L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Filosofia del diritto e sfide regolatorie postmoderne: il caso della cosiddetta Intelligenza Artificiale. Il problema del *deepfake* come rompicapo giuridico*

Raúl Madrid[1**], Claudio Sartea[2]

[1]*Pontificia Universidad Católica de Chile*
[2]*Università degli Studi di Perugia*

Abstract: **Philosophy of Law and Postmodern Regulatory Challenges: the Case of So-called Artificial Intelligence. The Problem of Deepfake as a Legal Puzzle**
The article explores the role of legal philosophy and jurists in the new landscape of technological society, offering in particular some reflections on the regulatory challenges arising from so-called artificial intelligence. The focus is on the automatic image, video, and audio processing algorithms that produce *deepfake* content, which pose a genuine regulatory puzzle due to their complex implications and powerful impact on established principles and fundamental rights.

Keywords*:* Philosophy of Law, Social Role of Jurist, Technological Civilization, Deepfake, Derrida.

Sommario: 1. Introduzione: la filosofia del diritto tra nottola di Minerva e rosa nella croce – 2. Le sfide della tecnologia e l'assedio all'umano – 3. Nuove tecnologie e trasformazione digitale del mondo: quale ruolo del giurista? – 4. *Deepfake*: definizione e problemi teorici – 5. *Deepfake*: la situazione regolatoria nel mondo – 6. Riflessioni conclusive.

1. Introduzione. La filosofia del diritto tra nottola di Minerva e rosa nella croce

In conclusione di un libro dimenticato troppo in fretta[1], Italo Mancini nel 1986 meditava in modo suggestivo le celebri immagini cui ricorse Hegel nell'introduzione alla sua *Filosofia del Diritto* per rappresentare il ruolo del pensiero rispetto alla realtà ed alla storia. Come tutti sanno, per il filosofo tedesco quel che è razionale è reale, e quel che è reale è razionale: ma ciò significa che la

* I contenuti del saggio sono condivisi dai due autori, introduzione e conclusioni sono scritti a quattro mani. Raúl Madrid ha redatto i paragrafi 4 e 5, Claudio Sartea i paragrafi 1, 2 e 3 nonché la parte dedicata all'ordinamento giuridico italiano del paragrafo 5.
** ANID/FONDECYT/Regular 1240924.
[1] I. Mancini, *Filosofia della prassi*, Morcelliana, Brescia, 1986, spec. pp. 421 ss.

filosofia, in quanto accesso razionale (obbiettivo, concettuale, e non arbitrario né soggettivo) al reale, è chiamata almeno a tre compiti[2].

Il primo: dedicarsi all'Essenza, e non alle molteplici relazioni che possono venire a determinarsi nell'esteriorità sensibile. Dunque, come scrisse Platone, può risparmiarsi i buoni consigli come quello di raccomandare alle balie di cullare i bambini; o anche, secondo l'insegnamento di Fichte, può sentirsi esentata dall'obbligo di perfezionare il passaporto di polizia.

Il secondo: la filosofia è chiamata a comprendere concettualmente ciò che è, quindi non è altra cosa che il proprio tempo catturato, colto, percepito col pensiero: dal che discende che affermare che la filosofia possa, o persino debba, andare oltre il proprio tempo è assurdo come pretendere che un individuo salti oltre se stesso, salti "oltre Rodi". In tal senso, ogni filosofo si confronta con la Realtà e cercare in essa la Ragione equivale a cercare la rosa nella croce della Presenza. Per questa via si perviene secondo Hegel a godere della Presenza, a riconciliarsi con la Realtà, rappacificarsi con essa in una forma calda, e non fredda come quella tentata dalle vie alternative del superficiale qualunquismo o della disperazione.

Il terzo ed ultimo compito della filosofia la vuole hegelianamente vocata a spiegarci il mondo: ma per farlo, essa deve attendere che la Realtà abbia completato storicisticamente il proprio processo di formazione, è dunque costretta ad arrivare sempre tardi, troppo tardi per insegnarci come la realtà *debba* o *dovrebbe* essere. Da qui proviene la nota metafora: "Quando la Filosofia dipinge il suo grigio sul grigio, allora una figura della vita si è fatta vecchia, e con grigio su grigio non possiamo ormai più ringiovanirla, ma solamente riconoscerla: la nottola di Minerva inizia il suo volo solo al crepuscolo"[3].

Le suggestioni hegeliane, a parere di Mancini, ci aiutano in maniera potente a chiarire il senso di ogni filosofia, e della filosofia del diritto in particolare. Ai fini del nostro articolo, le evochiamo con riferimento alla riflessione teorica e regolatoria sulle nuove tecnologie, specialmente quelle collegate con la cosiddetta "intelligenza artificiale", e, in seno ad essa, il problema concettuale e normativo del *deepfake*, o falso prodotto con algoritmi dotati di autonomia e capacità generativa. In primo luogo, confermano che la filosofia *deve occuparsi* delle sfide del presente, perché il suo compito è affrontare la realtà e comprendere il presente, assumerlo, catturarlo con il pensiero. In secondo luogo, le suggestioni hegeliane difendono una concezione del filosofo che va oltre i consigli comuni o le spiegazioni tecniche: quando si interessa di un settore del reale dovrà certamente averne una qualche conoscenza, ma non per ripeterci quello che già i cosiddetti esperti sono perfettamente in grado di dirci da soli, bensì per andare più in là, all'essenza – magari problematica e comunque provocatoria per il pensiero – del fenomeno considerato. In terzo luogo, non potremo rimproverare alla filosofia (e meno che meno alla filosofia del diritto) il fatto che arrivi tardi rispetto allo sviluppo

[2] G.W.F. Hegel, *Lineamenti di filosofia del Diritto*, (1820), trad. it., Bompiani, Milano, 2010, pp. 59 ss.
[3] *Ivi*, XXIII, p. 65.

tecnoscientifico da cui scaturiscono continuamente nuove sfide etiche e giuridiche, perché se c'è una caratteristica del pensiero, se qualcosa di assolutamente tipico esiste in questo peculiare impossessarsi del presente con il pensiero, questo qualcosa è proprio il suo levarsi al crepuscolo, il suo guardare a ciò che si è prodotto, realizzato, consumato – proprio perché è la sua consumazione ciò che il pensiero deve attendere per potersene appropriare.

In questo saggio, sosteniamo che il *deepfake* costituisce un rompicapo giuridico-filosofico non tanto perché moltiplica i falsi contenuti in rete, ma perché destabilizza tre presupposti sui quali poggiano la comunicazione pubblica e l'operatività del diritto: (a) la verità operativa del registro visivo, vale a dire la sua capacità di funzionare come ragione spontaneamente affidabile per credere ed agire; (b) l'identità e la paternità come condizioni di imputazione, vale a dire la possibilità di stabilire con certezza chi parla, chi agisce e chi deve rispondere; (c) il regime di rappresentazione, in cui volto e voce operano come segnali sociali di presenza e responsabilità. Da tutto ciò deriva che una risposta normativa adeguata non si può limitare a perseguire la tecnica, ma deve articolarsi in un'architettura multilivello – tra protezione dei dati, obblighi di trasparenza, regole sulla circolazione nelle piattaforme e garanzie probatorie – capace di ricostruire istituzionalmente il nesso tra segno, soggetto ed attribuzione.

2. Le sfide della tecnologia e l'assedio all'umano

Difficile è negare che tra le sfide più gravi che il pensiero deve oggi affrontare vi sono quelle lanciate dalle nuove tecnologie. In modo speciale, le nuove tecnologie informatiche (così come, sia notato qui solo per inciso, le tecnologie applicate al *bios*, per molti aspetti comunicanti[4]), per la loro capillarità e trasversalità di applicazione ed utilizzo, generano per la riflessione antropologica, il pensiero etico e la regolazione legale, sfide inedite quanto ai contenuti ed anche quanto alla forza di provocazione (promesse e minacce). Sin dagli anni Sessanta dello scorso secolo, denotando una chiaroveggenza non comune, Sergio Cotta ebbe a meditare "la sfida tecnologica"[5] portata al pensiero da quella che propose di chiamare "energia tecnologica"[6]. In essa si condensavano, a suo avviso, gli sviluppi ed i risultati già allora spettacolari della convergenza di scienza, tecnologia ed industria, ormai

[4] Al punto che un'autorità riconosciuta nel panorama filosofico contemporaneo come Luciano Floridi assimila bioetica ed etica dell'intelligenza artificiale per i quattro quinti dei principi etici cui bisognerebbe ricorrere: cfr. L. Floridi, *Etica dell'intelligenza artificiale*. *Sviluppi, opportunità, sfide*, trad. it., Raffaello Cortina, Milano, 2022, spec. pp. 92 ss.
[5] S. Cotta, *La sfida tecnologica*, il Mulino, Bologna, 1968.
[6] *Ivi*, p. 35.

saldamente intrecciate nel circolo analizzato da Gehlen come un fenomeno nuovo di quella che il pensatore tedesco definì "l'era della tecnica"[7].

Cotta in quel saggio così lontano e così attuale fece proprie le tesi di Heidegger sulla differenza essenziale tra tecnica tradizionale e tecnica moderna, aderendo alla critica del filosofo tedesco[8]. La nuova tecnica, quella che potremmo in modo pregnante ribattezzare "tecno-logia", non è più metodo o semplice mezzo (quello di cui da sempre gli esseri umani si servono per pro-durre, per accompagnare la natura nel mettere a disposizione dei bisogni umani quel che di meglio essa può dare, svelando così la propria verità più intima): adesso, in maniera inedita, la tecnica pro-voca la natura, la chiama fuori di sé perché la riduce a risorsa, a *fondo* disponibile in senso quantitativo. Questo appiattimento non risparmia nulla, nemmeno la natura umana, le cui manifestazioni sempre più spesso vengono trasformate in risorse (si pensi alla trapiantologia, alla sperimentazione sull'uomo, alla fecondazione artificiale con la circolazione e lo sfruttamento senza precedenti di cellule germinali, di embrioni, di ambienti uterini). Come aveva ipotizzato Heidegger, e come Gehlen non mancò di annotare, sarebbe arrivato il giorno in cui la colonizzazione tecnologica dei mondi vitali avrebbe fatto approdo all'umano: niente di sorprendente, dunque, nemmeno negli sviluppi transumanisti e postumanisti della cibernetica e dell'informatica, sviluppi che peraltro, a quel che si apprende, sarebbero già entrati in una fase di crisi[9].

La riflessione heideggeriana sulla tecnica, per accettazione o per contrasto, ha ispirato gran parte della filosofia della tecnica successiva per tutta la seconda metà del secolo ventesimo. Dalla scuola stessa del filosofo tedesco sono scaturiti alcuni dei contributi più rilevanti, *in primis* quello di Hans Jonas. Come noto, per l'autore de *Il principio responsabilità* le sfide tecnologiche ponevano l'esigenza di un'etica nuova perché derivavano da un impatto inedito del potere manipolativo umano sull'ambiente, tale da produrre problemi completamente nuovi. La minaccia macrocosmica legata allo sfruttamento dell'energia atomica, come quella microcosmica ma non meno insidiosa implicata dalla conoscenza e manipolazione del genoma dei viventi, umani compresi, aprivano nei decenni successivi alla Seconda guerra mondiale scenari caratterizzati non solo da spaventose minacce, ma anche proprio da novità assolute, che imponevano un nuovo pensiero morale e nuovi principi. Il principio responsabilità è quello dominante tra le proposte di Jonas, proiettato su uno sfondo a sua volta dominato dal criterio della precauzione:

[7] A. Gehlen, *L'uomo nell'era della tecnica. Problemi socio-psicologici della civiltà industriale*, (1957), trad. it. a cura di M.T. Pansera, Armando, Roma, 2003.

[8] In special modo riassunta nella celebre conferenza *La questione della tecnica*, del 1953, ora raccolta anche in M. Heidegger, *La questione della tecnica*, in Id., *Saggi e discorsi*, (1954), trad. it., Mursia, Milano, 1976.

[9] Si veda quello che sembra proprio l'epitaffio del "Future of Humanity Institute" dell'Università di Oxford, recuperato da https://www.futureofhumanityinstitute.org/, [Data di consultazione: 20/03/2026].

> All'immodestia dei suoi obiettivi, che mancano il bersaglio sotto il profilo sia ecologico che antropologico (com'è dimostrabile per l'uno e argomentabile filosoficamente per l'altro), il principio responsabilità contrappone il compito più modesto, dettato dalla paura e dal rispetto, di preservare all'uomo, nella residua ambiguità della sua libertà, che nessun mutamento delle circostanze può mai sopprimere, l'integrità del suo mondo e del suo essere contro gli abusi del suo potere[10].

Non si nascondeva già allora questo autore il carattere teoretico della questione, che non era mero problema di potenza ma vera e propria rivoluzione del rapporto tra l'essere umano e la realtà: "Il significato della rivoluzione tecnologica è parte del significato metafisico della rivoluzione scientifica, anzi, lo completa. La metafisica della scienza è venuta allo scoperto"[11].

Insistendo in modo del tutto analogo sul rapporto tra l'umana libertà ed i suoi abusi, diversi anni prima Romano Guardini aveva scritto che l'uomo contemporaneo

> si è liberato del dogma moderno che tutte le cose tendono per proprio conto al meglio. Per lui non esiste più l'ottimismo della fede nel progresso; egli sa che altrettanto facilmente, anzi più facilmente le cose possono volgere al peggio. Egli sa che il mondo è nelle mani della libertà; ne sente perciò responsabilità. E amore, un amore particolare: proprio perché il mondo è lanciato allo sbaraglio ed è distruttibile. Al sentimento del potere e della sua grandezza, alla familiarità con la tecnica ed alla volontà di farne uso, al fascino del pericolo si mescola una benevolenza, anzi una tenerezza per questa esistenza finita e così minacciata[12].

Dello stesso mutamento di paradigma di cui si occupa Jonas, dell'inatteso rovesciamento del rapporto con la natura (da madre – benevola e maligna al contempo – ad oggetto di dominio ed in qualche modo a creazione umana, o almeno ad universo sottoposto all'umana potenza) parlano dunque non pochi dei pensatori più profondi e sensibili del secolo in cui la tecnica è diventata tecno-logia, vero e proprio tecnomorfismo – come spiega Francesco D'Agostino:

> Che cos'è il tecnomorfismo? È l'idea che attraverso la tecnica, attraverso l'artificio, l'uomo possa dar liberamente forma a tutte le cose. È, in un certo senso, la più blasfema traduzione, nel lessico contemporaneo, dell'idea di

[10] H. Jonas, *Il principio responsabilità. Un'etica per la civiltà tecnologica*, (1979), trad. it. a cura di P.P. Portinaro, Einaudi, Torino, 2001, p. XXIX.
[11] H. Jonas, *Dopo il XVII secolo: il significato della rivoluzione scientifica e tecnologica*, in Id., *Dalla fede antica all'uomo tecnologico*, (1974), trad. it., Einaudi, Torino, 1991, p. 99.
[12] R. Guardini, *Il potere. Tentativo di un orientamento*, (1951), trad. it., Morcelliana, Brescia, 1963, p. 110.

> creazione. Negare Dio come creatore e riconoscere all'uomo un potere tecnomorfico sono fondamentalmente la stessa cosa[13].

Com'è chiaro, non si tratta di un problema quantitativo – la tecnica ha sempre fatto parte dell'esperienza vitale umana, ed anzi è possibile sostenere che l'essere umano sia naturalmente artificiale, sia tecnico per natura e non per cultura[14] – bensì qualitativo: è lo sguardo umano sul mondo che subisce un radicale mutamento, che con la già menzionata dottrina di Heidegger potremmo riassumere nel passaggio da pro-duzione a pro-vocazione[15], e con termini meno sofisticati si manifesta sotto le sembianze di un progressismo tecnocentrico. Qualunque aspetto negativo possa riscontrarsi e denunciarsi in questo atteggiamento nuovo, esso non è rappresentato dall'abilità tecnica umana. È ancora D'Agostino a sottolinearlo con espressioni particolarmente efficaci:

> Il cuore del problema non sta nel criticare la capacità progettuale, che – ripetiamolo – è tipica dell'intelligenza umana, ma nel riconoscere che esiste, nell'uomo, una dimensione che non è riducibile alla progettualità. Una dimensione nella quale la tecnica non può e non deve entrare, ma non perché non abbiamo ancora fatto gli studi necessari per impadronirci di quell'ambito, non perché siamo in una situazione di ritardo tecnologico, ma perché si tratta di una dimensione umanamente preziosissima, ma in linea di principio non gestibile in modo progettuale[16].

Ci sono dunque beni, valori, principi, elementi dell'umano che non risultano riproducibili macchinicamente, e non perché non siano replicabili dalle nuove strepitose potenzialità dei nostri strumenti tecnologici, che anzi già possono farlo per innumerevoli profili, magari superando l'originale: ma perché ostano ragioni di principio al vicariamento tecnico di quei beni, valori, principi, elementi dell'umano.

Quest'ultima chiarificazione rimanda a quello che possiamo considerare l'argomento principale dei sostenitori del cosiddetto "umanesimo digitale"[17] e più in generale della resistenza antropocentrica all'invadenza macchinica[18]: quello che già un maestro dell'informatica, il padre della prima *chatbot ante litteram*, *Eliza*,

[13] F. D'Agostino, *Bioetica e biopolitica. Ventuno voci fondamentali*, Giappichelli, Torino, 2011, p. 215.
[14] Lo ha sostenuto con vigore già Arnold Gehlen, *op. cit.*, p. 32, ed oggi in Italia riprende questa consolidata tesi, applicandola alla "novità" dell'intelligenza artificiale, tra gli altri M. Ferraris, *La pelle. Che cosa significa pensare nell'epoca dell'intelligenza artificiale*, il Mulino, Bologna, 2024, p. 34.
[15] M. Heidegger, *La questione della tecnica*, cit., p. 9.
[16] F. D'Agostino, *op. cit.*, p. 216.
[17] Si tratta di una galassia di autori e contributi che hanno probabilmente il loro fulcro riconosciuto nel cosiddetto "Manifesto di Vienna", del 2019, per informazioni sul quale rimandiamo a https://caiml.org/dighum/.
[18] Cfr. M. Bertolaso, A. Marcos, *Umanesimo tecnologico. Una riflessione filosofica sull'intelligenza artificiale*, Carocci, Roma, 2023.

considerava il punto capitale. Riassumendo in modo estremo il suo contributo, secondo Joseph Weizenbaum la questione non è se o quando le macchine *potranno* raggiungere e superare le capacità umane (di calcolo, memorizzazione, sintesi di dati, etc.), bensì quali siano le attività che noi esseri umani *dobbiamo* assegnare loro[19]. Per il celebre informatico del MIT, la responsabilità resta in ultima analisi sempre nostra, ed è in grado di orientare lo stesso sviluppo di queste tecnologie onde garantirle eticamente, giuridicamente e politicamente[20].

Non meno penetrante e significativo ai fini di una meditazione sulla tecnologia e le sue attuali derive antropologiche, è il contributo di un altro allievo di Heidegger, Günther Anders. A suo avviso,

> la nostra illimitata libertà prometeica di creare sempre nuove cose (costretti come siamo a pagare senza sosta il nostro tributo a questa libertà) ci ha portati a creare un tale disordine in noi stessi, esseri limitati nel tempo, che ormai proseguiamo lentamente la nostra via, seguendo di lontano ciò che noi stessi abbiamo prodotto e proiettato in avanti, con la cattiva coscienza di essere antiquati, oppure ci aggiriamo semplicemente tra i nostri congegni come sconvolti animali preistorici[21].

Come anche Anders rileva, tuttavia, la pervasività delle nuove tecnologie è tale da giungere ad assediare anche noi stessi, la nostra autopercezione, la nostra identità, il nostro corpo:

> Non già perché non sopporta più nulla che egli stesso non abbia fatto, vuole fare se stesso; ma perché non vuole essere qualche cosa di non-fatto. Non perché provi indignazione per essere fatto da altri (Dio, dèi, natura), ma perché non è fatto per nulla e, nella sua qualità di non-fatto, è inferiore a tutti i suoi prodotti fabbricati[22].

Proprio come suggerisce Hannah Arendt, altra allieva di Heidegger capace di lasciare traccia nella riflessione critica della tarda modernità pur non occupandosi direttamente di filosofia della tecnica, è la nascita (quella che Arendt chiama "natalità" come categoria antropologica) a marcare la distanza tra natura ed artificio, tra donato e prodotto: ma quello che per la pensatrice tedesca è una benedizione o un privilegio, per Anders è la nostra condanna.

[19] J. Weizenbaum, *Il potere del computer e la ragione umana. I limiti dell'intelligenza artificiale*, (1976), trad. it. a cura di F. Tibone, Edizioni Gruppo Abele, Torino, 1987.
[20] J. Weizenbaum, "Not Without Us", in *ECT. A Review of General Semantics*, 44 (1987), n. 1, pp. 44 ss.
[21] G. Anders, *L'uomo è antiquato. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, (1963), trad. it., Bollati Boringhieri, Torino, 2007, p. 24.
[22] *Ivi*, p. 33.

> Non si conosce alcuna religione per cui il fatto che l'uomo non sia un dio, ma solo una creatura, abbia rappresentato una giustificazione per l'indolenza morale; così, anche oggi, per la religione dell'industria e i suoi seguaci, il fatto che l'uomo è solo una creatura e non un prodotto non vale quale scusa del suo pigro perseverare nella sua insufficienza naturale. Deve assolutamente tentare di dimostrare la sua devozione alle cose, tentare una *imitatio instrumentorum*, una riforma di se stesso; procurare almeno di 'migliorarsi', quel tanto da ridurre al minimo possibile il 'sabotaggio' che egli opera a motivo del suo 'peccato originale', la nascita[23].

È per questa via che il mezzo diviene un fine, ed il fine – l'essere umano – un mezzo. Anche se si decide di discostarsi dal pessimismo e dal catastrofismo andersiani, è difficile contestare la precisione delle sue osservazioni su quello che la tecnologia stava già allora facendo alla nostra comprensione di noi stessi e del mondo. Il senso di inferiorità rispetto alle nostre macchine che egli ha denominato "vergogna prometeica" è puntualmente rappresentato in mille diffuse forme di apprezzamento ed ammirazione delle risorse tecnologiche, di delega ad esse di funzioni sempre più complesse e soprattutto sempre più delicate. Ne scaturisce un importante problema, non riducibile a quello della perdita di posti di lavoro[24]: è il problema di continuare a sapere chi siamo, che cosa ci caratterizza ultimativamente, ed è il problema di non poter più collocare questa "cosa" sul versante delle abilità o delle capacità, nemmeno di quelle più finemente intellettuali, se è vero che ormai siamo stati superati e soppiantati su tutti i fronti[25].

3. Nuove tecnologie e trasformazione digitale del mondo: quale ruolo del giurista?

Abbiamo già incontrato Sergio Cotta e la sua riflessione, acuta quanto rapidamente dimenticata, sull'era tecnologica e la sua sfida per l'antropologia filosofica, l'etica, il diritto. All'interno di un più ampio discorso di critica della tecnologia, questo autore offre una constatazione che può orientare la presente analisi del rapporto tra diritto e nuove tecnologie.

> Giunto forse per ultimo fra gli studiosi a tecnicizzare il proprio lavoro sotto la spinta di una sorta di umiliata fascinazione per le scienze naturali ed esatte,

[23] *Ivi*, p. 43.
[24] Che pure, ovviamente, è un problema significativo: cfr. J. Kaplan, *Le persone non servono. Lavoro e ricchezza nell'epoca dell'intelligenza artificiale*, (2015), trad. it., Luiss University Press, Roma, 2016.
[25] È noto che lo sviluppo delle tecnologie informatiche ha proceduto lungo i binari di una gara con le facoltà intellettive umane, in particolare nel loro esercizio ludico (dagli scacchi al Go): per una breve storia cfr. N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, il Mulino, Bologna, 2023.

> presunto modello di ogni scientificità (basti ricordare l'affannoso interrogarsi ottocentesco sulla scientificità della giurisprudenza), il giurista è oggi forse fra i primi ad essersi reso conto dei limiti di un tecnicismo assoluto. Orbene, proprio questo è l'atteggiamento che, come abbiamo visto, sembra richiesto dall'ardua impresa di costruire una civiltà che non ripudi la rivoluzione tecnologica, ma nemmeno si abbandoni ciecamente ad essa[26].

La "ardua impresa" di trovare il *giusto* equilibrio tra gli eccessi dell'ingenuità tecnofila e l'antistorico (ed antiumano, oltre che scarsamente realistico) misoneismo dei tecnofobi, sembra dunque ciò che attende il giurista in questo frangente culturale e sociale da tutti ormai ribattezzato nelle acque della cosiddetta intelligenza artificiale[27]. E si noti che qui la riflessione s'incentra non astrattamente sul senso o la funzione del *diritto* nell'era tecnologica, ma sul ruolo del *giurista*: dell'operatore giuridico, lo si intenda nella sua funzione di elaboratore del diritto (legislatore, amministratore pubblico), di interprete (giudice, avvocato, consulente ai vari possibili livelli), di studioso (accademico, teorico e tecnico). Come ricorda Cotta nel suo saggio, "la società tecnologica segna il completo superamento del modello giuspositivistico del giurista, di cui mette in luce il carattere intellettualistico, astrattamente ideologico e non reale"[28]. Per questo autore, quella giuspositivistica non è altro che una parentesi, tra un prima plurimillenario in cui il giurista è certamente alla guida della società ma non come governante bensì al più come consigliere dei potenti, e comunque sempre come coscienza critica del gruppo dirigente e come paladino dell'*ethos* comunitario (che Cotta non esita a chiamare anche "diritto naturale"), ed un poi che oggigiorno è tutto da immaginare, rinunciando ove necessario alla comoda tesi hegeliana ricordata all'inizio, secondo la quale la filosofia sarebbe adeguatamente rappresentata dal volo notturno della civetta, perché la rapidità dei cambiamenti e soprattutto la loro inedita potenza manipolativa esigono dalla filosofia del diritto e dai giuristi una riflessione capace anche di profezia, di proiezione verso il futuro.

Come scrive il filosofo fiorentino,

> la società tecnologica, in quanto è (come più volte ho sottolineato) una situazione di radicale ambivalenza impone a tutti una risposta globale, e quindi un recupero umanistico, una chiara assunzione di responsabilità. È un dovere che incombe ad ogni uomo di pensiero – filosofo, scienziato, poeta, artista – che non può più limitarsi, levandosi a volo sul far del crepuscolo come la nottola di Minerva, a interpretare e registrare il proprio tempo che fugge o a impegnarsi nella lotta del momento. Ma deve saper penetrare il fondamento

[26] S. Cotta, *op. cit.*, p. 152.
[27] Per innumerevoli autori la nostra è l'epoca/l'era dell'intelligenza artificiale: si vedano tra gli altri M. Ferraris, *op. cit.*, A. Fabris, *La filosofia nell'epoca dell'intelligenza artificiale. Come ci aiuta a vivere negli ambienti digitali*, Carocci, Roma, 2025; A. Pessina, *L'essere altrove. L'esperienza umana nell'epoca dell'intelligenza artificiale*, Mimesis, Udine-Milano, 2023.
[28] S. Cotta, *op. cit.*, p. 182.

> dell'esistenza per poter poi discernere le prospettive umane, di rischio e di speranza, in questo tempo racchiuse, offrendo così la propria lungimirante testimonianza di verità a quanti, immersi nell'azione, dall'azione sono travolti[29].

In che cosa consiste allora il ruolo del giurista all'interno di una simile contingenza storica e culturale? Che cosa la rivoluzione neotecnologica – che è stata anche definita la "quarta rivoluzione"[30] – richiede specificamente alle professioni della giustizia, per corrispondere alla propria funzione, alla propria "verità"? È ancora Cotta a cercare di dare una risposta all'ingombrante e spinosa domanda:

> Questo recupero umanistico ed etico, necessario per tutti, è richiesto anche al giurista. Non gli sarà certo difficile compierlo solo che sappia ispirarsi alla sua antica e sempre attuale funzione. In tutta la sua secolare vicenda, eccezion fatta per la fugace parentesi del positivismo giuridico, il giurista è stato il custode della volontà dell'individuo, attento a garantirne le possibilità di libera manifestazione, ad assicurarle efficacia e durata (persino oltre la morte del soggetto!), ma sempre nel rispetto e nel coordinamento delle necessità della convivenza, badando quindi a segnarne anche i limiti e le responsabilità e a proibirne le prevaricazioni[31].

Espressioni che potrebbero apparire generiche e di scarsa utilità, a distanza di tanti anni e con in mezzo il vertiginoso sviluppo di dispositivi tecnologici enormemente più pervasivi e perturbanti[32]: ma che acquistano tutto il loro rinnovato vigore teoretico, ed un'efficacia euristica persino accresciuta, se si comprende adeguatamente il senso di quel riferimento cottiano al "fondamento dell'esistenza" a partire dal quale procede il discernimento dei rischi e delle possibilità autenticamente umane che la nuova tecnologia racchiude. Come ha scritto acutamente Adriano Pessina,

> prima dell'etica occorre sviluppare, per così dire, un'antropologia delle nuove tecnologie. Solo così si può cercare di capire se la rete non renda di fatto impossibile trascendere totalmente la dimensione della pura funzionalità, evidenziando i limiti di tutti quei progetti di 'umanizzazione' della rete che pensano di poter simulare ed emulare al suo interno la vita sociale, nelle sue dimensioni formative, comunicative e affettive. La biosfera include l'infosfera, ma non il contrario[33].

[29] *Ivi*, p. 188.
[30] L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, (2016), trad. it., Raffaello Cortina, Milano, 2017.
[31] S. Cotta, *op. cit.*, p. 189.
[32] Si pensi al cosiddetto "effetto della valle perturbante" ben conosciuto dagli studiosi degli effetti delle tecnologie robotiche sul pubblico umano.
[33] A. Pessina, *op. cit.*, p. 101.

Forse è allora qui che la filosofia del diritto oggi può cercare e trovare il proprio posto, offrendo uno specifico ed insostituibile contributo epistemico all'intenso dibattito sulle nuove tecnologie, e segnatamente sulla cosiddetta intelligenza artificiale. Nel prosieguo di questo scritto, proveremo a mostrarlo con specifico riferimento ad una sua applicazione estremamente ambigua, quella del cosiddetto *deepfake*. Come emergerà con chiarezza, gli algoritmi che generano questi prodotti sfidano frontalmente alcuni dei beni giuridici su cui è stato edificato l'ordinamento giuridico liberaldemocratico che si è affermato nella maggior parte dei Paesi del mondo, e che in ogni caso viene additato come un modello ideale cui tutti dovrebbero aspirare. La libertà individuale – come del resto già Cotta ci ha spiegato – è tra questi beni quello forse più prominente ed in vista, ma senza dubbio il fenomeno del *deepfake* compromette alcuni gangli vitali della sua declinazione giuridica in termini di protezione, limitazione, imputazione delle azioni e responsabilità per i loro effetti (problema che potremmo definire di "autoria" o paternità, in ogni senso positivo e negativo). La verità obiettiva, principale argine della libertà individuale o, meglio, sua canalizzazione per la realizzazione della persona (si pensi alla formulazione dell'art. 2 della Costituzione italiana), è analogamente minacciata dall'abuso delle tecnologie di *machine learning* tutte le volte che queste intacchino o contaminino la narrazione del passato, a cui sono strutturalmente consacrate, manipolandolo secondo progetti ed intenzioni la cui stessa genealogia è spesso oscurata dai meccanismi di *black box* ben noti agli studiosi degli algoritmi di IA[34]. Se "ingiusto è lo stesso che falso", come afferma Aristotele nella *Retorica*, è evidente che il discorso giuridico è direttamente chiamato in causa dal fenomeno in questione. Vi è poi il profilo collettivo dei beni in gioco, che rimanda alla formazione dell'opinione pubblica: un processo delicatissimo e centrale nell'economia concettuale delle democrazie contemporanee, la cui protezione in termini di verità (dei contenuti circolanti) e di libertà (del pensiero, di opinione, e della loro espressione e diffusione) è a dir poco vitale per le medesime[35].

Queste ed altre sfide al mestiere del giurista, lanciate con inaudita aggressività dall'IA, evocano anche una questione che potremmo definire di metodo: in base ad una possibile prospettiva, di marca liberale, il nuovo, specie il nuovo tecnologico, dev'essere lasciato a se stesso ed alla propria attitudine autoregolatoria, sia perché i giuristi non fanno altro che arrivare in ritardo e rincorrere l'evoluzione tecnica, sia perché le dinamiche di tale evoluzione sono strettamente legate alle leggi del

[34] F. Pasquale, *The Black Box Society. The Secret Algorithms that control Money and Information*, Harvard University Press, Cambridge (MA), 2015. Di questo autore è pure molto interessante il più recente saggio, tradotto anche nella nostra lingua, F. Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale*, (2020), trad. it., Luiss University Press, Roma, 2021.

[35] Alcuni dei problemi sollevati dall'IA su questo versante sono analizzati in T. Casadei, S. Pietropaoli (a cura di), *Diritto e tecnologie informatiche. Questioni di informatica giuridica. Prospettive istituzionali e sfide sociali*, Wolters Kluwer, Milano, 2021.

mercato e dipendono prevalentemente dalla concorrenza tra aziende commerciali, dalla competizione di privati interessi. Ma vi è anche una prospettiva che contesta vigorosamente questo *laissez-faire*, soprattutto in base alla constatazione che il principale veicolo di realizzazione e circolazione dei contenuti digitali, Internet, nonostante la sua rilevante funzione pubblica è in mano privata e necessita di regolazione incisiva proprio per evitare che gli interessi commerciali sopravanzino il bene comune: questo atteggiamento inevitabilmente invoca il ruolo dei giuristi ed affida loro l'impegnativo compito di comprendere queste tecnologie e le loro minacce, per disciplinarle efficacemente e garantire agli utenti almeno la tutela elementare dei loro diritti fondamentali.

4. *Deepfake*: definizione e problemi teorici

Il vertiginoso sviluppo delle tecnologie di Intelligenza artificiale (IA) e apprendimento automatico (*deep learning*) ha modificato la maniera in cui le persone producono e consumano contenuti digitali[36]. Uno dei sottoprodotti di questo sviluppi è quello attualmente conosciuto come *deepfake*: un effetto secondario dei progressi nell'IA, in particolare nelle reti neurali e nel *deep learning* orientato alla generazione e manipolazione di immagini e video.

Il termine *deepfake* (la cui traduzione più appropriata in italiano è probabilmente "ultrafalsificazione") fu introdotto nel 2017 da un utente di Reddit, che pubblicava video falsi utilizzando tecniche di IA per cambiare i volti di personaggi famosi in sequenze pornografiche reali. Samantha Cole, giornalista di *Motherboard*, lo rese pubblico:

> C'è in internet un video di Gal Gadot che intrattiene relazioni sessuali con il fratellastro. Ma non è realmente il corpo di Gadot, ed anche il volto a malapena è il suo. Si tratta di un'approssimazione, con il suo viso modificato perché sembri che sia lei ad agire in un video pornografico a sfondo incestuoso[37].

Secondo "Deepfake", che preferì non rivelare la propria identità per evitare la pubblica gogna, il *software* si basava su varie biblioteche a codice aperto, come Keras con il *backend* di *TensorFlow*[38]. Per inserire i volti delle celebrità, utilizzò la

[36] D. Gamage, J. Chen, P. Ghasiya, K. Sasahara, *Deepfake and society: What lies ahead?*, in M. Khosravy, I. Echizen, N. Babaguchi (a cura di), *Frontiers in fake media generation and detection*, Springer, Singapore, 2022, pp. 3-44.

[37] S. Cole, "AI-assisted fake porn is here and we're all fucked", in *Motherboard (Vice)*, 2017, December 11, https://www.vice.com/en/article/ai-assisted-fake-porn-is-here-and-were-all-fucked.

[38] *Backend* si riferisce alla parte "retrostante" o non visibile di un sistema o di un'applicazione, che ha il compito di processare e gestire la logica, i dati e le funzionalità necessarie perché l'utente interagisca con il sistema in modo fluido. Nel contesto dello sviluppo di *software* e *machine learning*, il *backend* è l'ambiente in cui si realizzano i calcoli e le operazioni necessarie affinché il *frontend* (la parte visibile all'utente) funzioni in modo adeguato. Nel caso del *deep learning*,

ricerca di immagini di Google e video di YouTube. Allenò l'algoritmo con video pornografici ed il volto di Gal Gadot, e dopo il sufficiente "addestramento", i nodi si organizzarono per completare un incarico specifico, quello di manipolare un video in maniera convincente ed in tempo reale.

In base a tutto ciò, i *deepfake* possono essere definiti come registrazioni di audio o video alterate digitalmente, in cui il volto e/o la voce di una persona vengono sovrapposti al corpo di un'altra, creando un'evidenza ingannevole di eventi che non si sono mai verificati[39]. In altre parole, si tratta di materiale audiovisivo manipolato con ricorso all'IA[40], che si avvale della capacità di un sistema di interpretare correttamente dati esterni, apprendere da tali dati e utilizzare questi apprendimento per perseguire obiettivi e realizzare compiti specifici mediante un adattamento flessibile[41], così che il risultato sia una tecnica di sintesi di immagini basata sull'IA che implica la creazione di contenuto video falso ma altamente realistico, tramite cui è possibile cambiare l'aspetto di una persona, un oggetto o il contesto[42].

Tutte queste definizioni o spiegazioni hanno un nucleo condiviso che si ripete, con parole diverse ogni volta, e si caratterizza poiché (a) riguarda un contenuto audiovisivo (video o audio; immagine/voce) che fa da supporto al fenomeno; (b) in tutte vi è un'idea di alterazione, manipolazione o generazione digitale del materiale, non mera interpretazione umana, bensì produzione tecnica del contenuto; (c) questa alterazione o manipolazione è sintetizzata in ogni caso col ricorso all'IA; (d) il punto focale è il volto o la voce (elementi identificativi), visto che si tratta della sovrapposizione, sostituzione o modificazione di come si presenta una persona; (e) il prodotto è un effetto altamente realistico e verosimile, proprio perché intende sortire l'effetto di una prova o di una rappresentazione persuasiva; (f) questo risultato produce evidenza ingannevole o apparenza di avvenimenti o azioni che non si sono mai verificati (o, più in generale, una rappresentazione falsa idonea a provocare affidamento), e (g) si ripete l'idea che non ci si limita a "editare", ma si riconfigura la presentazione (persona, oggetto, ambiente) nel materiale impiegato.

La creazione di *deepfake* si basa su un tipo di algoritmo di *deep learning* conosciuto come reti generative antagoniste (*Generative Adversarial Networks*, GANs). Consiste in una modalità di addestramento di modelli generativi (vale a dire, modelli che creano contenuti), una innovazione dell'apprendimento

frameworks come *TensorFlow* operano come *backend*, fornendo l'ambiente e gli strumenti affinché i modelli realizzino calcoli complessi e processino grandi volumi di dati.

[39] R. Rini, L. Cohen, "Deepfake, deep harms", in *Journal of Ethics and Social Philosophy*, 22 (2022), n. 2, p. 143.

[40] A. Godulla, C.P. Hoffmann, D. Seibert, "Dealing with deepfake. An interdisciplinary examination of the state of research and implications for communication studies", in *Studies in Communication and Media*, 10 (2021), n. 1, p. 75.

[41] A. Kaplan, M. Haenlein, "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", in *Business Horizons*, 62 (2019), n. 1, p. 17.

[42] A. Godulla, C.P. Hoffmann, D. Seibert, *op. cit.*, p. 76.

automatico che utilizza un metodo speciale per allenarsi ed evitare problemi che si manifestavano in altri metodi più risalenti, e che consente la creazione di contenuti audiovisivi iperrealistici e personalizzati in maniera istantanea. Il suo addestramento si realizza come un gioco di competizione tra due modelli: (a) un modello generatore, che cerca di creare immagini realistiche o contenuto falso, e (b) un modello discriminatore, che cerca di capire se le immagini o il contenuto che riceve sono reali o falsi. Entrambi si allenano insieme in un gioco a somma zero (uno vince, l'altro perde), in cui il generatore migliora continuamente per ingannare il discriminatore, e il discriminatore migliora per indovinare i falsi intenti del generatore, come se si trattasse dei ruoli di falsario e di poliziotto, rispettivamente[43]. Il modello generativo sarebbe equivalente a quelli che cercano di produrre denaro falso, mentre il modello discriminatore sarebbe equivalente agli agenti di polizia che cercano di impedirlo. Questa dialettica di confronto è quella che produce falsificazioni sempre più realistiche, fino al punto che il modello discriminatore non è più in grado di stabilire se il risultato sia falso oppure no.

Le GANs sono utili per generare immagini realistiche seppur sintetiche, visto che evitano alcuni problemi matematici presenti nei metodi di allenamento precedenti, e rendono più facile ed efficiente l'apprendimento e la generazione di contenuto. Queste immagini sintetiche prodotte mediante GANs, all'inizio relativamente instabili, ultimamente hanno raggiunto una resa eccezionale[44]. Esse ed i modelli di reti neurali profonde sono responsabili dell'addestramento che sostituisce l'originale e fornisce come risultato un *deepfake* ottenuto mediante una combinazione ottimizzata di tecniche di codificazione e decodificazione. La codificazione riduce l'immagine ad uno spazio latente di dimensione minore, e la decodificazione la ricostruisce a partire dalla rappresentazione latente. Nel creare *deepfake*, lo spazio latente viene utilizzato per modellare le caratteristiche chiave e decodificare l'oggetto. Questo spazio deve includere le caratteristiche principali dell'oggetto, come i lineamenti del viso e le posizioni del corpo, in modo che il modello possa sovrapporre l'obiettivo sui lineamenti del viso e del corpo sottostanti nel video originale[45].

In origine, l'IA e il *deep learning* si erano concentrati sulla soluzione di problemi come il riconoscimento delle immagini, il miglioramento di fotografie e la sintesi di contenuto per applicazioni creative ed educative[46]. Tuttavia, le stesse tecniche potevano anche permettere di manipolare immagini e voci in maniera iperrealistica, come è avvenuto nell'uso di questi strumenti per creare contenuti

[43] I.J. Goodfellow *et al.*, "Generative Adversarial Networks", in *arXiv:1406.2661*, 2014.
[44] D. Saxena, J. Cao, "Generative adversarial networks (GANs): Challenges, solutions, and future directions", in *ACM Computing Surveys*, 54 (2021), n. 3, Article 63.
[45] D. Gamage, J. Chen, P. Ghasiya, K. Sasahara, *op. cit.*
[46] Sebbene in questo contributo vengano analizzate prevalentemente le utilizzazioni negative dell'algoritmo, e criticate sullo sfondo teoretico che la filosofa del diritto può offrire, queste tecniche come ogni altra hanno anche possibili impieghi positivi: per una sintesi descrittiva cfr. N. Ahmad, P. Chauhan, "Deepfake: Risks and Opportunities", in *Computer*, June 2024, pp. 141-144.

deepfake senza che fosse questo necessariamente l'intento iniziale. Le tecniche di trasferimento di stile e di interpolazione di immagini sviluppate in *deep learning* permisero inoltre di combinare elementi visivi di diverse immagini o video, rendendo possibile la manipolazione di elementi facciali e movimenti in nuovi video. Questi strumenti furono inizialmente programmati per usi artistici e di intrattenimento, ma con la loro applicazione alla creazione di video falsi diedero luogo a quel che oggi chiamiamo *deepfake*.

I *deepfake* si realizzano con un *software open source* (OSS in sigla per iniziali inglesi), la cui caratteristica è che il "codice fonte" (quello che permette che il *software* funzioni) è gratuitamente a disposizione del pubblico. Questo apre la possibilità che chiunque veda, modifichi e distribuisca detto *software* in accordo con i termini della licenza. Il vantaggio di questa modalità è quello di favorire la collaborazione e la trasparenza, visto che gli utenti possono migliorare il *software*, adattarlo alle proprie necessità e condividere le modifiche con la comunità degli utenti. Sono OSS tra gli altri Linux (un sistema operativo), Apache (un *web server*) e Mozilla Firefox (un navigatore *web*). Questo codice aperto è un algoritmo di "apprendimento automatico" (*machine-learning algorithm*). Con tale espressione s'intende un insieme di istruzioni che permettono ad una macchina o sistema meccanico di "apprendere" dai dati e migliorare la propria attività in compiti specifici con il trascorrere del tempo, senza necessità di essere espressamente programmato per ogni situazione possibile. Questi algoritmi analizzano schemi in grandi insiemi di dati, generando modelli che possono fare previsioni o prendere decisioni in funzione di quei dati.

La tecnologia di cui stiamo parlando genera o potrà generare in futuro un'enorme quantità di problemi etici e giuridici. Menzioniamo di seguito tre gruppi di principi che entrano in crisi a causa dello sviluppo e della diffusione di queste tecnologie, dai quali deriveranno inediti dilemmi regolatori:

a) Crisi di autenticità, caratterizzata dall'erosione della fiducia e della verità operativa. Non si produce soltanto un aumento esponenziale delle falsità in circolazione, ma anche la degradazione del valore probatorio spontaneo dell'evidenza audiovisiva[47]. Dinnanzi all'incertezza circa l'autenticità del registro, i cittadini cessano di ritenere la semplice esistenza di fotografie, audio o video ragione sufficiente per formarsi una convinzione ed agire. Questa perdita di fiducia incide sull'essenza stessa della comunità civile, e si ripercuote in maniera critica

[47] C. Vaccari, A. Chadwick, "Deepfake and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news", in *Social Media + Society*, 6 (2020), n. 1, pp. 1-13.

sulla deliberazione democratica[48], i processi giudiziari[49] ed il coordinamento sociale[50].

b) Opacità dell'identità personale e del cosiddetto "io digitale". I *deepfake* non soltanto presuppongono una sfida tecnica probatoria, ma anche una frattura nella stessa concezione del soggetto di diritto. Sottopongono ad interdetto la nozione di identità in tre dimensioni che si intrecciano. In primo luogo, spezzano la continuità narrativa (la costruzione del "chi sono io" attraverso la coerenza delle mie azioni pubbliche), giacché permettono di iniettarvi episodi vitali iperrealistici ma apocrifi che alterano retrospettivamente la biografia di una persona[51]. In secondo luogo, vulnerano la autenticità (il vincolo legittimo tra l'individuo e le sue espressioni), facilitando il "furto d'identità generativo", con il quale si usurpano gli elementi biometrici falsificando il contenuto del messaggio. In terzo luogo, manomettono la *agency*: attraverso una sofisticata mimesi funzionale, vengono fabbricate azioni che materialmente sembrano appartenere all'individuo, attribuendogli condotte e dichiarazioni alle quali non ha mai acconsentito né partecipato[52]. Questa dissociazione radicale tra la rappresentazione fisica e la volontà reale sfocia in una domanda filosofico-giuridica ineludibile per la teoria dell'imputazione: che cosa significa essere autore di un'azione quando l'apparenza materiale di paternità può venir fabbricata sinteticamente? Se le istituzioni legale si sono strutturate attorno alla presunzione che il registro audiovisivo diretto è un riflesso inequivoco della *agency* umana[53], il paradigma dell'IA generativa obbliga a ripensare i fondamenti stessi della responsabilità.

c) Crisi della rappresentazione: immagine/voce come segni. Un terzo effetto – meno appariscente ma forse più profondo – è che i *deepfake* alterano il regime ordinario di rappresentazione personale su cui riposano molteplici pratiche sociali e giuridiche. Nella vita quotidiana, il volto e la voce operano come marcatori privilegiati della persona: vedere un viso o udire una voce non equivale soltanto a ricevere informazione sensibile, bensì attiva un meccanismo pratico di attribuzione da cui si traggono, quasi automaticamente, inferenze sulla paternità, intenzione, presenza, e, pertanto, sulla responsabilità. Detto in altri termini: il volto e la voce

[48] P. Raj, "The impact of deepfake technology on elections and democratic trust", in *Indian Journal of Law*, 3 (2025), n. 6, pp. 56-64.
[49] R.A. Delfino, *Deepfake on trial 2.0: A revised proposal for a new Federal Rule of Evidence to mitigate deepfake deceptions in court*, Loyola Law School, Los Angeles Legal Studies Research, Paper No. 2025-10, 2025, February 15.
[50] Europol Innovation Lab., *Facing reality? Law enforcement and the challenge of deepfake: An observatory report from the Europol Innovation Lab*, Publications Office of the European Union, 2024.
[51] A. Verma, "Deepfake and the crisis of digital authenticity: Ethical challenges in the age of synthetic media", in *Journal of Information, Communication and Ethics in Society*, 24 (2025), n. 1, pp. 59-76.
[52] Europol Innovation Lab., *Facing reality? Law enforcement and the challenge of deepfake: An observatory report from the Europol Innovation Lab*, cit.
[53] R.A. Delfino, *op. cit.*

funzionano come segni che, in condizioni normali, *economizzano* l'identificazione e riducono i costi di verifica. I *deepfake* intervengono proprio nel punto in cui la comunità ripone una bassa soglia di dubbio ed una soglia elevata di fiducia. Essi riconfigurano questo meccanismo in tre passaggi. Dapprima, rendono "trasferibili" i lineamenti identificativi: estraggono la parvenza audiovisiva della persona e la separano dal contesto reale in cui quella persona ha parlato o agito. Il segno – volto o voce – cessa di essere associato all'"esser stato lì" e viene a dipendere dalla capacità tecnica di produrre una somiglianza funzionale. In secondo luogo, come conseguenza di detta separabilità, il materiale audiovisivo perde parte del proprio valore probatorio spontaneo: non evapora la sua forza persuasiva (anzi, potrebbe persino intensificarsi), ma si trasforma in una forza contaminata da una domanda preliminare ("sarà autentico?"). Infine, questo trasferimento si ripercuote su categorie giuridiche che presuppongono una connessione relativamente stabile tra segno e soggetto: il consenso (di che cosa sono io l'autore, quando la mia immagine o la mia voce possono venir riutilizzate e ricontestualizzate), la testimonianza (quale peso ha una dichiarazione audiovisiva se il suo indice personale è riproducibile), la reputazione (la mia identità pubblica può essere riscritta tramite segni che "sembrano provenire da me"), e la responsabilità (chi è colui che parla o agisce in una registrazione). Da questa prospettiva, il *deepfake* non è soltanto un "nuovo tipo di menzogna", bensì un intervento sul dispositivo sociale di attribuzione personale: si appropria del modo in cui la cultura identifica il soggetto attraverso segni sensibili e lo trasforma in un meccanismo replicabile.

Questo fenomeno può essere interpretato con speciale chiarezza a partire da Derrida, perché la teoria derridiana del segno viene elaborata proprio contro la supposizione che il senso e la identità siano garantiti da una presenza originaria pienamente accessibile. Sullo sfondo della fiducia quotidiana verso il volto e la voce opera una intuizione "metafisica": che certi segni, soprattutto la voce, ci mettono in contatto privilegiato con la presenza del soggetto (con la sua interiorità o la sua intenzione). Derrida diagnostica questo privilegio come parte di quello che lui chiama fono-logocentrismo: la tradizione occidentale tende a trattare la voce come figura di auto-presenza, come se nell'"udirmi parlare" il senso si desse in maniera immediata, senza mediazione. Il concetto derridiano decisivo per comprendere il *deepfake* non è solo la critica al privilegio della voce, ma anche la iterabilità. Derrida sostiene che ogni segno, per funzionare come tale, deve essere ripetibile e capace di operare fuori del proprio contesto originario[54]; da qui discende la sua idea secondo cui nessun contesto può chiudere definitivamente il funzionamento del marcatore ("*No context can enclose it*"). Questa struttura si formula in modo particolarmente utile per il nostro problema: il marcatore deve poter "funzionare"

[54] J. Derrida, "Firma evento contesto", in *Margini della filosofia*, (1971), trad. it. e cura di M. Iofrida, Einaudi, Torino, 1997, pp. 393-424.

anche in assenza del destinatario o dell'emittente, cioè, al di là della presenza empirica che, nell'esperienza ordinaria, sembra garantirla[55].

In questa cornice, il *deepfake* può essere descritto come una tecnologia di iterabilità applicata alla identità audiovisiva. Ciò che il montaggio, l'imitazione o il doppiaggio già mostravano su scala minore – che cioè il segno può venir separato dalla sua origine –, si trasforma in massivo, preciso ed economico tramite il *deepfake*, poiché prende il segno (volto e voce) come marcatori suscettibili di estrazione e reinserimento, e li fa circolare come se conservassero la presenza del soggetto. Quel che si intensifica non è un'anomalia esterna al segno, bensì una possibilità strutturale del segno stesso, cioè la sua capacità di operare "al di là" della presenza. Questa intensificazione apre, inoltre, un registro che Derrida esplora con la nozione di spettrale. In *Spettri di Marx* egli formula una tesi che risulta particolarmente espressiva nel caso del *deepfake*: "pedinare" non equivale ad essere presente; ci sono effetti di presenza che operano senza la presenza stessa[56]. Il *deepfake* provoca esattamente questo tipo di "apparizione": una presenza audiovisiva che incide su reputazioni, scatena decisioni e tenta di adempiere la funzione testimoniale, sebbene non sia supportata dall'avvenimento e dalla *agency* che fa apparire.

I *deepfake* non pongono un insieme differenziato di rischi, ma un problema comune: producendo segni audiovisivi (volto e voce) capaci di funzionare come presenza, disaccoppiano la rappresentazione del soggetto originale. Questo disaccoppiamento riconfigura le categorie con cui il diritto attribuisce fatti, imputa responsabilità e protegge beni personali[57]. Essi colpiscono in modo diretto i diritti della personalità, perché consentono di attribuire ad una persona comportamenti o detti che non ha mai realizzato, compromettendo il suo onore e la sua reputazione; inoltre, implicano l'uso non autorizzato del volto e, con sempre maggior frequenza, l'impiego della voce come elemento identificativo, trattabile come dato personale ed anche biometrico[58]. La difficoltà è tipicamente duplice: da un lato, quella di quantificare danni che molto spesso sono diffusi (perdita di opportunità, stigmatizzazione, umiliazione, danno morale), dall'altro, distinguere tra parodia legittima (o uso satirico/artistico) e lesione alla persona, specialmente quando il realismo del *deepfake* cancella i segnali contestuali che in condizioni normali

[55] J. Derrida, *Limited Inc.*, (1988), trad. it. e a cura di N. Perullo, Raffaello Cortina Editore, Milano, 1997.

[56] J. Derrida, *Spettri di Marx. Stato del debito, lavoro del lutto e nuova Internazionale. Nuova edizione*, (1993), trad. it. di G. Chiurazzi e A. Romani, Raffaello Cortina Editore, Milano, 2025.

[57] R. Chesney, D.K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security", in *California Law Review*, 107 (2019), n. 6, pp. 1753-1819, https://scholarship.law.bu.edu/faculty_scholarship/640/.

[58] R. Rini, L. Cohen, *op. cit.*; European Parliament and the Council of the European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Official Journal of the European Union*, 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

permetterebbero di riconoscere il genere o l'intenzione. Così, se la reputazione può venire intesa come un "archivio" di impronte circolanti (dichiarazioni, gesti, scene), la possibilità di fabbricare impronte altamente verosimili rende l'identità pubblica più vulnerabile all'altrui riscrittura: il soggetto può restare rappresentato da marcatori che lo imitano senza che lui sia nemmeno comparso.

Nella pratica, molti *deepfake* vengono costruiti con dati estratti da reti: fotografie, video ed audio disponibili pubblicamente o filtrati. Da qui scaturiscono problemi circa la base legale del trattamento, limiti al riutilizzo ed alla reidentificazione, e un campo particolarmente delicato: la "biometria derivata" (modelli o schemi di voce e volto costruiti a partire da tracce)[59]. A tutto ciò si aggiungono doveri di sicurezza e minimizzazione: quanto più persuasivi risultano i modelli, maggiore sarà l'incentivo a raccogliere e conservare grandi volumi di dati identificativi, il che incrementa l'esposizione a filtrazioni ed usi secondari. In questo scenario, il consenso diviene problematico sia nella sua definizione che nella sua prova. Ormai non basta un "sì" generico: ho dato il mio consenso per una foto, ma non per un modello di volto capace di generare scene potenzialmente infinite. Ho dato il consenso all'imitazione della voce in un determinato contesto, ma non in un altro. Che cosa significa "consenso informato" quando l'uso futuro è imprevedibile e la tecnica permette nuove forme di reinnesto?[60] Se l'immagine e la voce operano come marcatori riutilizzabili, il consenso smette di essere un atto semplice associato ad un contesto chiuso e diviene una questione di limiti di ricontestualizzazione: quali trasformazioni, quali usi, quali scopi e quali ambienti. Il diritto si vede costretto, di conseguenza, a precisare che cosa significhi autorizzare l'uso di lineamenti identificativi quando tali lineamenti possono essere tecnicamente "citati" ed innestati in scene nuove.

In sede giudiziaria, il *deepfake* colpisce al cuore il sistema probatorio: incide sull'autenticità, la catena di custodia ed i metadati; nelle perizie, sugli standard e sui costi di verifica; e nel rischio di esclusione ingiusta di una prova reale. La figura stabilisce una contrapposizione tra sicurezza probatoria ed accesso alla giustizia. Se elevare allo standard tecnico di autenticazione rende proibitive le attività peritali, si penalizzano le parti processuali dotate di minori risorse; se si mantengono standard lassi, si apre la porta alla contaminazione del processo da parte di materiale falso[61]. Parte della forza del registro audiovisivo proveniva dal fatto che palesava un "essere stato lì". La iterabilità tecnica introduce una sorta di "spettralità" in questa funzione: conserva la forma testimoniale (voce/volto, gesto, intonazione), ma separa questa

[59] *Ivi*; F. Romero-Moreno, "Deepfake detection in generative AI: A legal framework proposal to protect human rights", in *Computer Law & Security Review*, 58 (2025), pp. 106-162.
[60] Reuters, *Spain moves to curb AI deepfake, tighten consent rules on images*, January 13, 2026, https://www.reuters.com/business/media-telecom/spain-moves-curb-ai-deepfakes-tighten-consent-rules-images-2026-01-13/
[61] National Center for State Courts (2026), "AI-generated evidence is a threat to public trust in the courts, 24/02/2026", in *NSC*. Recuperato da: https://www.ncsc.org/resources-courts/ai-generated-evidence-threat-public-trust-courts, [Data di consultazione: 30/03/2026].

forma dalla presenza e dall'avvenimento. Perciò la discussione tende a spostarsi dal contenuto ("che cosa si è detto o si è fatto"), alla condizione di possibilità ("chi sta parlando qui", "in che senso questo è attribuibile"), forzando il sistema probatorio a dipendere sempre di più da dispositivi di autenticazione e tracciabilità.

In questo modo, i *deepfake* non si limitano ad amplificare la disinformazione: toccano un livello più basilare, quello dell'economia simbolica della presenza. Producendo segni di persona capaci di operare socialmente come presenza senza presenza, obbligano il diritto a ripensare – ed a rafforzare mediante dispositivi istituzionali – quel che prima si presumeva spontaneamente: la relazione tra rappresentazione, attribuzione e responsabilità. Per questo i diversi fronti (personalità, dati, area penale, prova, responsabilità civile, libertà d'espressione, proprietà intellettuale, etc.) possono venire interpretati come variazioni di uno stesso problema matrice, che è quello di identificare criteri di imputazione e tutela quando il segno di identità diventa tecnicamente replicabile.

5. *Deepfake*: la situazione regolatoria in Europa e negli Stati Uniti d'America

Di seguito, e conclusivamente, verranno presentate le principali prospettive legislative che affrontano il *deepfake*, insieme con alcune proposte affacciate sul piano internazionale, con speciale riferimento ad Europa e Nordamerica[62]. Come punto di partenza, conviene ricordare che l'Unione Europea definisce il *deepfake* – o "ultraimitazione" – come "un contenuto di immagine, audio o video generato o manipolato da una IA che assomiglia a persone, oggetti, luoghi, entità o avvenimenti reali e che può indurre una persona a pensare erroneamente che siano autentici o veridici"[63]. A partire da questa definizione, le strategie regolatorie mostrano due diversi stili: la normativa europea si presenta come una regolazione multilivello, mentre l'approccio nordamericano risulta frammentato e settoriale.

5.1. Europa: regolazione a scaglioni. La regolazione europea dei *deepfake* non si organizza ancora come un codice unico ed autosufficiente. Si configura, piuttosto, come una struttura a strati in cui il fenomeno viene disciplinato da: (a) il Regolamento Generale per la Protezione dei Dati (GDPR), quando la creazione o

[62] Per qualche ulteriore informazione e spunto in materia regolatoria, in particolare per riferimenti ai criteri normativi cui si è fatto ricorso in Cina, possono vedersi tra i tanti contributi quelli di F.V. Valenti, "Il deep fake: la nuova sfida dell'intelligenza artificiale generativa", in *Derecom*, 37 (2024), pp. 9-18, e di A. Orlando, "La regolamentazione del *deepfake* in Europa, Stati Uniti e Cina", in *MediaLaws*, numero speciale 2024, pp. 307-327 (https://www.medialaws.eu/rivista/la-regolamentazione-del-deepfake-in-europa-stati-uniti-e-cina/).

[63] European Parliament and the Council of the European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act). Official Journal of the European Union*, art. 3(60*)*, 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

diffusione del *deepfake* implica il trattamento di dati personali e biometrici[64]; (b) l'AI Act (2024), che contiene obblighi specifici di trasparenza per i contenuti sintetici e di ultraimitazione[65]; e (c) il Regolamento dei Servizi Digitali (DSA) (2022), che incide sull'ambiente di circolazione, specialmente nel disegno di interfacce e pratiche che ingannano o manipolano l'utente[66]. A ciò si aggiunge il diritto interno degli Stati europei (civile e penale), che solitamente inquadra il *deepfake* come una modalità tecnologica di illeciti già noti (contro la reputazione, contro la propria immagine, contro la diffusione non autorizzata di materiale intimo, etc.).

5.1.1. GDPR: i *deepfake* come trattamento di dati personali e biometrici. Il primo pilastro rilevante è il GDPR (27 aprile del 2016), il cui oggetto è la protezione delle persone fisiche circa il trattamento dei dati personali e l'individuazione di norme per la loro libera circolazione (*European Parliament & Council of the European Union*, 2016, art. 1). La connessione con i *deepfake* è diretta: nella misura in cui un *deepfake* di solito richiede la raccolta, archiviazione, elaborazione o diffusione di materiale che permette l'identificazione di una persona (immagine, video o audio), molte delle sue fasi entrano nell'ambito del regolamento (art. 2). Il GDPR definisce "dati personali" come ogni informazione su una persona identificata o identificabile (art. 4.1.). Nel caso del *deepfake*, il punto decisivo è la categoria dei dati biometrici, definita come dati personali ottenuti mediante un trattamento tecnico specifico relativi a caratteristiche fisiche, fisiologiche o comportamentali che permettono o confermano l'identificazione unica, includendovi espressamente "immagini facciali" (art. 4.14). In questo modo, il viso – spesso elemento centrale del *deepfake* – rientra chiaramente tra i dati biometrici; e, secondo l'utilizzo tecnico (per esempio, identificazione), anche la voce può operare come dato personale identificativo. Una volta stabilito questo quadro di riferimento, il GDPR passa ad esporre i principi di trattamento, le basi per la liceità, i diritti dell'interessato e gli obblighi del responsabile. In primo luogo, l'art. 5 esige che il trattamento sia lecito, leale e trasparente; che rispetti la limitazione di finalità; che si attenga a minimizzazione, esattezza, limitazione del tempo di conservazione e integrità e confidenzialità (art. 5.1.). Detti principi sono particolarmente rilevanti per il *deepfake*, perché la pratica è solita appoggiarsi sulla riutilizzazione di materiale identificativo e sulla sua ricontestualizzazione, forzando la limitazione di

[64] European Parliament and the Council of the European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)*, cit.
[65] European Parliament and the Council of the European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act)*, cit.
[66] European Parliament and the Council of the European Union, *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 (Digital Services Act). Official Journal of the European Union*, art. 25, 2022, https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

finalità e la minimizzazione, oltre ad aggravare i rischi di sicurezza e di esposizione a filtraggi.

In secondo luogo, il GDPR impone una base di liceità per trattare questi dati (art. 6). Nell'ambito dei *deepfake*, quando il trattamento è lecito, lo scenario più tipico è quello del consenso del titolare. Detto consenso deve soddisfare esigenze di chiarezza e separabilità, deve poter essere provato dal responsabile ed è per di più revocabile (artt. 6 e 7). Tale regime viene rafforzato quando si tratta di minori nei servizi della società dell'informazione (art. 8).

Nel capo III del GDPR vengono riconosciuti i diritti dell'"interessato" – la persona i cui dati vengono trattati – che risultano significativi rispetto ai *deepfake*, perché in essi il danno non proviene solo dall'ottenimento del dato (immagine o voce), bensì del suo utilizzo per generare contenuti e, soprattutto, dalla sua diffusione. Anzitutto, gli articoli da 12 a 14 stabiliscono un regime di trasparenza e informazione: il responsabile del trattamento deve comunicare, in maniera chiara ed accessibile, chi tratta i dati, con quale finalità, quali categorie di dati vengono impiegate, per quanto tempo, a chi vengono comunicati e quali diritti proteggano l'interessato. Quest'obbligo assume peculiare rilievo quando il trattamento avviene mediante applicazioni, piattaforme o servizi che catturano o processano volto e voce (per esempio, strumenti di "face swap" o clonazione della voce), poiché l'opacità informativa è esattamente quel che di solito impedisce alle persone di comprendere e controllare l'uso della propria identità digitale (artt. 12-14). L'art. 15 riconosce il diritto di accesso, che permette di pretendere dal responsabile la conferma del fatto che si stiano o meno trattando dati personali e, in caso affermativo, di accedere ad una copia di essi ed all'informazione dettagliata sul trattamento. Nel contesto dei *deepfake*, questo diritto può operare come un meccanismo di "apertura della scatola nera", abilita cioè l'interessato a verificare se un servizio immagazzina materiale associato alla sua identità (per esempio, immagini originali, registrazioni della voce, elementi biometrici o profili derivati), con quale base giuridica lo fa e con quali scopi. L'art. 16, da parte sua, consacra il diritto di rettifica, che, sebbene sia stato concepito principalmente per correggere dati inesatti, può acquistare rilevanza quando esistono errori di attribuzione o riferimento a registri o metadati vincolati all'elaborazione o diffusione del contenuto (artt. 15-16). Infine, il diritto più decisivo per il *deepfake* è il diritto di cancellazione, nell'art. 17 – frequentemente descritto come "diritto all'oblio" – perché il danno tipico si intensifica con la pubblicazione e la replicazione. L'interessato può sollecitare la cancellazione quando il trattamento sia illecito, quando ritira il consenso, quando i dati non siano più necessari per la finalità invocata o quando si manifesti un'opposizione legittima al trattamento. In caso di *deepfake*, tali causali diventano specialmente pertinenti, visto che molte volte l'immagine o la voce sono impiegate senza autorizzazione o vengono riutilizzate fuori da qualsiasi finalità compatibile. Inoltre, l'art. 19 aggiunge un elemento cruciale per scenari di circolazione. Quando si debba procedere a cancellazione o rettifica, il responsabile deve comunicarlo ai destinatari a cui abbia divulgato i dati, salva l'impossibilità o lo sforzo sproporzionato. Quest'obbligo mira a mitigare il

cosiddetto "effetto moltiplicatore" caratteristico del *deepfake*: non basta eliminare il contenuto all'origine, se il materiale si è ormai propagato a terzi o si è inserito in circuiti di redistribuzione (artt. 17 e 19). Il regolamento impone obblighi organizzativi e tecnici al responsabile (*privacy by design* e per difetto, registri, cooperazione con autorità), prevede risarcimento dei danni (art. 82), sanzioni amministrative (art. 83), e pretende di conciliare la protezione dei dati con la libertà di espressione ed informazione (art. 85), questione specialmente delicata per *deepfake* satirici o informativi quando si invoca la finalità espressiva.

Come si può constatare, il GDPR non regola il *deepfake* come tecnica, ma sì lo regola in quanto presuppone il trattamento dei dati personali e biometrici, articolando principi, liceità, diritti, doveri di sicurezza e rimedi.

5.1.2. AI Act (2024): la trasparenza sui contenuti sintetici e l'ultraimitazione. Il secondo pilastro europeo è l'AI Act (13 giugno del 2024), che stabilisce regole armonizzate per i sistemi di IA. La sua rilevanza per i *deepfake* si manifesta, da un lato, nella speciale attenzione alla biometria ed a determinate pratiche proibite; e, dall'altro lato, in un regime esplicito di trasparenza per i contenuti sintetici e l'ultraimitazione. L'AI Act include una definizione di dati biometrici coerente con quella del GDPR, ribadendo che le immagini facciali possono essere biometriche quando si ottengono mediante trattamento tecnico specifico (art. 3.34). Inoltre, proibisce pratiche come la creazione o l'ampliamento di basi di dati di riconoscimento facciale mediante estrazione non selettiva di immagini facciali da internet o CCTV per fini specifici, il che evidenzia una politica restrittiva nei confronti della cattura massiva di biometria (art. 5.1e). Peraltro, il nucleo più direttamente applicabile ai *deepfake* è la trasparenza: l'AI Act impone obblighi ogniqualvolta i sistemi generino o manipolino contenuti sintetici (audio, immagine, video, testo), e, in particolare, quando costituiscono ultrafalsificazione/*deepfake*, esigendo che si informi che il contenuto è stato generato o manipolato artificialmente, con sfumature ed eccezioni che vanno lette in armonia con la libertà di espressione (art. 50). Questo disegno cerca di ridurre l'inganno mediante *disclosure* più che mediante una generale proibizione della tecnica.

5.1.3. DSA (2022): piattaforme, interfacce e manipolazione. Il terzo "strato" regolatorio è il DSA (2022), il cui apporto non è quello di definire il *deepfake*, quanto piuttosto di incidere sul suo ecosistema di circolazione. In particolare, l'art. 25 proibisce che i fornitori di piattaforme in linea disegnino o operino con le proprie interfacce in modo da ingannare o manipolare i destinatari del servizio oppure distorcano materialmente la loro capacità di prendere decisioni libere ed informate (art. 25). Questa disposizione è rilevante per il caso che ci occupa quando il servizio agevola l'efficacia dell'inganno, avvenga questo per disegno dell'interfaccia, per assenza di segnali contestuali o per meccanismi che ostacolano le decisioni informate.

5.1.4. L'ultimo livello è il diritto interno europeo. Insieme alla cornice del diritto UE, il diritto interno degli Stati membri colloca il *deepfake* tra le modalità classiche di illecito. In Francia, il *Code pénal* sanziona la divulgazione di un montaggio realizzato con parole o immagini di una persona senza il suo consenso, quando non risulti evidente che si tratta di un montaggio o non venga indicato espressamente (*Code pénal*, art. 226-8). In Spagna, la Ley Orgánica 1/1982 prevede la protezione civile dell'onore, dell'intimità personale e familiare e della propria immagine nei confronti di intromissioni illegittime, offrendo un rimedio civile applicabile per gli usi non autorizzati dell'immagine e di danni alla reputazione (Ley Orgánica 1/1982, de 5 de mayo, art. 1). Da parte sua, il Regno Unito – uscito dall'UE, ma rilevante in Europa – combina il *Data Protection Act* 2018 con l'*Online Safety Act* 2023, rafforzando meccanismi di sicurezza, rapporto e rimozione in servizi regolati, con impatto pratico nei riguardi dei contenuti dannosi che circolano sulle piattaforme (*Data Protection Act* 2018, c. 12; *Online Safety Act* 2023, c. 50).

In Italia, l'intera problematica è stata disciplinata dall'art. 26, lettera c), della legge n. 132 del 23 settembre 2025, che introduce l'art. 612-*quater* del Codice Penale ed il nuovo delitto di "Illecita diffusione di contributi generati con sistemi di intelligenza artificiale". Al di là dei dettagli regolatori, è interessante di questa disciplina evidenziare il ricorso a principi chiave da cui discendono poi i numerosi interventi di modifica del Codice Civile e penale che accolgono, nello spirito dell'AI Act europeo[67], la rivoluzione digitale provocata dall'avvento dei sistemi di IA. Tali principi sono enumerati nell'art. 3 della legge ("Principi generali"), dove al primo comma si esige un utilizzo dei sistemi di IA "nel rispetto dei diritti fondamentali e delle libertà previste dalla Costituzione, del [*sic*] diritto dell'Unione europea e dei [*sic*] principi di trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, accuratezza, non discriminazione, parità dei sessi e sostenibilità".

Il secondo comma insiste sui dati e processi realizzativi, "di cui devono essere garantite e vigilate la correttezza, l'attendibilità, la sicurezza, la qualità, l'appropriatezza e la trasparenza", mentre il terzo comma sottolinea la necessità di preservare "l'autonomia ed il potere decisionale dell'uomo", "assicurando la sorveglianza e l'intervento umano", ed il quarto comma mira a salvaguardare "lo svolgimento con metodo democratico della vita istituzionale e politica". Sempre in relazione all'Italia, merita di essere ricordato – perché a suo modo paradigmatico di uno stile regolatorio multilivello come quello che stiamo presentando – il ruolo sempre più incisivo che hanno assunto le autorità amministrative indipendenti, *in subjecta materia* il Garante per la Protezione dei Dati Personali[68]. Questo ente ha

[67] Come traspare anche nell'articolo più rappresentativo della legge, il n. 3, al comma 6, dove l'accento viene posto sulla cybersicurezza e sull'approccio "proporzionale e basato sul rischio".
[68] Il sito istituzionale del Garante è https://www.garanteprivacy.it/home, dove è possibile reperire documentazione abbondante.

negli anni elaborato una normativa ed una giurisprudenza molto accurate, che già costituiscono un *corpus* di riferimento ricchissimo di indicazioni e rivelatore di orientamenti.

5.2. Stati Uniti e Canada. In Nordamerica, la regolazione dei *deepfake* si è sviluppata in forma assai diversa rispetto alla strategia europea. Il modello dominante – specialmente negli USA – è frammentario, giacché convivono risposte statali (per esempio, pornografia non consentita ed elezioni), iniziative federali orientate alla trasparenza ed ai diritti sulle replicazioni digitali, e una regolazione amministrativa che persegue la falsificazione e la frode con strumenti di settore. Questa traiettoria asimmetrica si spiega, almeno in parte, per il peso del Primo Emendamento, che rende costose le restrizioni ampie del discorso; perciò, il legislatore tende a formulare norme ritagliate in vista di finalità (specialmente violenza sessuale basata sull'immagine) o di doveri di *disclosure*[69].

5.2.1. Stati Uniti: Stati, Congresso ed agenzie. Le prime norme esplicite sono sorte a livello statale, con l'obiettivo di tutelare l'intimità nei confronti della pornografia sintetica. Un primo intervento fu quello della Virginia (2019), dove la *House Bill* 2678 aggiornò la regolazione della pornografia non consentita per includere premesse compatibili con contenuti sintetici, inaugurando in tal modo una via che oggi si riconosce come centrale nel trattare determinati *deepfake* come *Non-Consensual Intimate Imagery* (NCII), o immagini intime diffuse senza consenso (Virginia General Assembly, 2019).

Sul piano federale, le proposte si sono organizzate attorno a tre perni. Il primo è quello della trasparenza e della tracciabilità: il *Deepfakes Accountability Act* (2023) non proibisce la tecnica, ma esige una marcatura ed una identificazione del contenuto sintetico e attribuisce conseguenze all'omessa divulgazione quando vi sia una finalità malevola[70]. Il secondo perno è quello della protezione dell'identità: il *No Fakes Act* (2024; reintrodotto nel 2025) cerca di riconoscere un diritto federale sulla voce e sull'immagine in contrasto con "repliche digitali" non autorizzate, con

In relazione al fenomeno del *deepfake*, già nel 2020 è stato diffuso un Vademecum (https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9512226), integrato più recentemente (https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10190269). Per una recente determinazione può utilmente leggersi il Provvedimento del 18 dicembre 2025 (https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10207132).

[69] Y.D. Clarke (2023), "Clarke leads legislation to regulate deepfake", in *Yvette D. Clarke*. Recuperato da: https://clarke.house.gov/clarke-leads-legislation-to-regulate-deepfake/, [Data di consultazione: 30/03/2026].

D. Durbin (2026), *Durbin successfully passes bill to combat nonconsensual, sexually-explicit deepfake images*, in *U.S. Senator Dick Durbin of Illinois*. Recuperato da: https://www.durbin.senate.gov/newsroom/press-releases/durbin-successfully-passes-bill-to-combat-nonconsensual-sexually-explicit-deepfake-images, Data di consultazione: 30/03/2026].

[70] Y.D. Clarke, *op. cit.*

la sfida di conciliare questa tutela con eccezioni legate al Primo Emendamento ed al *fair use*[71]. Il terzo perno – quello di sviluppo legislativo più verificabile alla data attuale – è il *Defiance Act*, orientato a *deepfake* sessuali non consentiti, che attribuisce una causa di azione civile federale per le vittime di falsificazioni intime[72].

Parallelamente, le agenzie hanno operato con strumenti preesistenti. Nel febbraio del 2024 la FCC ha sostenuto che le voci generate dall'IA vanno considerate "artificiali" agli effetti della TCPA, rafforzando la persecuzione delle *robocalls* e di truffe basate sulla clonazione della voce[73]. In materia elettorale, l'avanguardia continua ad essere statale: norme come quella della California (AB 2655) introducono obblighi associati alla disinformazione sintetica in periodi critici di campagna, sebbene queste iniziative affrontino spesso sfide costituzionali e tensioni sulla responsabilità degli intermediari[74].

5.2.2. Canada: strategia precoce e difficoltà di consolidare una legge integrale. Il Canada costituisce un caso eloquente di sfasamento tra *leadership* strategica precoce e consolidamento normativo vincolante. Nel 2017, il Paese adottò la *Pan-Canadian Artificial Intelligence Strategy*, iniziativa coordinata dal CIFAR e orientata a rafforzare capacità nazionali di ricerca, formazione del talento e sviluppo di ecosistemi di innovazione sull'IA[75]. Tuttavia, il fatto di aver articolato con grande anticipo una politica pubblica di incentivo non si è tradotto, finora, in una cornice legislativa integrale idonea ad imporre doveri generali e rimedi sistematici di fronte ai rischi associati a contenuti sintetici, piattaforme e sistemi di IA. Gli obiettivi più ambiziosi di "giuridificazione" di quest'agenda si sono concentrati in due progetti recenti. In primo luogo, l'*Online Harms Act* (*Bill* C-63),

[71] C. Coons, M. Blackburn, A. Klobuchar, T. Tillis, *No Fakes Act of 2024* (One-pager). https://www.coons.senate.gov/wp-content/uploads/media/doc/no_fakes_act_one-pager.pdf.

P. Manatt, L.L.P. Phillips, *Senators officially introduce No Fakes Act with digital replica protections* (Client alert, 2024, August 5). https://www.manatt.com/insights/newsletters/client-alert/senators-officially-introduce-no-fakes-act-with-di.

[72] D. Durbin, *op. cit.*

The 19th. (2026, January 13). *Senate moves to let victims of sexually explicit deepfake sue*. https://19thnews.org/2026/01/senate-defiance-act-nonconsensual-images-deepfakes/?utm_source=chatgpt.com

[73] Federal Communications Commission (2024), "FCC makes AI-generated voices in robocalls illegal", in *Federal Communications Commission*. Recuperato da: https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal, [Data di consultazione: 30/03/2026].

Federal Communications Commission, "Implications of artificial intelligence technologies on protecting consumers from unwanted robocalls and robotexts" (FCC 24-17; Declaratory Ruling), consultabile in https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf.

[74] California Legislative Information, "Defending Democracy from Deepfake Deception Act of 2024, Assem. B. 2655, 2023–2024 Reg. Sess. (Cal. 2024)", consultabile in. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2655.

[75] CIFAR (2017), "Pan-Canadian Artificial Intelligence Strategy", in *CIFAR*. Recuperato da: https://cifar.ca/ai/, [Data di consultazione: 30/03/2026].

presentato nel 2024, che ha cercato di stabilire un regime di doveri e responsabilità per i servizi in linea con riguardo ai contenuti dannosi, con particolare incidenza sulla protezione dei minori. Ciononostante, il progetto non ha raggiunto lo *status* di legge, perché la procedura di approvazione si è interrotta a seguito della proroga parlamentare dell'inizio del 2025, la qual cosa imporrebbe di reintrodurlo se si intende continuare il processo di approvazione[76]. In secondo luogo, il *Artificial Intelligence and Data Act* (AIDA), incluso nel *Bill* C-27, aspirava a regolare i sistemi di IA ad alto impatto mediante obblighi di governo, mitigazione dei rischi e responsabilità istituzionale. Nemmeno l'AIDA è culminato nell'adozione legislativa e, per il medesimo effetto procedimentale, richiede di ricominciare il proprio *iter* per trasformarsi in norma vigente[77].

Il caso canadese mostra che i progressi normativi non dipendono solamente dalla diagnosi tecnica o dalla disponibilità di strumenti concettuali (strategie nazionali), ma anche da condizioni politiche e dall'architettura legislativa. L'inserimento di AIDA in un progetto troppo ampio (*Bill* C-27) e l'interruzione parlamentare che ha colpito tanto C-63 quanto C-27, riflettono un problema ricorrente nella regolazione degli ecosistemi digitali, vale a dire la difficoltà di allineare in un unico testo normativo oggetti eterogenei (protezione di minori, doveri delle piattaforme, governo delle IA ad alto impatto) senza accrescere frizioni procedimentali e controversie politiche. Il risultato, per il momento, è che il Canada allo stato attuale non ha un regime legale integrale e stabile comparabile al tentativo regolatorio europeo[78].

6. Riflessioni conclusive

Come è emerso nell'analisi delle incipienti risposte regolatorie a livello globale, il diritto si trova dinnanzi ad uno scenario inedito. Dalla spiegazione che precede è possibile percepire che la tecnologia *deepfake* si mostra come un caso limite che rende specialmente evidente il problema generale della tecnologia contemporanea: la tendenza a trasformare la realtà in una superficie manipolabile fino a mettere in crisi la distinzione tra originale e copia.

[76] Government of Canada, "Proposed bill to address online harms (Bill C-63 status note)", 2025, consultabile in https://www.canada.ca/en/canadian-heritage/services/online-harms.html.

[77] D. Fraser. S. Anderson Dykema (2025), "The demise of the Artificial Intelligence and Data Act (AIDA): 5 key lessons", in *McInnes Cooper*. Recuperato da: https://www.mcinnescooper.com/publications/the-demise-of-the-artificial-intelligence-and-data-act-aida-5-key-lessons/, [Data di consultazione: 30/03/2026].

Parliament of Canada (2025), "Bill C-27 (44-1): Digital Charter Implementation Act, 2022", in *LEGISinfo.* Recuperato da: https://www.parl.ca/legisinfo/en/bill/44-1/c-27 [Data di consultazione: 30/03/2026].

[78] CIFAR (2017), *op. cit.*; Government of Canada (2025). *op. cit.*; Parliament of Canada (2025). *op. cit.*

Non si tratta soltanto di una nuova tecnica di falsificazione nell'ordinario significato della parola, ma di un dispositivo idoneo a produrre segni di presenza (volto, voce, gesto, intonazione) che operano socialmente anche in assenza della persona o dell'avvenimento rappresentati. Il risultato non è semplicemente un aumento di falsità, ben di più attinge un carattere strutturale che incide sulla verità operativa delle immagini e dei suoni, cioè sulla loro idoneità a valere come ragioni naturalmente sicure per la decisione e l'azione. Quando la forma della testimonianza può essere generata senza l'evento, l'ordine discorsivo diventa più vulnerabile a causa non solo della falsità in senso materiale, ma anche dell'incertezza simbolica inerente all'operatività della fiducia, e di conseguenza la base stessa della comunità politica.

Occupandosi di altro tema, quello della metamorfosi delle relazioni interpersonali come esito letale della digitalizzazione e della datificazione[79] universali, Byung-Chul Han osserva in maniera pertinente anche al nostro discorso che

> la comunicazione digitale danneggia considerevolmente le relazioni umane. Oggi siamo ovunque collegati senza tuttavia essere *legati* gli uni agli altri. La comunicazione digitale è estensiva. Le manca qualsiasi intensità. Connessione non è sinonimo di *relazione*. Il *tu* viene sostituito ovunque dall'*Es*. La comunicazione digitale abolisce l'*interlocutore personale*, il *volto*, lo *sguardo*, la *presenza corporea*. In tal modo, accelera la *scomparsa dell'Altro*. Gli spettri abitano l'*inferno dell'Uguale*[80].

Nel caso del *deepfake*, queste critiche assumono un rilievo ancor più decisivo: non è solo la forma della relazione che viene tradita dalla sua replica digitale, ma lo stesso contenuto. Viene intaccato dalla manipolazione informatica non solo il veicolo comunicativo, con il tradimento della presenza; ma anche, e con inedita gravità, il contenuto veritativo che si vorrebbe condividere. In altri termini, la nuova tecnologia volta le spalle sia all'esperienza della presenza che alla comunicazione della verità: e tanto più allora valgono le considerazioni che altrove sviluppa il filosofo sudcoreano, secondo il quale "nell'era delle fake news, della disinformazione e delle teorie del complotto, stiamo perdendo la realtà e le verità fattuali. 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"[81].

[79] Sul problema della datificazione rimandiamo a C. Sarra, *Il mondo-dato. Saggi su datificazione e diritto*, Cleup, Padova, 2022.
[80] B.-C. Han, *Le non cose. Come abbiamo smesso di vivere il reale*, (2021), trad. it., Einaudi, Torino, 2023, p. 60.
[81] B.-C. Han, *Infocrazia. Le nostre vite manipolate dalla rete*, (2021), trad. it., Einaudi, Torino, 2023, p. 60.

Questo fenomeno è destinato, secondo la lettura che di Nietzsche ci offre lo stesso Han, a corrompere del di dentro la società, la vita pubblica, l'esperienza politica.

> Nietzsche constaterebbe [oggi] che la conoscenza ha perso completamente la *spinta alla verità*, la *volontà di verità*. Solo una società intatta sviluppa la spinta alla verità. L'affievolirsi della spinta alla verità e la disintegrazione della società sono reciprocamente dipendenti. La crisi della verità si diffonde laddove la società si disgrega in raggruppamenti o tribù, tra i quali non è più possibile alcuna intesa, alcuna *designazione vincolante delle cose*. Nella crisi della verità si perde il mondo comune, anzi il linguaggio comune. La verità è un regolatore sociale, un'idea regolatrice della società[82].

La cosa è tanto più seria se la società concepisce se stessa come democratica, come fondata su interazioni comunicative, sul dialogo tra i singoli e tra le generazioni, e sull'affidabilità dei contenuti discorsivi di cui s'intesse questa comunicazione continua.

Se le cose stanno così, siamo in effetti davanti ad una sfida enorme per il diritto, la cui azione si basa sulla determinazione genealogica dell'identità di soggetti ed azioni, e la cui funzione coesistenziale è al contempo generata dalla fiducia ed alla fiducia orientata. Detta trasformazione incide direttamente su categorie giuridiche fondamentali. L'identità e la paternità, che reggono l'imputazione e la responsabilità, non possono più appoggiarsi con la stessa semplicità sull'immediata evidenza, perché la prova audiovisiva diminuisce la propria forza presuntiva, la reputazione si trasforma in un *file* esposto a scritture altrui, e il consenso, da atto specifico, diventa un problema di ricontestualizzazione (quali usi, quali metamorfosi, in quali circostanze).

Così, l'erosione della differenza tra originale e copia – la cui epitome è il *deepfake* – sembra obbligare il diritto a riconoscere che la stabilità del vincolo tra segno e soggetto non è un dato naturale, bensì una elaborazione istituzionale che richiede di divenire sempre più esplicita, orientarsi verso il tecnico e cercare nuove forme di riconoscimento. Il *deepfake* non è un'eccezione, ma la figura più chiara di questo paradigma in cui la tecnica incrementa indefinitamente il potere della copia, fino a rendere opaca la stessa nozione di origine. Questo è forse il punto più problematico che deriva dalla confusione di entrambi i termini del binomio, e quello che potrebbe avere maggiori effetti normativi. L'opposizione "autentico/falso" corre il rischio di risultare insufficiente, perché ciò che scompare è la possibilità sociale di stabilire l'autenticità con mezzi ordinari. Da qui che la risposta normativa al fenomeno che ci occupa non può ridursi alla dimensione persecutoria o punitiva, né all'astratta proibizione di una tecnica destinata a mutare rapidamente – ed è proprio questo che sembra succedere con la perplessità che caratterizza gli strumenti legali che abbiamo passato prima in rassegna. Si rende necessario,

[82] *Ivi*, pp. 61 s.

piuttosto, un ripensamento della risposta a diversi livelli: protezione dei dati e biometria, obblighi di trasparenza e tracciabilità, regole di circolazione e disegno di piattaforme, e un rafforzamento delle garanzie probatorie mediante standard condivisi, catene di custodia e meccanismi di autenticazione adeguati.

Tutto questo sul piano pratico. Peraltro, è soprattutto indispensabile provvedere a una riformulazione dei principi legati alla preminenza dell'essere umano considerato come una realtà sostanziale permanente, nonostante la volatilità dell'universo digitale. Questo richiede un accurato lavoro, che deve realizzarsi secondo le fogge di una collaborazione trasversale di esperti ed autorità pubbliche di aree differenti (molto al di là del diritto nazionale), che ripropongano la possibilità di mediare nel riconoscimento attraverso strumenti giuridici nuovi (molto al di là del diritto internazionale). Tutto ciò, a partire dalla premessa che l'autenticità non è più un valore spontaneo nella vita in comune, ma semmai il risultato che va garantito istituzionalmente, tenendo come punto di appoggio la singolarità dell'uomo concreto del mondo analogico.

Per tutto quel che si è detto, azzardiamo l'ipotesi che il *deepfake* sia la rappresentazione più nitida della sfida tecnologica contemporanea: una tecnica che non si limita a cambiare ciò che possiamo fare con le immagini, ma si spinge fino a mutare quel che le immagini possono significare per i soggetti umani.