue tra "tecniche subliminali" e "tecniche volutamente manipolative o ingannevoli", senza fornire, tuttavia, indicazioni utili a comprendere in cosa consistano tali applicazioni di IA. Per quanto riguarda le tecniche subliminali, un ausilio a fini interpretativi perviene dal considerando n. 29 del regolamento 2024/1689, il quale menziona le tecniche di manipolazione basate sull'IA che impiegano «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 sia in grado di controllarle o resistervi o possa evitare l'inganno». La definizione sembrerebbe riferirsi, in sostanza, a tutti i sistemi di IA in grado di influenzare le decisioni e i comportamenti dei soggetti in modo subdolo, ossia senza che questi ultimi ne abbiano coscienza, compromettendone la capacità di agire autonomamente e consapevolmente.

Appare invece meno agevole capire cosa contraddistingua le "tecniche volutamente manipolative o ingannevoli". L'espressione sembra porre l'enfasi sull'intenzione del produttore (o utilizzatore) di un sistema di IA di ingannare deliberatamente gli utenti, sfruttandone i possibili fattori di vulnerabilità con l'obiettivo di influenzarne o controllarne il comportamento. Ciò posto, la distinzione rispetto alle semplici tecniche subliminali rimane ambigua, dal momento che, alla luce della formulazione del già citato art. 5, la volontà del fornitore o del *deployer* di distorcere il comportamento delle persone non rileva al fine di accertare l'illeceità di una determinata pratica, essendo sufficiente provare il compimento dell'effetto manipolatorio. In effetti, come precisato nelle linee guida pubblicate dalla Commissione sulle pratiche di intelligenza artificiale vietate ai sensi dell'*AI Act*[34], ciò che accomuna le varie tecniche proibite risiede nel risultato che esse intendono perseguire, ossia indurre gli utenti a tenere un comportamento o a prendere una decisione che altrimenti non avrebbero assunto.

Quest'ultimo rappresenta l'aspetto più controverso della disposizione *de qua*, poiché presuppone che il comportamento umano in questione costituisca una diretta conseguenza della manipolazione effettuata tramite l'utilizzo dell'intelligenza artificiale. Pertanto, nel caso in cui l'IA si limitasse a condizionare in modo generico l'operato dei soggetti, il divieto *ex* art. 5 non troverebbe applicazione, dovendosi riscontrare, viceversa, «un impatto sostanziale sul comportamento» di una persona tale da comprometterne l'autonomia e la libertà di scelta[35]. Detto altrimenti, soltanto l'esistenza di un nesso causale tra la tecnica subliminale utilizzata e l'alterazione del comportamento umano renderebbe illegittima siffatta applicazione di IA. Si tratta, come risulta evidente, di un punto piuttosto problematico della disposizione in parola, suscettibile di restringerne sensibilmente la sfera d'efficacia.

---

[34] Comunicazione della Commissione, *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA)*, 29 luglio 2025, C(2025) 5052 fin.
[35] *Ibidem*, par. 76.

Con riferimento ai profili d'interesse in questa sede, il divieto si estenderebbe ai sistemi di IA in grado di incidere in modo significativo sulle preferenze di voto degli elettori. Al riguardo, vale la pena soffermarsi sulla vicenda rumena evocata in precedenza. In quel caso, infatti, la Corte costituzionale ha sostenuto che la diffusione *online* di contenuti disinformativi su tematiche di rilievo politico (di cui era peraltro taciuta la matrice elettorale), agevolata dai sistemi di raccomandazione di TikTok, avesse impedito agli elettori di ricevere informazioni puntuali sui candidati, così da formarsi liberamente un'opinione senza essere influenzati in modo illecito e sproporzionato. Tale ragionamento sembrava sottintendere una relazione diretta tra la campagna disinformativa condotta sfruttando gli algoritmi della piattaforma e il risultato elettorale.

Ciò detto, non si può non considerare che la decisione della Corte costituzionale sia maturata in uno scenario di grande incertezza politica, giacché, appena due giorni prima della pronuncia, un *dossier* dei servizi di *intelligence* aveva rilevato interferenze russe durante la campagna elettorale. Tale circostanza ha indotto il giudice costituzionale ad intervenire *ex officio* prima che si svolgesse il secondo turno delle elezioni – in programma quattro giorni dopo la pubblicazione del *dossier* – penalizzando la solidità dell'argomentazione giuridica[36]. Pur costituendo un precedente di rilievo, la vicenda non consente, proprio per le circostanze specifiche che l'hanno accompagnata, di pervenire a facili conclusioni. A livello generale, infatti, ci sembra improbabile che la divulgazione di informazioni false possa configurare un fattore di distorsione del comportamento nell'accezione accolta dall'art. 5, par. 1, lett. *a)*, dell'*AI Act*. Del resto, pure di fronte a un'intensa strategia disinformativa, sarebbe possibile sostenere che, in assenza di tali azioni intrusive, il risultato elettorale sarebbe stato diverso? Per quanto verosimile in linea teorica, rimane un'ipotesi difficile da dimostrare con certezza, o, quantomeno, con un certo grado di plausibilità.

È pur vero, come suggeriscono le linee guida della Commissione[37], che la nozione di «distorsione materiale del comportamento» possa essere interpretata in parallelo agli effetti distorsivi causati dalle pratiche commerciali sleali, ingannevoli e aggressive di cui agli artt. 5-9 della direttiva 2005/29/CE[38] nei confronti dei consumatori. Orbene, come precisato dalla Corte di giustizia, per essere ritenuta ingannevole è sufficiente che una pratica *sia idonea* a indurre il consumatore ad assumere una decisione commerciale che altrimenti non avrebbe preso, anche qualora il comportamento del soggetto non sia stato effettivamente falsato nel caso specifico[39]. A tal proposito, peraltro, la Corte ha avuto modo di evidenziare le criticità derivanti dalle pratiche che mirano a «sfruttare l'effetto psicologico creato nella mente del consumatore […]»[40]. Tale aspetto appare particolarmente significativo nel caso dei sistemi di

---

[36] Oltre a ciò, la sentenza ha ricevuto critiche, sotto il profilo procedurale, per l'intervento *ex officio* anziché su istanza di parte, per le ristrette tempistiche con cui la procedura si è svolta e per il mancato coinvolgimento delle parti interessate. Su questo dibattito, si rinvia a F. ROSA, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, cit.

[37] *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate*, cit., par. 80 ss.

[38] Direttiva 2005/29/CE del Parlamento europeo e del Consiglio, dell'11 maggio 2005, relativa alle pratiche commerciali sleali tra imprese e consumatori nel mercato interno e che modifica la direttiva 84/450/CEE del Consiglio e le direttive 97/7/CE, 98/27/CE e 2002/65/CE del Parlamento europeo e del Consiglio e il regolamento (CE) n. 2006/2004 del Parlamento europeo e del Consiglio, in *GUCE* L 149 dell'11 giugno 2005, p. 22 ss.

[39] Sentenza della Corte di giustizia (Quinta Sezione) del 26 ottobre 2016, *Canal Digital Danmark A/S*, causa C-611/14, ECLI:EU:C:2016:800, punto 37.

[40] Sentenza della Corte (Sesta Sezione) del 18 ottobre 2012, *Purely Creative e a.*, causa C-428/11, ECLI:EU:C:2012:651, punto 49.

raccomandazione utilizzati dalle piattaforme, i quali, come ricordato in precedenza, fanno leva sulla partecipazione emotiva degli utenti, valorizzando i contenuti più coinvolgenti, suscettibili di plasmare le percezioni degli individui senza che ne abbiano consapevolezza. Questi sviluppi giurisprudenziali in materia di protezione dei consumatori offrono un riferimento cui la Corte potrebbe attingere qualora fosse chiamata, ad esempio nell'ambito di un rinvio pregiudiziale *ex* art. 267 TFUE, a chiarire la nozione di «distorsione materiale del comportamento» ai sensi dell'art. 5, par. 1, lett. *a*). Tuttavia, dal momento che le preferenze di voto sono spesso volubili, rimarrebbe comunque complesso dimostrare l'esistenza di un nesso causale tra la pratica di IA impiegata e l'alterazione del comportamento elettorale.

In ogni caso, anche qualora un simile nesso causale fosse verificabile, i sistemi di raccomandazione delle piattaforme rientrerebbero difficilmente nel perimetro applicativo dell'art. 5 del regolamento 2024/1689. Ciò perché, secondo il dettato di detta disposizione, la causazione di un «danno significativo», o, quantomeno, la circostanza per cui siffatto scenario risulti «ragionevolmente» prevedibile, rappresenta una *conditio sine qua non* affinché una specifica tecnica manipolatoria basata sull'IA debba essere vietata. Il già citato considerando n. 29 precisa che per "danni significativi" debbano intendersi gli «effetti negativi sufficientemente importanti sulla salute fisica, psicologica o sugli interessi finanziari». Per i nostri fini, l'unica tipologia potenzialmente rilevante è costituita dai danni di tipo psicologico, in quanto correlati ai sistemi di intelligenza artificiale «che sfruttano le vulnerabilità cognitive ed emotive e influenzano i comportamenti di una persona»[41]. Come precisano le linee guida pubblicate dalla Commissione, tali danni comprendono gli «effetti negativi sulla salute mentale e sul benessere psicologico ed emotivo di una persona»[42]: sebbene la circolazione di contenuti disinformativi possa inficiare la capacità delle persone di distinguere il vero dal falso, ci sembra difficile che l'applicazione dell'IA per scopi disinformativi possa comportare un danno nel senso inteso dalla disposizione in parola[43].

Ricapitolando, vista la necessità di provare la contestuale presenza dei requisiti analizzati, appare improbabile che l'utilizzo dell'intelligenza artificiale per finalità di disinformazione – specialmente tramite i sistemi di raccomandazione – possa configurare una pratica vietata ai sensi dell'art. 5 dell'*AI Act*[44].

4. *Il contrasto alla disinformazione nell'ambito dei sistemi di intelligenza artificiale ad alto rischio*

Ai fini del contrasto alla disinformazione *online*, vale la pena soffermarsi anche sulla regolamentazione riservata ai sistemi di intelligenza artificiale ad alto rischio di cui all'art. 6 dell'*AI Act*. Come ricordato in precedenza, la disciplina di tali sistemi costituisce la parte più significativa dell'approccio basato sul rischio introdotto dal regolamento 2024/1689. Le

41 *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate*, cit., par. 88.
42 *Ibidem*.
43 Si noti che, durante la fase di discussione della proposta di regolamento sull'IA presentata dalla Commissione, il Comitato economico e sociale europeo ha suggerito di includere, tra le pratiche vietate *ex* art. 5, par. 1, lett. *a*), quelle che causano «pregiudizio ai diritti fondamentali, compresi quelli alla salute e alla sicurezza fisiche o psicologiche, di un'altra persona o di un gruppo di persone, o alla democrazia e allo Stato di diritto» (parere del Comitato economico e sociale europeo sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione, COM(2021) 206 fin., 2021/0106 (COD), par. 4.3). Tale modifica, tuttavia, non è stata accolta dal legislatore.
44 In tal senso, M. MESARČÍK, N. SLOSIAROVÁ, *Regulating AI for a truthful tomorrow*, cit., p. 8.

applicazioni di IA ricomprese in questa categoria presentano rischi elevati per la salute, la sicurezza e i diritti fondamentali: ragion per cui l'immissione sul mercato unico, la messa in servizio e l'utilizzo di tali applicazioni sono subordinati al rispetto di determinati requisiti tecnici, cui si accompagnano specifici obblighi di *compliance* posti in capo ai fornitori e agli utilizzatori. Il regolamento non intende bandire siffatti sistemi dal mercato, bensì ridurne i rischi potenziali, cosicché la loro progettazione e distribuzione non comporti pericoli eccessivi per gli individui e la collettività.

Per quanto d'interesse in questa sede, rileva il contenuto dell'allegato III, il quale, dando seguito all'art. 6, par. 2, del regolamento *de quo*, elenca, distinguendoli per settore di applicazione, una serie di sistemi di IA ritenuti ad alto rischio[45]. La sezione dedicata all'amministrazione della giustizia e ai processi democratici menziona, *inter alia*, i sistemi «destinati a essere utilizzati per influenzare l'esito di un'elezione o di un referendum o il comportamento di voto delle persone fisiche nell'esercizio del loro voto alle elezioni o ai referendum»[46]. La qualificazione di tali sistemi come ad alto rischio deriva dalla necessità di prevenire eventuali effetti negativi sulla democrazia e sullo Stato di diritto, nonché «i rischi di indebite interferenze esterne sul diritto di voto» sancito dall'art. 39 della Carta[47].

Si noti, tuttavia, che la riconducibilità di uno strumento di IA a una delle ipotesi descritte dall'allegato III non ne determina *ipso facto* la classificazione come sistema ad alto rischio, giacché, per effetto dell'art. 6, par. 3, tale dicitura è limitata ai sistemi che presentino un «rischio significativo di danno» per la salute, la sicurezza o i diritti fondamentali delle persone fisiche[48], valutato alla luce della gravità del possibile danno, nonché della probabilità che esso si verifichi[49]. Si tratta, come appare evidente, di un criterio piuttosto sfuggente, giacché la valutazione di "significatività" non si basa sulla gravità di un danno effettivo, bensì sulla mera prevedibilità di quest'ultimo. Ne consegue che i fornitori dispongano di un apprezzabile margine di discrezionalità nel valutare se il loro sistema rientri nell'ambito di applicazione della deroga *ex* art. 6, par. 3, dal momento che la decisione di un fornitore di invocare tale clausola, pur dovendo essere documentata, non è soggetta ad alcun controllo esterno da parte di un'autorità di vigilanza[50].

Venendo al regime normativo applicabile, le tecniche di IA suscettibili di interferire con i processi elettorali sono soggette, al pari degli altri sistemi inclusi nella fascia di rischio

---

[45] Ai sensi dell'art. 6, par. 1, sono inoltre considerati ad alto rischio i sistemi di IA destinati a essere utilizzati come componente di sicurezza di un prodotto, o che costituiscano essi stessi un prodotto, già disciplinati dalla normativa di armonizzazione dell'Unione elencata nell'allegato I.

[46] Regolamento (UE) 2024/1689, cit., allegato III, par. 8, lett. *b*). Rimangono esclusi dalla qualifica di sistemi ad alto rischio i «sistemi di IA ai cui output le persone fisiche non sono direttamente esposte, come gli strumenti utilizzati per organizzare, ottimizzare e strutturare le campagne politiche da un punto di vista amministrativo e logistico».

[47] *Ibidem*, considerando n. 62.

[48] La disposizione precisa che un sistema non possa essere ritenuto ad alto rischio se *non* influenza «materialmente il risultato del processo decisionale»; viceversa, «è sempre considerato ad alto rischio qualora esso effettui profilazione di persone fisiche». Per una lettura critica del concetto di "danno" come soglia per la protezione dei diritti fondamentali, cfr. L. GROZDANOVSKI, J. DE COOMAN, *Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union*, in *Rutgers Computer and Technology Law Journal*, n. 2, 2023, p. 207 ss.

[49] Regolamento (UE) 2024/1689, cit., considerando n. 52.

[50] *Ibidem*, art. 6, par. 4. A tal proposito, si ricorda che, in linea generale, l'immissione sul mercato dei sistemi di IA ad alto rischio non è subordinata al rilascio di una licenza o a una forma di autorizzazione preventiva, bensì avviene sulla base di un'autovalutazione *ex ante* da parte del fornitore volta ad accertare la conformità del sistema da esso sviluppato ai requisiti tecnici stabiliti dal regolamento (art. 43).

elevato, al sistema di gestione dei rischi di cui all'art. 9 del regolamento 2024/1689. Tale quadro normativo, evidentemente mutuato dal meccanismo di gestione dei rischi sistemici istituito dagli artt. 34-35 del DSA, impone ai fornitori di identificare i potenziali rischi per la salute, la sicurezza e i diritti fondamentali derivanti dall'impiego del sistema di IA da essi progettato, nonché di apportare, se necessario, le opportune misure correttive[51]. Oltre a ciò, i fornitori dei sistemi di IA in questione sono tenuti, prima della loro immissione sul mercato, a valutarne la conformità rispetto a specifici *standard* di sicurezza, nonché a prevedere adeguati meccanismi di supervisione umana in grado di scongiurare eventuali usi impropri ed effetti indesiderati.

Anche i *deployer*, ossia gli utilizzatori di una tecnologia di IA, siano essi organismi di diritto pubblico o enti privati che forniscono servizi pubblici – nel nostro caso, ad esempio, i partiti politici o le agenzie di comunicazione che organizzano le campagne elettorali – sono destinatari di obblighi specifici. In particolare, ai sensi dell'art. 27 dell'*AI Act*, tali soggetti sono tenuti a valutare se il sistema di intelligenza artificiale da essi impiegato possa arrecare un pregiudizio ai diritti fondamentali[52]. Logicamente, i diritti fondamentali maggiormente attenzionati varieranno di volta in volta in ragione del settore in cui opera il sistema di IA implementato. Va da sé che, con riferimento ai sistemi volti ad influenzare l'andamento delle consultazioni elettorali, la valutazione d'impatto dovrà accertare se l'impiego di quel sistema comporti rischi significativi soprattutto in termini di pregiudizio alla libertà di espressione, alla tutela della privacy e al diritto a un'informazione libera e pluralistica. È in questa sede, pertanto, che i pericoli correlati all'utilizzo dell'intelligenza artificiale per scopi disinformativi dovrebbero essere mitigati.

Ci sembra ancora prematuro stabilire se uno schema normativo così delineato sia in grado di affrontare efficacemente i rischi per la democrazia derivanti dalle strategie disinformative messe in campo tramite l'IA. Tuttavia, non si può fare a meno di osservare che il regolamento sconti, sotto questo profilo, taluni vizi strutturali già segnalati in dottrina[53]. In particolare, occorre domandarsi se la prospettiva che ispira l'*AI Act*, basata sull'accettabilità del rischio quale criterio di riferimento, sia funzionale ad arginare il fenomeno della disinformazione *online*. Come già accennato, il regolamento muove dalla premessa secondo cui il rischio altro non sia che la combinazione della probabilità ed entità dell'eventuale danno: di conseguenza, affinché il livello di rischio atteso sia tollerabile, è necessario che tali fattori rispettino uno *standard* minimo accettabile. Una simile impostazione – debitrice degli strumenti di valutazione del rischio elaborati nell'ambito della legislazione in materia di sicurezza dei prodotti – presuppone che il danno eventuale risulti misurabile e prevedibile attraverso modelli statistici, nonché univoco, nel senso di costituire l'unico possibile effetto indesiderato correlato a un determinato rischio.

Difficilmente siffatto schema potrebbe rivelarsi idoneo a contrastare efficacemente i pericoli per la democrazia connessi all'utilizzo dei sistemi di IA per ragioni disinformative. La capacità di un contenuto disinformativo di condizionare in via autonoma i processi

---

[51] Per un'analisi di tali profili del DSA, cfr. E. BIRITTERI, *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, p. 52 ss.

[52] Per approfondire la valutazione d'impatto sui diritti fondamentali prevista dall'*AI Act*, 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 Law & Security Review*, vol. 54, 2024, p. 1 ss.

[53] Cfr. M. ALMADA, N. PETIT, *The EU AI Act*, cit., p. 103 ss.; E. CIRONE, *L'*AI Act *e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, cit., p. 61 ss.

elettorali, dunque senza l'interazione di altri eventi esterni, risulta, infatti, piuttosto difficile da prevedere *a priori*, vista l'incertezza che accompagna solitamente le intenzioni di voto degli individui. Inoltre, gli effetti negativi causati dalla disinformazione *online* potrebbero concretizzarsi solamente nel lungo periodo, peraltro attraverso manifestazioni non facilmente tangibili. Ad esempio, la circolazione di informazioni manipolate o fuorvianti su questioni di interesse politico-sociale potrebbe indebolire la fiducia delle persone nei confronti delle istituzioni, minando le fondamenta dei sistemi democratici[54]. In queste circostanze, l'approccio adottato dall'*AI Act* non impedirebbe la realizzazione dell'evento dannoso, poiché quest'ultimo non sarebbe riconducibile a una specifica applicazione dell'intelligenza artificiale, bensì costituirebbe l'esito di un articolato processo di interazione uomo-macchina[55].

5. *Gli obblighi di trasparenza per i sistemi di IA a rischio minimo e le disposizioni specifiche in materia di* deep fake

Alla base della struttura piramidale predisposta dall'*AI Act* si situano i sistemi di intelligenza artificiale a rischio minimo, i quali, ai sensi dell'art. 50, sono sottoposti soltanto ad obblighi di trasparenza. Al di là delle varie ipotesi specifiche previste da detta disposizione, la normativa risponde, nel complesso, a una precisa finalità: rendere i contenuti e gli strumenti basati sull'intelligenza artificiale facilmente riconoscibili agli utenti che si trovassero ad interagirvi. A tal proposito, l'art. 50, par. 1, del regolamento 2024/1689 obbliga i fornitori dei sistemi destinati a interagire direttamente con le persone fisiche a progettare e sviluppare tali sistemi in modo tale che queste ultime «siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo». Come osservato in dottrina, tale deroga introduce un elemento di flessibilità potenzialmente suscettibile di limitare la portata degli obblighi informativi incombenti ai fornitori[56].

Per i nostri fini, in ogni caso, più delle applicazioni di IA volte ad interfacciarsi direttamente con gli utenti (i *chatbot*, ad esempio), a suscitare interesse sono soprattutto i contenuti generati tramite IA. I sistemi di intelligenza artificiale, infatti, in quanto capaci di riprodurre fedelmente luoghi e personaggi, possono essere utilizzati per raffigurare in modo verosimile situazioni in realtà mai accadute, ingannando gli utenti sull'autenticità degli eventi rappresentati: logicamente, più un contenuto realizzato tramite l'ausilio dell'IA appare realistico, più sarà difficile per i soggetti riconoscerne l'origine artificiale. Se impiegati in contesti politico-elettorali, tali strumenti, impattando negativamente «sull'integrità e sulla fiducia [delle persone] nell'ecosistema dell'informazione»[57], possono trasformarsi in potenti mezzi di manipolazione dell'opinione pubblica. Proprio in virtù di ciò, è fondamentale che i soggetti abbiano consapevolezza della natura autentica o artefatta dei contenuti con cui si

---

[54] Al riguardo, per quanto concerne i *deep fake*, cfr. R. CHESNEY, D. K. CITRON, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in *California Law Review*, 2019, p. 1753 ss.
[55] M. ALMADA, N. PETIT, *The EU AI Act*, cit., p. 108.
[56] R. PALLADINO, *L'"approccio europeo" al contrasto alla disinformazione digitale e alla protezione dei valori democratici*, cit., p. 313.
[57] Regolamento (UE) 2024/1689, cit., considerando n. 133.

interfacciano. A tal fine, i fornitori dei sistemi di IA che generano contenuti audio, immagini, video o testuali sintetici sono tenuti a marcare «in un formato leggibile meccanicamente» gli *output* prodotti dal proprio sistema, così da consentire di rilevare se questi ultimi siano stati «generati o manipolati artificialmente»[58].

Specifica attenzione è riservata, inoltre, ai *deep fake*[59]. Secondo la nozione accolta dall'*AI Act*, l'espressione si riferisce a «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»[60]. Come spiegato dal Garante per la protezione dei dati personali, si tratta, in sostanza, di contenuti multimediali generati tramite sofisticati *software* di intelligenza artificiale capaci, attingendo a immagini e audio reali, di «modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce»[61], cosicché una persona sembri pronunciare parole o compiere azioni che, in realtà, non ha mai detto o fatto. Ciò rende tali contenuti particolarmente efficaci in contesti elettorali, in quanto possono essere creati allo scopo di screditare o, al contrario, promuovere un determinato personaggio politico, proponendo narrazioni del tutto manipolate. Assicurare la riconoscibilità dei *deep fake* da parte degli utenti non risponde, pertanto, soltanto a un generico requisito di trasparenza, bensì anche alla necessità di proteggere il funzionamento dei sistemi democratici, evitando storture nel sistema dell'informazione in grado di alterare i meccanismi di consolidamento del consenso.

Al riguardo, l'art. 50, par. 4, primo comma, dell'*AI Act*, obbliga gli utilizzatori di un sistema di IA in grado di realizzare *deep fake* a rendere noto che «il contenuto è stato generato o manipolato artificialmente»[62]: tuttavia, qualora il contenuto in questione «faccia parte di

---

[58] *Ibidem*, art. 50, par. 2.

[59] Per una prospettiva d'insieme sul fenomeno dei *deep fake* e sulla sua regolamentazione nell'ordinamento dell'Unione, cfr. M. J. BLOCK, *A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union*, in *Journal of European Consumer and Market Law*, n. 4, 2024, p. 184 ss.; G. PROIETTI, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del* deep fake, in *MediaLaws*, 2024, p. 328 ss.; A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, 2024, p. 407 ss.; M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai* deepfakes, in *MediaLaws*, 2023, p. 170 ss.

[60] Regolamento (UE) 2024/1689, cit., art. 3, par. 60. Secondo CONTISSA e GALLI, tale definizione solleva talune perplessità interpretative. In particolare, essa non chiarisce quale sia il livello di manipolazione richiesto affinché un contenuto possa essere classificato come *deep fake*: «ad esempio, nel caso delle immagini, posto che un'immagine scattata da una fotocamera digitale non mostra mai la realtà (vera), ma è il risultato della catena di elaborazione algoritmica dell'immagine e delle impostazioni scelte dal fotografo, la questione è se potrebbe basterebbe l'applicazione sull'immagine di un qualsiasi filtro digitale basato su AI (diffusissimi nei comuni smartphone) perché essa sia considerata manipolata, e quindi un deepfake, qualora ricorressero anche le altre condizioni» (G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, cit., p. 150). Anche il concetto di somiglianza non appare sufficientemente nitido: ci si chiede, al riguardo, se occorra riscontrare la somiglianza di un contenuto con un elemento effettivamente esistente, oppure se basti rilevare il *realismo* di una figura, a prescindere che essa coincida o meno con un soggetto/oggetto realmente esistente. Per approfondire tali profili, si rinvia a K. MEDING, C. SORGE, *What Constitutes a Deep Fake? The Blurry Line Between Legitimate Processing and Manipulation Under the EU AI Act*, in *Proceedings of the 2025 Symposium on Computer Science and Law*, 2025, p. 152 ss.

[61] Garante per la protezione dei dati personali, *Deepfake. Il falso che ti «ruba» la faccia e la privacy*, 28 dicembre 2020.

[62] Il medesimo obbligo si applica nei confronti del *deployer* di un sistema di IA «che genera o manipola testo pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico», a meno che il contenuto generato dall'IA sia stato sottoposto «a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione del contenuto» (regolamento (UE) 2024/1689, cit., art. 50, par. 4, secondo comma).

un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi», si applica un obbligo di trasparenza meno rigoroso, poiché limitato a rivelare l'esistenza di tale contenuto generato artificialmente, «senza ostacolare l'esposizione o il godimento dell'opera». Viste le finalità del presente scritto, vale la pena soffermarsi su tale disposizione.

Innanzitutto, occorre evidenziare che il legislatore abbia preferito non disciplinare in modo eccessivamente stringente il fenomeno dei *deep fake*. I sistemi di IA in grado di realizzarli, infatti, non sono stati né proibiti, né inseriti tra i sistemi ad alto rischio, nonostante tali contenuti possano comportare evidenti pericoli in termini di mistificazione della realtà. A nostro avviso, un generale divieto di creazione e divulgazione dei *deep fake* sarebbe stato eccessivo, giacché, proibendo qualsiasi utilizzo di tali tecnologie, ne avrebbe impedito anche eventuali applicazioni positive[63]. Del resto, ad apparire controversa non è la tecnologia in sé, di fatto neutrale, bensì gli scopi per la quale essa viene implementata[64]. Inoltre, un divieto di tale portata, censurando anche eventuali contenuti parodici o satirici, avrebbe determinato una sproporzionata restrizione della libertà di manifestazione del pensiero.

Il fatto che i *deep fake* possano costituire una minaccia per l'ordine democratico, pur non potendo configurare, di per sé, un motivo valido per vietarne qualsiasi forma di riproduzione, rende necessario regolamentarne il funzionamento in modo da limitare i possibili rischi per la collettività. Sotto questo profilo, la disciplina introdotta dall'*AI Act*, sostanzialmente basata sull'imposizione di meri obblighi di informazione in capo ai *deployer*, cui non si associano misure di *enforcement* di natura sanzionatoria[65], non pare particolarmente incisiva.

Il dovere di segnalare i *deep fake* al fine di renderli facilmente riconoscibili per gli utenti può contribuire a tenere sotto controllo il fenomeno, ma difficilmente può rappresentare un deterrente utile a scongiurarne eventuali utilizzi per scopi disinformativi[66]. In questi casi, infatti, tali contenuti sono realizzati proprio allo scopo di manipolare gli utenti: ragion per cui gli autori avranno interesse a nasconderne la natura artefatta. Com'è stato opportunamente notato altrove, se l'autore di un *deep fake* è disposto a rivelare l'origine artificiale della propria opera, è plausibile che tale operazione non celi un intento malevolo: in effetti, qualora fosse mossa da una finalità ingannevole, ad esempio poiché architettata nell'ambito di una mirata campagna di disinformazione, la diffusione di *deep fake* potrebbe essere gestita tramite strumenti informatici in grado di mascherare l'identità degli autori (come *bot* o *account* falsi), con la conseguenza di rendere piuttosto complessa l'identificazione dei soggetti in capo ai quali ricondurre la responsabilità in caso di violazione degli obblighi di notifica di cui all'art. 50, par. 4, dell'*AI Act*[67].

---

[63] In tal senso, cfr. F. R. MORENO, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in *International Review of Law, Computers & Technology*, 2024, p. 1 ss.

[64] Non è possibile, in questa sede, approfondire tutti i possibili risvolti dei *deep fake*, compresi quelli positivi. Basti ricordare che taluni studi in materia hanno evidenziato che tali tecnologie possono trovare applicazione anche per finalità creative, ad esempio in settori come il cinema, la moda e la pubblicità, nonché per scopi di sperimentazione, specialmente in ambito medico. Su questi aspetti, cfr. M. FEENEY, *Deepfake laws risk creating more problems than they solve*, Regulatory Transparency Project, 1° March 2021, reperibile *online*; J. SILBEY, W. HARTZOG, *The Upside of Deep Fakes*, in *Maryland Law Review*, 2019, p. 960 ss.

[65] In proposito, PALLADINO lamenta la mancata previsione «dell'obbligo di rimozione dei contenuti manipolati o l'introduzione di sanzioni da comminare in capo a coloro che svolgono attività di diffusione dei *deep fake* con intenti manipolatori» (R. PALLADINO, *L'"approccio europeo" al contrasto alla disinformazione digitale e alla protezione dei valori democratici*, cit., p. 315).

[66] M. LABUZ, *Regulating Deep Fakes in the Artificial Intelligence Act*, in *Applied Cybersecurity & Internet Governance*, n. 1, 2023, p. 1 ss.

[67] A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale*, cit., p. 422.

Per ovviare a questo problema, dal momento che i *deep fake* circolano prevalentemente sui *social network*, si potrebbe pensare di obbligare le piattaforme digitali a fornire indicazioni puntuali sull'origine dei contenuti pubblicati, ad esempio segnalando i casi in cui la condivisone fosse opera di un *bot*[68]. Una tale previsione normativa – da concretizzare, presumibilmente, nell'ambito del DSA – pur non rimuovendo il rischio che i *deep fake* realizzati per scopi disinformativi possano sfuggire all'obbligo di notifica, consentirebbe agli utenti di interfacciarsi in modo più consapevole con i contenuti creati dall'intelligenza artificiale.

Peraltro, il regolamento non chiarisce in che modo i *deployer* dovrebbero adempiere in concreto agli obblighi di trasparenza ad essi incombenti. L'art. 50, par. 5, si limita a specificare che la segnalazione relativa alla natura manipolata di un contenuto dovrebbe avvenire «in maniera chiara e distinguibile al più tardi al momento della prima interazione o esposizione». La notifica, in altre parole, non dovrebbe lasciare spazio a dubbi, permettendo agli utenti di cogliere immediatamente l'origine artificiale di un contenuto, né essere tardiva, giacché in tal caso l'effetto manipolatorio potrebbe essersi già concretizzato, sia pur inconsciamente. Al di là di ciò, tuttavia, la disposizione lascia ai *deployer* ampia discrezionalità circa le modalità tramite cui effettuare la segnalazione. Sotto questo profilo, una maggiore uniformità sarebbe stata preferibile, così da scongiurare il rischio che gli utenti meno accorti e meno avvezzi alle tecnologie digitali possano essere indotti in errore.

Per affrontare tali carenze, il legislatore ha optato per la via della co-regolamentazione, già sperimentata nel quadro del DSA[69]. L'art. 50, par. 7, affida all'Ufficio per l'IA[70] il compito di incoraggiare l'elaborazione di codici di buone pratiche, adottati dalla Commissione come atti di esecuzione ai sensi dell'art. 56, par. 6, volti a «facilitare l'efficace attuazione degli obblighi relativi alla rilevazione e all'etichettatura dei contenuti generati o manipolati artificialmente». Tali codici dovrebbero offrire, in sostanza, un supporto operativo ai destinatari degli obblighi di *compliance* stabiliti dal regolamento: tuttavia, trattandosi di strumenti di *soft law*, a cui fornitori e utilizzatori possono peraltro aderire in modo del tutto volontario, non sembrano sufficienti a colmare le lacune normative evidenziate in precedenza.

Tali criticità, infine, sono accentuate dal fatto che l'ambito di applicazione del regolamento 2024/1689 non si estenda agli utilizzi non professionali dei sistemi di IA. Ai sensi dell'art. 2, par. 10, infatti, allorché l'uso dell'intelligenza artificiale sia riconducibile a un'attività puramente personale, gli obblighi sanciti dal regolamento nei confronti dei *deployer* non trovano applicazione. Tale eccezione, benché coerente con la *ratio* dell'*AI Act*, inquadrabile in un'ottica di armonizzazione del mercato interno, solleva qualche perplessità in relazione al fenomeno dei *deep fake*[71]. In effetti, vista la crescente accessibilità dei *software* di intelligenza artificiale, unita alla rapidità di diffusione dei contenuti nella rete, appare difficile escludere *a priori* che un contenuto manipolato, seppur creato in un contesto prettamente privato senza apparenti finalità disinformative, possa generare conseguenze distorsive su larga scala.

---

[68] M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione*, cit., p. 186.

[69] Sul tema, si rinvia a E. BIRITTERI, *Contrasto alla disinformazione, Digital Services Act e attività di private* enforcement, cit.

[70] Come precisato dall'art. 3, par. 47, dell'*AI Act*, l'espressione si riferisce «alla funzione della Commissione volta a contribuire all'attuazione, al monitoraggio e alla supervisione dei sistemi di IA e dei modelli di IA per finalità generali, e della governance dell'IA prevista dalla decisione della Commissione del 24 gennaio 2024».

[71] G. PROIETTI, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri*, cit., p. 347.

6. *Conclusioni*

L'analisi condotta ha evidenziato che l'*AI Act* affronti solo parzialmente i problemi connessi all'uso dell'intelligenza artificiale per finalità disinformative. Tale circostanza è una diretta conseguenza della scelta compiuta in proposito dal legislatore dell'Unione. Nel complesso, infatti, il regolamento si propone di promuovere lo sviluppo dei sistemi di IA, armonizzandone la progettazione, lo scambio e l'utilizzo nel mercato interno, sia pur nel rispetto dei valori sanciti dall'art. 2 TUE. Tale bilanciamento si concretizza tramite un approccio basato sul rischio che ricalca quello affermatosi nell'ambito della legislazione in materia di sicurezza dei prodotti, dunque primariamente orientato a minimizzare gli effetti negativi dell'IA sul godimento dei diritti fondamentali. L'attenzione al dato individuale rischia, tuttavia, di lasciare in secondo piano quei processi non immediatamente lesivi di un diritto fondamentale della persona, eppure parimenti dannosi nel lungo periodo. È il caso, per l'appunto, dei fenomeni disinformativi, i quali, veicolando narrazioni manipolate, sono suscettibili non solo di alterare i meccanismi di formazione del consenso e di interferire con il regolare svolgimento dei processi elettorali, ma anche di polarizzare l'opinione pubblica e accrescere la disaffezione nei confronti della politica: sintomi evidenti di una progressiva erosione dei valori democratici.

Da questo punto di vista, la disciplina dettata dall'*AI Act* risulta piuttosto timida. Il fatto che i sistemi idonei ad influenzare l'esito delle elezioni siano inclusi, in linea di principio, tra quelli ritenuti ad alto rischio riflette la consapevolezza del legislatore circa la pericolosità dell'IA per la tenuta degli ordinamenti democratici. A tale previsione di carattere generale si accompagna, tuttavia, uno schema regolatorio che individua nella causazione di un danno significativo – di natura fisica, psicologica o economica – il presupposto necessario ai fini dell'attivazione degli obblighi di *compliance* imposti a fornitori e *deployer*. Difficilmente tale modello consentirà di contrastare adeguatamente i rischi di derive illiberali connessi all'impiego dei sistemi di IA per ragioni disinformative, giacché siffatti fenomeni producono conseguenze di natura sistemica, non contemplate nella nozione di danno accolta dal regolamento, calibrata invece sulla sfera individuale[72].

Anche la selezione delle pratiche vietate di cui all'art. 5 del regolamento 2024/1689 non sembra offrire adeguate garanzie sul piano del contrasto alla disinformazione *online*. Come si è visto, affinché possa trovare applicazione siffatto divieto, è necessario dimostrare l'esistenza di un nesso causale diretto tra l'applicazione di IA utilizzata e l'alterazione del comportamento umano. Nel nostro caso, tale operazione rischia di rivelarsi sempre vana, dal momento che le decisioni di voto degli elettori sono estremamente volubili, nonché influenzate da una molteplicità di fattori: ne consegue che risulti difficile riscontrare, con un

[72] Tali criticità sono solo parzialmente compensate dalla disciplina riservata ai modelli di IA a scopo generale, i quali, in ragione della molteplicità degli ambiti in cui possono trovare applicazione, possono servire un'ampia gamma di finalità e, di conseguenza, comportare rischi sistemici non prevedibili *ex ante*, tra cui rientrano, *inter alia*, gli «eventuali effetti negativi, effettivi o ragionevolmente prevedibili, sui processi democratici», nonché «da diffusione di contenuti illegali, mendaci o discriminatori» . Al riguardo, l'art. 55 del regolamento *de quo* affida ai fornitori dei sistemi di IA il compito di individuare e mitigare i rischi sistemici che potrebbero derivare dai loro modelli, avvalendosi semmai dei codici di buone pratiche, strumenti di *soft law* formulati dall'Ufficio per l'IA ai sensi dell'art. 56. Tale approccio, affidandosi alla responsabilizzazione degli attori privati, non pare particolarmente efficace, rischiando, peraltro, di dare origine a forme di tutela difformi all'interno degli Stati membri. Sul punto, si rinvia a G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, cit., p. 150 ss.

certo grado di ragionevolezza, una connessione diretta tra una campagna disinformativa avvalsasi dell'intelligenza artificiale e la distorsione del comportamento elettorale delle persone. Benché comprensibile in un'ottica di certezza del diritto, tale requisito appare inidoneo ad arginare le conseguenze negative derivanti dalla disinformazione *online*, la cui natura diffusa e non misurabile non si presta ad essere ricondotta entro un rigido schema di causa-effetto.

Nel complesso, riservando maggiore attenzione all'impatto prodotto dall'intelligenza artificiale sui diritti individuali, il regolamento rischia di sottovalutarne i possibili effetti negativi sul piano collettivo – come la regressione dei principi democratici – i quali, sebbene menzionati a più riprese, non integrano sufficientemente la nozione di danno su cui, in fin dei conti, si basano sia i divieti che le limitazioni. Vero è che, a proposito del contrasto alla disinformazione, il legislatore debba intervenire con cautela, atteso che, in tale ambito, ciascun intervento restrittivo comporta una compressione della libertà di espressione e informazione. Eppure, continuare a percorrere la via europea all'evoluzione tecnologica significa, innanzitutto, agire in chiave preventiva, affinché siano neutralizzate tutte le minacce potenzialmente in grado di nuocere ai valori fondamentali di cui all'art. 2 TUE, *ivi* compresa la disinformazione *online*.