LUISS

TESI DI LAUREA MAGISTRALE A CICLO UNICO

DIPARTIMENTO DI GIURISPRUDENZA

CATTEDRA DI MACCHINE INTELLIGENTI E DIRITTO

IA E DEEPFAKES: LA PROTEZIONE DELLA REPUTAZIONE DIGITALE NELL'ERA DELLE REPLICHE CREDIBILI

Chiar.mo Prof. Angelo Costanzo
RELATORE

Chiar.mo Prof. Giovanni Buonomo
CORRELATORE

Sergio Pompilio

CANDIDATO

Matricola n. 167033

A.A. 2024/25

A me stesso.

Al "percorso" e a chiunque ne abbia fatto parte, lasciando un pezzo di sé in me.

Alle "comuni" persone:

a chi colora la nostra vita in silenzio,

sentendosi buono in tutto, ma bravo in niente.

L'importante è restare protagonisti per sé stessi.

“We are the people that rule the world,

A force running in every boy and girl”

(Empire Of The Sun – We Are The People)

INDICE

INTRODUZIONE

Negli ultimi anni, la rapida evoluzione dei sistemi di intelligenza artificiale (IA) ha introdotto nuovi strumenti in grado di modificare profondamente molteplici aspetti della vita istituzionale e sociale. È dunque inevitabile la reazione del diritto, che ha dovuto fronteggiare questioni inedite e rivisitare istituti giuridici storicamente consolidati.

In un simile scenario, tra le applicazioni dell'IA che presentano maggiori criticità si collocano i c.d. *deepfakes*. Attraverso tali tecnologie, di cui sarà fornita una definizione compiuta più avanti (v. *infra*, cap. I, 1.), è possibile generare contenuti audiovisivi fortemente realistici che riproducono l'immagine o la voce di un soggetto a prescindere dal consenso di quest'ultimo. Ne deriva che il principale bene giuridico leso è l'identità personale e, in particolare, due dei suoi elementi costitutivi, vale a dire l'immagine e la voce.

Di conseguenza, il presente elaborato, prima ancora di affrontare tecnicamente il fenomeno dei *deepfakes*, intende soffermarsi sui profili giuridici connessi. Nel dettaglio, il primo capitolo assume una duplice chiave di lettura. Anzitutto, sono trattati i concetti di «identità», «immagine» e «voce», nonché la loro evoluzione nel corso dei secoli in una prospettiva teorico-filosofica. Segue un inquadramento del diritto all'identità personale, sia come inteso nelle principali Carte fondamentali ─ quali la Convenzione Europea dei Diritti dell'Uomo (CEDU), la Carta dei Diritti Fondamentali dell'Unione Europea (CDFUE, o Carta di Nizza) e la Costituzione italiana ─ sia nei principali approdi della giurisprudenza italiana relativi alla «reputazione» e alla «dignità», beni giuridici che assumono grande rilevanza, non solo in quanto spesso erroneamente sovrapposti all'identità personale, ma anche perché, come sarà evidenziato (v. *infra*, cap. I, 3.), possono essere lesi dai *deepfakes* singolarmente o congiuntamente.

Ricostruiti i profili giuridici interessati, il secondo capitolo approfondisce l'analisi del fenomeno. A tal fine, esso comincia da un inquadramento della tecnologia alla base dei contenuti sintetici, ossia l'IA generativa, oltre che delle sue applicazioni e dei principali modelli sviluppatisi nell'ultimo decennio.

Successivamente, l'attenzione si concentra sui *deepfakes* in senso stretto, esaminandone la storia, le diverse finalità perseguite, i criteri per distinguerli dai contenuti reali e le tecnologie attualmente utilizzate per la loro creazione. Tra queste ultime, la tecnica del c.d. *voice cloning*, sottocategoria delle voci sintetiche e utilizzata nell'ambito dei *deepfake* audio, è oggetto di uno specifico paragrafo. Inoltre, per una migliore comprensione del funzionamento dell'IA generativa, si introduce il concetto di *dataset*, il cui utilizzo per la generazione di contenuti sintetici può dar luogo a un'ipotesi di *training* non autorizzato.

Tuttavia, è importante chiarire che il presente scritto, pur avendo ad oggetto i *deepfakes*, non ha come unico obiettivo quello di fornirne un'analisi descrittiva. Difatti, il lavoro di ricerca si interroga principalmente sulla capacità degli ordinamenti di regolare il fenomeno, verificando quali siano gli strumenti normativi disponibili e se questi, allo stato attuale, risultino adeguati allo scopo.

A tali quesiti è fornita risposta nel terzo capitolo, fulcro dell'elaborato e incentrato su alcuni dei principali sistemi giuridici. Anzitutto, è esaminato l'approccio assunto nell'Unione Europea, la quale regola i *deepfakes* sia indirettamente ─ mediante il Regolamento 2016/679 (GDPR), il Regolamento 2022/2065 (Digital Services Act) e la normativa comunitaria in materia di diritto d'autore e diritto all'immagine ─ sia direttamente con il più recente Regolamento 2024/1689, meglio noto come *AI Act*. In secondo luogo, rileva il contesto degli Stati Uniti, in cui il fenomeno è stato inizialmente disciplinato da strumenti preesistenti come il *Copyright Act* e l'istituto del *right of publicity*, per giungere solo di recente a una regolazione espressa, seppur ancora limitata (v. *infra*, cap. III, 2.3.), con il *Take It Down Act*. Ancora, è considerato l'ordinamento cinese, caratterizzato da un modello di regolamentazione opposto a quello c.d. *multistakeholder* occidentale e che, come avrà modo di vedersi (v. *infra*, cap. III, 3.), si distingue a livello mondiale per la rapidità e completezza degli interventi normativi in materia di *deepfakes*. Infine, si volge uno sguardo all'ordinamento italiano, primo fra gli Stati membri dell'Unione Europea ad essersi adeguato

all'AI Act con la legge n. 132/2025, incriminando la diffusione di «contenuti generati o alterati con sistemi di intelligenza artificiale»[1].

Dopo aver messo a confronto i diversi approcci in relazione ai *deepfakes*, il terzo capitolo si conclude affrontando il loro impatto a livello processuale, capace di mettere in crisi non solo i metodi tradizionali di valutazione della prova, ma anche la credibilità del materiale audiovisivo nel processo, sempre più esposto al rischio di confusione tra realtà e finzione.

Il quarto e ultimo capitolo, dedicato alle nuove proposte normative in risposta al fenomeno, completa il quadro normativo fornito in precedenza. Infine, sono presenti riflessioni sulla possibilità di un'utilizzazione etica dei *deepfakes*, evidenziando l'insufficienza della sola dimensione normativa a tale scopo.

[1] Questa la formulazione assunta dal nuovo art. 612-quater c.p., introdotto dall'art. 26 della L. 132/2025, il quale sarà oggetto di specifico approfondimento nel terzo capitolo dell'elaborato.

CAPITOLO I

L'IDENTITÀ E L'EVOLUZIONE GIURIDICA

SOMMARIO: 1. Identità, immagine, voce: concetto ed evoluzione. – 1.1. Identità. – 1.2. Immagine. – 1.3. Voce. – 2. Il diritto all'identità nelle carte fondamentali (CEDU, Carta di Nizza, Costituzione italiana). – 3. Profili giurisprudenziali di identità personale, reputazione e dignità.

1. Identità, immagine, voce: concetto ed evoluzione

L'umanità sta varcando la soglia di un'epoca storica in cui il digitale non costituirà più mero accessorio alla propria quotidianità, ma ne andrà a rappresentare il perno centrale. Non a caso si parla sempre più frequentemente di "società digitale", espressione che indica il costante aumento delle relazioni tra gli individui, le imprese e le pubbliche amministrazioni fondate sullo scambio in rete di dati e informazioni[2]. Sebbene questo fenomeno sia in atto ormai da tre decenni, ciò che costituisce un discrimine tra il passato (neanche troppo remoto) e il presente è l'avvento dell'intelligenza artificiale (IA).

Una delle applicazioni dell'IA (in particolare di quella c.d. "generativa") più sofisticate e, soprattutto, controverse è data dai c.d. *deepfakes*. Di questi ultimi, così come dell'IA generativa, si parlerà compiutamente più avanti. Per quel che ora ci interessa, occorre brevemente indicare cosa si intende per *deepfakes*. Una definizione esaustiva proviene dal Garante per la protezione dei dati personali: essi «sono foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce»[3]. La formulazione del Garante è molto chiara e permette di

[2] F. Pizzetti, *Con AI Verso la Società digitale*, in Federalismi, n. 23, 2023, p. 4

[3] Garante per la protezione dei dati personali, *Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*, 2020. Reperibile al sito: https://garanteprivacy.it/home/docweb/-/docweb-display/docweb/9512278

individuare l'elemento concretamente leso a seguito della creazione di un *deepfake*, ovverosia l'identità personale. Infatti, tale prodotto dell'intelligenza artificiale non si limita a generare un contenuto falso, ma arriva ad intaccare ciò che consente a un individuo di riconoscersi e distinguersi nel contesto sociale.

In definitiva, arrivare a comprendere cosa è un *deepfake* in senso lato rende imprescindibile analizzare la sfera dell'identità in sé, ma non solo: per il presente scritto, sarà sufficiente porre attenzione alle nozioni richiamate indirettamente nella definizione del Garante, ossia immagine e voce. La ragione di tale limitazione risiede nella natura del *deepfake*, che non falsifica ogni aspetto dell'identità[4], ma partendo da volto e voce si limita a manipolare fotografie, video o *file* audio, mirando così alla compromissione di quelle componenti che consentono «l'individuazione inequivoca della persona»[5].

1.1. Identità

Con il termine «identità» si fa riferimento a un concetto il quale vive nel mondo della filosofia da ormai alcuni secoli, ma che solo dal Novecento, assumendo significato politico, è entrato nel linguaggio comune e in vari contesti come i giornali, i social media o la televisione[6]. Va detto sin d'ora che parlare di identità non è missione affatto semplice, avendo tale espressione una notevole quantità declinazioni ed essendo presente in moltissimi universi, da quello della matematica alla filosofia, o ancora da quello antropologico a quello della psicologia.

Ciononostante, la distinzione fra identità personale e sociale può essere un tentativo utile di interpretazione sul punto. La prima costituisce «la riflessività

[4] In effetti, immagine e voce sono particolarmente rilevanti per l'analisi dei *deepfakes*, ma non possono elevarsi a uniche componenti costitutive dell'identità di un individuo, dovendosi fare menzione anche di elementi quali il nome o i codici attribuiti dalle pubbliche amministrazioni (come il codice fiscale).

[5] G. Resta, *Identità personale e identità digitale*, in Diritto dell'Informazione e dell'Informatica, n. 3, 2007, pp. 511 ss.

[6] R. Petri, *Identità: parabole politiche di un concetto*, in Memoria e Ricerca, n. 1, 2023, p. 7.

del Sé»[7], ossia la capacità di pensare a sé stessi come un'entità distinta dalle altre: è con la celebre espressione «*cogito ergo sum*» di Cartesio che, secondo molti, nasce l'idea di soggettività[8]. Da tale linea di pensiero sono emerse in seguito molteplici declinazioni: si pensi alle riflessioni seicentesche di John Locke, per il quale l'identità personale non si sostanzia nel corpo, ma nel fatto che in un individuo permanga la coscienza di sé e che questa sia continua nel tempo[9]. Ebbene, tale concezione verrà successivamente mutata prima da autori come Heidegger, secondo cui invece l'identità personale non può considerarsi statica e data una volta per tutte, ma deriva dal rimanere fedeli o meno alla propria autenticità, scelta che si concretizza nell'accettazione o nel rifiuto del mondo (comprensivo di vita quotidiana, tradizioni, abitudini) in cui si vive[10]; e poi, in tempi più recenti, da visioni come quella di Erikson, per il quale in ogni fase di vita dell'individuo subentra una "crisi di identità" che deve essere risolta per passare allo stadio successivo. Tale crisi, aggiunge lo psicologo tedesco, si accentua fortemente nell'età adolescenziale, stadio in cui l'uomo assiste a numerosi cambiamenti dal punto di vista fisico, sociale e cognitivo, che lo spingono a chiedersi quale sia il senso della propria esistenza[11]. Emerge dunque un concetto dinamico di identità personale, cioè in continua trasformazione e che mai si cristallizza.

Tra gli anni Settanta e Ottanta del secolo scorso si assiste a una nuova considerazione: vi è una parte dell'identità di ognuno che nasce dal senso di appartenenza a uno o più gruppi, e dal valore emotivo che ne consegue. È la teoria

[7] Ivi, p. 9

[8] I. Sibella, *Identità narrativa: racconto o significato?*, Psicologia Fenomenica, 2019. Reperibile al sito: https://www.psicologiafenomenologica.it/identita-narrativa-racconto-significato/

[9] J. Locke, *An Essay Concerning Human Understanding* (1690), in *Philosophical Works,* vol. 1, Henry G. Bohn, London, 1854, pp. 448, 466

[10] M. Heidegger, *Sein und Zeit*, Max Neimeyer, Tübingen, 1967, pp. 20, 175-180

[11] E. H, Erikson, *Young Man Luther: A Study in Psychoanalysis and History*, Norton & Company, New York, 1993, pp. 13, 18, 21-22, 282

della c.d. "identità sociale"[12], che non si eleva a concetto di identità assoluto, ma che va semplicemente a porsi quale uno dei tre livelli di "categorizzazione" del sé[13], assieme all'identità personale e all'identità "umana", (che guarda al sé come essere umano)[14]. Ciascuna di queste tre tipologie di identità, poi, assumerà carattere preminente o meno rispetto alle altre a seconda delle differenti circostanze in cui si svolge il vivere umano. Così, ad esempio, nelle relazioni dirette (come tra marito e moglie) l'identità personale assumerà maggior rilevanza, mentre nel caso di un confronto fra gruppi (si guardi a due squadre di calcio), la dimensione sociale dell'identità sarà prevalente[15].

Dall'originaria elaborazione di identità sociale sono poi emerse, come avvenuto per quella personale, ulteriori ramificazioni teoriche. Una delle più note sostiene che per un individuo rilevi non solo la mera appartenenza a più gruppi: nel dettaglio, dai rapporti instaurati con ciascuno di questi, scaturiscono differenti percezioni per il singolo le quali sono frutto di un processo di sovrapposizione (quanti membri di un gruppo appartengono anche agli altri) e similarità (quanto si somigliano i membri di gruppi distinti)[16]. Combinando variamente sovrapposizione e similarità, in base alle circostanze del caso concreto, l'interessato svilupperà diversi modi di concepire la propria identità sociale[17].

---

[12] H. Tajfel, J. C. Turner, *An integrative theory of intergroup conflict*, in W. G. Austin, S. Worchel, *The social psychology of intergroup relations*, Brooks/Cole Publishing Company, California, 1979, pp. 33-47

[13] Cfr. J. C. Turner, *Rediscovering the Social Group: A Self-Categorization Theory*, Basil Blackwell, Oxford, 1987

[14] M. Rubini, E. Crocetti, *Una o più identità? Articolando identità personale e sociale*, in Psicologia sociale, n. 1, 2018, p. 31

[15] Ivi, p. 35

[16] S. Roccas, M. B. Brewer, *Social identity complexity*, in Personality and Social Psychology Review, n. 6, 2002, pp. 88-106

[17] Nel dettaglio, si possono ottenere quattro configurazioni di identità sociale, che assumono un grado crescente di complessità: intersezione (l'individuo riflette sé in chi è parte di tutti i suoi gruppi di appartenenza); dominanza (si considera un gruppo come il principale, mentre gli altri assumono ruolo secondario); compartimentalizzazione (il singolo pone le diverse identità sociali allo stesso piano,

Come accennato la differenziazione fra personale e sociale, per quanto significativa, non può esaurire la complessità del concetto di identità. Pertanto, risulta opportuno tener brevemente conto anche di alcune delle prospettive che ad essa sono state accostate nel tempo.

Da segnalare, anzitutto, è l'influenza dell'identità "narrativa". Teorizzata da Paul Ricoeur, essa non è altro che la narrazione della propria esistenza da parte di un individuo, frutto della interazione fra "idem" (ossia l'identità oggettiva e statica, data dagli elementi che restano invariati come la memoria di sé o il nome) e "ipse" (l'identità dinamica, il sé che si evolve tramite decisioni etiche ed esistenziali)[18]. In altri termini, identità personale e alterità si intrecciano e la prima non può considerarsi separata dalla seconda: tale visione è perfettamente rappresentata dal titolo dell'opera di Ricoeur "Sé come un altro"[19].

Ulteriore tipologia, che mostra ancora la costante interazione fra identità personale e sociale, è quella dell'identità "culturale", per cui le culture (intese come quelle familiari, di ceto, di status) contribuiscono all'edificazione del Sé attraverso elementi come le tradizioni, la religione, i valori, la lingua[20]. La conseguenza, come sostenuto da alcuni autori, è che l'identità diviene così "finta", ossia risultato di una costruzione umana, la quale è tuttavia «operazione irrinunciabile»: l'individuo ha bisogno di ricercare sé in «forme culturali che garantiscono un'identità»[21]. Non manca, inoltre, chi ritiene che tale forma di identità (come quella personale) sia ormai in crisi, poiché a causa del fenomeno

---

attivandone una o alcune a seconda dei diversi contesti); fusione (le differenti identità si integrano, e l'individuo riconosce come parte del suo gruppo i soggetti che ne hanno almeno una in comune). Per approfondimenti, cfr. M. Rubini, E. Crocetti, *Una o più identità? Articolando identità personale e sociale*, in Psicologia sociale, n. 1, 2018, p. 33-34

[18] F. Ciotti, C. Morabito, *La narrazione come incontro*, Firenze University Press, Firenze, 2022, p. 130

[19] P. Ricoeur, *Sé come un altro* (a cura di D. Iannotta), Jaca Book, Milano, 2005, p. 78. L'autore, nell'illustrare il senso del titolo, ha cura di precisare che il "come" non assume funzione comparativa, ma di implicazione, rendendo la connessione tra "sé" e "altro" ancora più evidente.

[20] P. Del Core, *Rapporto tra identità e cultura, una coordinata essenziale per «comprendere l'altro» nella prospettiva dell'intercultura*, in EDUCA, n. 2, 2016, p. 122

[21] F. Remotti, *Contro l'identità*, Editori Laterza, Roma-Bari, 2001, p. 97

della globalizzazione si è assistito a una crescente omologazione delle culture. La soluzione potrebbe essere la ricerca della già citata alterità, tramite la quale le differenti culture si aprirebbero al dialogo con le altre, ma senza cancellarsi e ripristinando un'identità dinamica[22].

Un breve accento, poi, può porsi per l'identità "di genere": oggetto di dibattiti fortemente attuali, con essa si intende il senso di aderenza di un individuo a un genere in cui esso si identifica[23]. La nozione di identità di genere, di impostazione relativamente recente, ha scardinato l'antica credenza per cui "sesso" e "genere" fossero considerati sinonimi: mentre il primo è dato dagli elementi genetici e anatomici, dunque oggettivi, il secondo costituisce la dimensione «psicologica e culturale associata al sesso biologico»[24].

Sebbene non manchino altre classificazioni meritevoli di considerazione, per collegarsi al tema dei *deepfakes* ci si deve infine focalizzare sull'identità "digitale": l'ordinamento italiano, nel Codice dell'Amministrazione Digitale (CAD), la definisce come «la rappresentazione informatica della corrispondenza tra un utente e i suoi attributi identificativi, verificata attraverso l'insieme dei dati raccolti e registrati in forma digitale»[25]; ancora, il Regolamento eIDAS, che pone una normativa a livello comunitario per l'identificazione elettronica e i servizi digitali, parla di «processo per cui si fa uso di dati di identificazione personale in forma elettronica che rappresentano un'unica persona fisica o giuridica»[26].

Si comprende che l'identità digitale, più che una formulazione meramente teorica, è uno strumento concreto che permette sia una identificazione univoca

---

[22] A. Perotti, *La via obbligata dell'interculturalità*, Editrice Missionaria Italiana, Città di Castello (Perugia), 1994, p. 18

[23] K. Cooper, L.G.E. Smith, A.J. Russell, *Gender Identity in Autism: Sex Differences in Social Affiliation with Gender Groups*, in Journal of Autism and Developmental Disorders, vol. 48, 2018, pp. 3995-4006

[24] T. Schechner, *Gender identity disorder: a literature review from a developmental perspective*, in Israel Journal of Psychiatry and Related Sciences, vol. 47 (2), 2010, pp. 132-138

[25] Codice dell'Amministrazione Digitale (D.lgs. n. 82/2005), art. 1, lett. u-quater)

[26] Regolamento (UE) n. 910/2014, art. 3, n. 1)

del soggetto interessato, sia l'accesso di quest'ultimo ai servizi digitali che richiedano l'autenticazione, come avviene per le pubbliche amministrazioni. Addirittura, la Strategia Italia Digitale per il 2026 prevede che ogni cittadino italiano dovrà disporre di un'identità digitale con cui «accedere a tutti i servizi digitali pubblici e privati in Italia e in Europa»[27]. Ad oggi, gli strumenti di identità digitale nel nostro ordinamento sono tre: lo SPID (Sistema Pubblico di Identità Digitale), che permette ai cittadini l'accesso ai servizi digitali delle pubbliche amministrazioni e dei privati aderenti[28] mediante una richiesta gratuita presso provider qualificati (come Poste Italiane o Aruba); la CIE (Carta di Identità Elettronica), emessa dal Ministero dell'Interno ed evoluzione della carta d'identità cartacea[29], che accerta l'identità del titolare e permette la fruizione di servizi in Italia e in Europa grazie al chip interno e alle c.d. "credenziali CIE" di cui è dotata[30]; e la CNS (Carta Nazionale dei Servizi), valida per usufruire dei servizi delle pubbliche amministrazioni ulteriori a quelli per cui è sufficiente lo SPID[31], e che a differenza di quest'ultimo costituisce anche un documento fisico di identificazione[32].

[27] Agenzia per l'Italia Digitale (AGID), *Identità digitale*, 2025. Reperibile al sito: https://www.agid.gov.it/it/ambiti-intervento/identita-digitale

[28] Agenzia per l'Italia Digitale (AGID), *SPID - Sistema Pubblico di Identità Digitale*, 2025. Reperibile al sito: https://www.agid.gov.it/it/piattaforme/spid

[29] L'utilizzazione della CIE, in tutta l'Unione Europea, diverrà in pochi anni esclusiva. Il regolamento UE 2019/1157, al suo art. 5, dispone che ogni supporto cartaceo si considererà non valido a partire dal 3 agosto 2031. Nel frattempo, non mancano misure gradualmente restrittive verso le tradizionali carte d'identità: si pensi che già dal prossimo 3 agosto 2026 queste ultime perderanno la loro efficacia per l'espatrio (art. 5, par. 2, lett. a).

[30] Ministero dell'Interno, *Carta di Identità Elettronica*, 2025. Reperibile al sito: https://www.cartaidentita.interno.gov.it/la-carta/

[31] Nel dettaglio, la CNS è particolarmente usata in ambito sanitario: in molte regioni italiane, ad esempio, la Tessera Sanitaria può costituire CNS grazie al chip di cui è dotata, sebbene a tale scopo sia necessario espletare un'apposita procedura di attivazione della TS.

[32] Agenzia per l'Italia digitale, *Carta Nazionale dei Servizi*, 2024. Reperibile al sito: https://www.agid.gov.it/it/piattaforme/carta-nazionale-servizi

Dall'analisi di tali strumenti emerge una importante differenza con l'identità fisica. Quest'ultima, nonostante vi siano più strumenti per certificarla come passaporto o patente, è composta da elementi stabili e invariabili (nome o data di nascita). Al contrario, l'identità digitale può assumere molteplici forme contemporaneamente[33], ciascuna per ogni piattaforma che si utilizza. Per fare un esempio, l'utente che gestisce più account, dai *social network* ai servizi di posta elettronica, può accedere in ciascuno di questi con credenziali differenti, generando dunque più profili ricondotti alla stessa persona. Oltre a questo i dati immessi dall'utente, per scelta sua o della piattaforma, possono essere in un secondo momento aggiunti, cancellati o modificati. Tale ipotesi riflette il carattere dinamico dei dati in un'identità digitale, a differenza di quanto avviene per quella fisica.

Affinchè l'identità digitale possa considerarsi sicura e affidabile, è necessario verificare che l'utente sia la persona che dichiara di essere: questo avviene con il processo di autenticazione[34], che si articola in tre principali fasi. Anzitutto l'interessato fornisce informazioni personali a un ente autorizzato, il quale verifica l'identità dell'utente tramite documenti tradizionali (come la carta di identità) o con riconoscimento biometrico; a seguito della verifica dell'autenticità di questi dati, l'ente assegna un profilo digitale con credenziali univoche; grazie a queste ultime, infine, l'utente potrà accedere in modo sicuro ai vari servizi online[35]. Espletata questa procedura, occorre poi specificare che esistono diversi strumenti per l'autenticazione: qualcosa che l'utente conosce (come la password), più comune ma anche più semplice da rubare o decifrare; qualcosa che l'utente possiede (come una smart card), che tuttavia deve sempre essere a portata di mano dell'utente e che può essere soggetta a furto; e qualcosa

[33] G. Dragoni, *Identità Digitale: cos'è, come funziona e come si usa*, Osservatori.net, 2025. Reperibile al sito: https://www.osservatori.net/blog/digital-identity/identita-digitale-significato-funzionamento/

[34] M. Kosinski, *Cos'è l'autenticazione?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/it-it/think/topics/authentication

[35] *Identità Digitale (Digital Identity): Cos'è, Come Funziona e gli Strumenti in Italia*, Workinvoice, 2024. Reperibile al sito: https://www.workinvoice.it/identita-digitale

che l'utente è: si parla dei metodi di autenticazione biometrica come le impronte digitali e il riconoscimento vocale[36].

In conclusione, si può asserire che anche l'identità digitale concorre alla formazione del sé, ma non senza pericoli per la reputazione dell'utente. È qui che si può finalmente fare cenno ai *deepfakes*, che proprio nel furto di identità[37], nella frode, nella minaccia e nella disinformazione trovano terreno fertile[38]. Se con l'intelligenza artificiale si può clonare un timbro vocale con precisione o sostituire l'aspetto di una persona in un video, si comprende bene la semplicità nell'eludere i sistemi di autenticazione. Il paradosso è che i più volte richiamati dati biometrici, da una parte percepiti come comodi e sicuri dalla community di internet, dall'altra suscitano preoccupazione sempre maggiore tra gli esperti, che invece spingono per una regolamentazione più severa su come tali dati sono raccolti e utilizzati[39]. Ecco comparire nuovamente la necessità di andare oltre l'identità e considerare quegli elementi identitari, come l'immagine e la voce, suscettibili di lesione dai *deepfakes*. Saranno questi temi, infatti, ad interessare i paragrafi a seguire.

---

[36] F. B. Schneider, *Authentication of People*, Cornell University, 2005. Reperibile al sito: https://www.cs.cornell.edu/courses/cs513/2005fa/NNLauthPeople.html

[37] Esemplificativo è quanto successo a Hong Kong nel 2024, dove una multinazionale britannica è stata truffata per 25 milioni di dollari. Nel dettaglio, i criminali hanno indotto un dipendente ad autorizzare tale transazione tramite una videochiamata con persone che credeva fossero i dirigenti, ma che in realtà erano delle ricostruzioni *deepfake*. Cfr. K. Magramo, *British engineering giant Arup revealed as $25 million deepfake scam victim*, CNN, 2024. Reperibile al sito: https://edition.cnn.com/2024/05/16/tech/arup-*deepfake*-scam-loss-hong-kong-intl-hnk/index.html

[38] *Deepfake Deception in Digital Identity*, Identity Management Institute. Reperibile al sito: https://identitymanagementinstitute.org/*deepfake*-deception-in-digital-identity

[39] S. He, Y. Lei, Z. Zhang, Y. Sun, S. Li, C. Zang, J. Ye, *Identity Deepfake Threats to Biometric Authentication Systems: Public and Expert Perspectives*, arXiv, 2025, p. 3

1.2. Immagine

Così come visto per l'identità, il concetto di immagine non può ritenersi assoluto, né scevro da numerose classificazioni e ambiti di applicazione. Tuttavia, anche in questo caso è possibile individuare un punto di partenza, dato dal significato etimologico del termine: immagine deriva da *imago-gĭnis*, che sembra rinviare alla sola "figura", intesa come la superficie delle cose, la forma esteriore di oggetti e soggetti[40]. In realtà, l'espressione latina rimanda anche al pensiero, allo spettro, all'ombra, all'idea, mostrando di possedere una ricchezza semantica che va ben oltre la mera dimensione visiva.

In effetti l'immagine non è realtà, ma uno strumento che la incornicia, costituendo un processo di imitazione che offre solo una delle sue possibili versioni. Conseguenza è che l'immagine non solo risulta esposta alla manipolazione di chi le genera, ma anche alla reinterpretazione soggettiva di chi la osserva: si può addirittura ritenere che «tutte le immagini costituiscono mondi»[41], sfuggendo a ogni tentativo di attribuire loro un senso oggettivo.

Sulla base di tali premesse è possibile individuare una delle numerose declinazioni dell'immagine, in particolare quella antropologica: tra i suoi fautori vi è Hans Belting, che vede nel processo di *picture-making* la costante interazione di tre elementi. Anzitutto, l'immagine «radicata nella nostra mente», che grazie al pensiero, al sogno, la memoria o l'immaginazione influenza il giudizio di un soggetto nei riguardi delle immagini esterne, determinandone la conseguente attrazione o repulsione[42]. In secondo luogo vi è il corpo vivente, che produce l'immagine o vi interagisce[43]. Vi è infine il *medium* il quale, nel rapporto tra

[40] R. Siconolfi, *La magia dell'immagine*, Adv Media Lab, 2024. Reperibile al sito: https://www.advmedialab.com/la-magia-dell-immagine/#toc_La_magia_dellimmagine_definizione_oltre_la_superficie

[41] A. Testa, *Le vie del senso*, Garzanti, Milano, 2021, p. 116

[42] D. Agrillo, E. Amedeo, A. Di Nobile, C. Tarallo, *L'immagine nel mondo, il mondo nell'immagine: nuove prospettive per un approccio pluridisciplinare alla rappresentazione testuale ed extra-testuale*, Quaderni della ricerca – 1, UniorPress, Napoli, 2016, p. 12

[43] E. Macaluso, *Che cosa sono le immagini? Riflessioni su una questione radicale*, in Rivista interdisciplinare di comunicazione, n. 3, 2021, p. 204

corpo e immagine, non può considerarsi un mero supporto (come la pietra o la fotografia), assumendo significato solo tramite lo sguardo dell'osservatore e che «solo se viene animato» da quest'ultimo «diventa un'immagine»[44].

L'analisi condotta non solo rimarca la soggettività innata dell'immagine, ma permette anche di riconoscerne ulteriori forme: oltre quelle di cui si è appena fatto cenno, come l'immagine mentale (la rappresentazione interna) o quella visiva (data da ciò che appare all'esterno, come una fotografia o un dipinto), può assumere rilievo accennare anche all'immagine sociale.

L'immagine sociale rappresenta il modo in cui il soggetto è o vorrebbe essere percepito dagli altri, mutando in tutto o in parte a seconda degli interlocutori o delle circostanze[45]. Se si pensa che l'individuo appartiene a più gruppi sociali, e che il conseguente senso di appartenenza è influenzato dall'idea che gli altri si fanno di esso, la connessione con la già trattata identità sociale (v. *supra*, 1.1.) è evidente. La differenza fra identità e immagine sociale, tuttavia, è che mentre la prima è frutto di una percezione interna, di come il soggetto riconosce sé all'interno di quei gruppi, la seconda dipende da una percezione esterna, cioè come si appare agli altri.

Non si può analizzare pienamente l'immagine sociale senza menzionare la c.d. *impression management* (gestione dell'impressione): teorizzata da Erving Goffman[46], essa indica il processo in cui ciascuna persona cerca di controllare e condizionare l'impressione di sé che lascia negli altri[47]. Il sociologo canadese ricorre alla metafora della maschera, di cui si dota ognuno per adattare di volta in volta il proprio «sé esteriore», combinando alla "performance" costruita per

---

[44] H. Belting, *Antropologia delle immagini*, Carocci editore, Roma, 2011, p. 42

[45] D. Usera, C*ommunicating to Connect: Interpersonal Communication for Today*, LibreTexts, 2025. Reperibile al sito: https://socialsci.libretexts.org/Bookshelves/Communication/Introduction_to_Communication/Communicating_to_Connect_Interpersonal_Communication_for_Today_(Usera)/03%3A_Identity/3.02%3A_Self-Image

[46] Cfr. E. Goffman, *La vita quotidiana come rappresentazione* (trad. it. di M. Ciacci), Il Mulino, Bologna, 1997, opera da cui è tratta la celebre espressione

[47] M. Vigarani, *"Giù la maschera": fenomenologia e rischi del Self Impression Management nel processo di Recruitment*, Meliusform, 2020. Reperibile al sito: https://www.meliusform.it/giu-la-maschera-fenomenologia-e-rischi-del-self-impression-management-nel-processo-di-recruitment.html

gestire le impressioni l'occultamento di quegli aspetti del sé che andrebbero a stonare con le richieste della specifica situazione[48]. Si immagini il candidato a un colloquio di lavoro: è nel suo interesse offrire un'immagine presentabile di sé, ad esempio vestendosi in maniera formale, mantenendo un atteggiamento calmo e dialogando con sicurezza. Allo stesso momento, dovrà però nascondere l'agitazione, la paura di non essere assunto o tratti della sua personalità (si pensi all'introversione) ritenuti convenzionalmente inidonei a un simile contesto[49].

In coerenza con lo schema del paragrafo precedente, si può concludere il discorso sull'immagine introducendo la nozione di immagine digitale. Differente dalla storica immagine analogica, catturata e conservata su un supporto fisico tramite un processo chimico, come avviene per le tradizionali fotografie, quella digitale è composta da un insieme di puntini disposti su una griglia bidimensionale che prendono il nome di pixel (acronimo di Picture Elements). Tali nuove rappresentazioni sono generate da dispositivi elettronici quali fotocamere digitali o scanner, per essere poi visualizzate su schermi o supporti altrettanto digitali come computer o smartphone[50].

La nascita dell'immagine digitale segna un definitivo punto di rottura col passato: si passa dall'analogico, il quale costituiva traccia fisica della realtà, a un'immagine prodotta da un algoritmo, un insieme di numeri modificabili all'infinito[51]. Forti sono dunque le preoccupazioni di autori che, paragonando le immagini informatizzate a «dipinti o disegni, piuttosto che a delle fotografie»[52], vedono nell'immagine informatizzata uno strumento di manipolazione della

---

[48] J. Solomon, A. Solomon, N. Joseph, S. Norton, *Impression management, myth creation and fabrication in private social and environmental reporting: Insights from Erving Goffman*, in Accounting, Organizations and Society, vol. 38, n. 3, 2013, pp. 195-213

[49] C. Genova, *Il cerchio nello spazio. Ipotesi e strumenti per un'analisi della ri–significazione dei luoghi*, in Lexia. Rivista di semiotica 9-10, 2011, p. 193-209

[50] *Differenze tra immagini digitali e analogiche: una guida completa*, Giornalettismo, 2025. Reperibile al sito: https://www.giornalettismo.com/immagine-digitale-e-immagine-analogica-differenze/

[51] R. Debray, *Vita e morte dell'immagine* (trad. it. A. Pinotti), Editrice Il Castoro, Milano, 1999, pp. 230-231

[52] S. Pezzano, *L'immagine digitale. Una vera-falsa "nuova immagine"*, in Leitmotiv, n. 4, 2004, p. 86

realtà, e proprio in questo delicato tema si osserva il filo rosso che lega immagine e *deepfakes*. Invero, le immagini digitali possono essere duplicate o manipolate con estrema semplicità, ed è assai difficile se non impossibile esaminare il *file* per stabilire se vi siano state tracce di interventi, a differenza di ciò che accade con un classico negativo[53]. Se si aggiunge poi che la manipolazione di tali immagini è possibile anche senza acquisire particolari competenze, il quadro si fa decisamente più complesso[54].

Con la costante proliferazione di immagini digitali, dalla pubblicità ai video online alla televisione, il rischio per la società è quello di una «confusione tra reale e immaginario», con gli utenti che potrebbero accontentarsi di vivere solo attraverso tali immagini, senza la necessità di ricorrere alle esperienze dirette[55]. Tale pericolo è da considerarsi ormai concreto, con l'eccezione dei prodotti dell'intelligenza artificiale: quest'ultima, notevolmente più recente, rende (almeno per ora) sufficientemente palese la sua natura, ma essendo una tecnologia in crescita esponenziale e che assume tratti sempre più realistici, la minaccia di una «fusione delle immagini stesse con l'ambiente reale che ci circonda fino a prenderne progressivamente il posto»[56] aleggia incessantemente.

Da tali riflessioni nasce la necessità di sviluppare una «grammatica dell'immagine digitale»[57] che introduca principi e regole per una sua utilizzazione critica e consapevole, onde evitare un consumo incontrollato di immagini e privo di senso etico.

---

[53] Ivi, p. 83

[54] *Science & Tech Spotlight: Deepfakes*, U.S. Government Accountability Office, 2020. Reperibile al sito: https://www.gao.gov/products/gao-20-379sp

[55] S. Pezzano, *L'immagine digitale. Una vera-falsa "nuova immagine"*, in Leitmotiv, n. 4, 2004, p. 88

[56] J.-J. Wunenburger, *Filosofia delle immagini* (trad. it. S. Arecco), Einaudi, Torino, 1999, pp. 362-363

[57] V. Codeluppi, *Il ruolo sociale delle immagini digitali*, in Mediascapes Journal, n. 19, 2022, p. 226

1.3. Voce

Dopo aver esaminato i concetti di identità e immagine è opportuno concludere il presente paragrafo richiamando la nozione di voce: difatti, come indicato nella definizione del Garante per la protezione dei dati personali (v. *supra*, 1.), essa contribuisce alla creazione dell'identità di un soggetto, e in quanto tale costituisce uno di quegli elementi che possono essere lesi dalla creazione di un *deepfake*.

Si può definire la voce come il timbro naturale e distintivo emesso da ogni individuo che ricorre al linguaggio o, in senso più generico, un qualunque suono prodotto dalla vibrazione delle corde vocali[58]. Deve inoltre essere pensata come un fenomeno complesso, in quanto concorrono diversi elementi alla sua produzione. Vi è infatti chi distingue fra caratteristiche qualitative e dinamiche della voce[59]: le prime, come la tensione muscolare o la nasalità, trovano origine dagli aspetti fisici e anatomici di ciascuno, essendo in quanto tali innate e non scelte consapevolmente; al contrario le seconde, quali l'intonazione o il ritmo, trovano sviluppo durante il percorso di vita e si modellano in base alle varie interazioni sociali, potendo essere controllate consapevolmente da chi dialoga[60].

In sintesi, mentre le componenti qualitative hanno natura stabile ed extralinguistica, quelle dinamiche sono a breve o medio termine e dipendono da informazioni linguistiche, come la differenza tra domande e affermazioni, o paralinguistiche, come uno stato emotivo momentaneo[61]. Tuttavia una simile differenziazione, per quanto efficace sul piano teorico, potrebbe sfumare nel momento in cui l'interessato decide di controllare, seppur temporaneamente,

---

[58] R. L. Trask, *A dictionary of phonetics and phonology*, Routledge, London, 1996, pp. 378-379

[59] D. Abercrombie, *Elements of general phonetics*, Edinburgh University Press, Edinburgh, 1967

[60] Cfr. P. Ladefoged, *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech*, Oxford University Press, London, 1967, p. 61. L'autore propone una suddivisione analoga a quella appena citata, distinguendo fra caratteristiche "organiche", assimilate alle qualitative; e "funzionali", corrispondenti a quelle dinamiche.

[61] B. Paye Herrero, *Voice and Identity: A contrastive study of identity perception in voice*, Universitat de València, Servei de Publicacions, Valencia, 2009, p. 20

aspetti inerenti le caratteristiche qualitative: si pensi a chi sussurra appositamente all'amico che non ama i rumori forti, o nel caso di uno spot televisivo per promuovere la vendita di un certo prodotto[62].

Modellare la voce a seconda delle circostanze contribuisce alla genesi di quella "maschera" che determina il sé esteriore cui si è già fatto cenno (v. *supra*, 1.2.). Da tali esempi, dunque, risulta impossibile ignorare il forte nesso che lega identità, immagine e voce. Tuttavia, se quest'ultima rappresenta uno dei tasselli che definiscono l'identità di un individuo, è anche vero che non agisce in isolamento, necessitando di interazioni faccia a faccia accompagnate da gesti o espressioni facciali. E se si considera che una delle sfere dell'identità, nel dettaglio quella culturale (di cui già discusso, v. *supra*, 1.1.), trova fondamento anche nel linguaggio, è facile intuire come anche la cultura abbia influenza in un tale contesto[63].

Le informazioni che si ricavano dalla voce, nelle conversazioni quotidiane, coordinano istintivamente l'interlocutore nel decidere «chi ci piace, di chi fidarci e con chi fare affari»[64]. Ma quali indicazioni può fornire il solo dialogo sull'identità del singolo? I dettagli maggiormente intuibili sono quelli di tipo fisico[65]: numerosi sono stati gli studi che hanno osservato un nesso tra profondità della voce e robusta corporatura[66], fra l'avanzamento dell'età e caratteristiche vocali specifiche come la ruvidità e una minore qualità armonica[67],

---

[62] Cfr. M.M. Saz Rubio, B. Pennock-Speck, *Male and Female Stereotypes in Spanish and British Commercials*, in M. Muñoz-Calvo, C. Bueso-Gómez, M. Á. Ruiz-Moneva, *New Trends in Translation and Cultural Identity*, pp. 377-378

[63] B. Paye Herrero, *Voice and Identity: A contrastive study of identity perception in voice*, Universitat de València, Servei de Publicacions, Valencia, 2009, pp. 29-30

[64] C. Nass, *Machine voices*, PBS, 2005. Reperibile al sito: https://www.pbs.org/speak/ahead/technology/voiceinterface/#

[65] Cfr. J. Laver, *The Gift of Speech: Papers in the Analysis of Speech and Voice*, Edinburgh University Press, Edinburgh, 1991

[66] J. Fay, W.C. Middleton, *Judgement of Kretschmerian Body Types on the Voice Transmitted over a Public Address System*, in Journal of Social Psychology, n. 12, 1940, pp. 151-162

[67] Cfr. R. Luschinger, G.E. Arnold, *Voice, Speech, Language: Clinical Communicology: Its Physiology and Pathology*, Wadsworth Publishing, Belmont, 1965

o ancora fra l'alterazione vocale e lo stato di salute, come una tonalità nasale tipica di un raffreddore[68]. Altra sfera di rilevanza è poi quella sociale: si utilizza la voce per esplicitare l'appartenenza a un determinato gruppo[69]; per la pratica della cortesia (si pensi alla critica posta con intonazione appropriata affinchè non risulti aggressiva), indispensabile ai fini di una convivenza pacifica in comunità[70]; o per acquisire autorevolezza con gli interlocutori, ad esempio aumentando l'intensità vocale[71]. Infine, da quasi un secolo si è giunti, con metodo scientifico, alla conclusione che tramite la voce possano trapelare informazioni anche di natura psicologica[72]: tra i molti studi sul punto, si pensi a quelli secondo cui tonalità vocali deboli e acute indicano sottomissione o introversione[73] o quelli per cui una caduta di tono alta o bassa determina coinvolgimento attivo o disinteresse[74].

Questa digressione serve a comprendere quanto la voce possa comunicare sull'identità di un soggetto, al punto che è stata definita «faccia uditiva»[75]. La forte connessione fra voce e identità fa sorgere il timore che una manipolazione della prima comprometta gravemente anche la seconda, rischio reso oggi concreto proprio a causa dei *deepfakes*.

---

[68] B. Paye Herrero, *Voice and Identity: A contrastive study of identity perception in voice*, Universitat de València, Servei de Publicacions, Valencia, 2009, pp. 33-34

[69] Cfr. N. Mendoza-Denton, *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*, Blackwell, Malden, Mass., 2008. Nell'opera citata l'autrice delinea come le ragazze delle gang latine ricorrano a un'utilizzazione peculiare della loro voce per definire la propria identità.

[70] B. Paye Herrero, *Voice and Identity: A contrastive study of identity perception in voice*, Universitat de València, Servei de Publicacions, Valencia, 2009, pp. 34-35

[71] Cfr. T. van Leeuwen, *Speech, music, sound*, Macmillan, Basingstoke, 1999

[72] G.W. Allport, H. Cantril, *Judging Personality from Voice*, in Journal of Social Psychology, n. 5, 1934, p. 50

[73] H. Eckert, J. Laver, *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*, Beltz/Psychologie Verlags Union, Weinheim, 1994, p. 7

[74] A. Cruttenden, *Intonation*, Cambridge University Press, Cambridge, 1997, p. 91

[75] A. Schirmer, *Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing*, in Social Cognitive and Affective Neuroscience, n. 13, 2018, pp. 1-13

La creazione di questi ultimi solleva non pochi dubbi morali poiché essi interrompono il processo di costruzione sociale dell'identità delle loro vittime, utilizzando la loro immagine o voce per generare rappresentazioni su cui non possono avere voce in capitolo e che comunque mai approverebbero[76]. Riprendendo quanto detto in sede di immagine, la semplicità di accesso ai sistemi di IA generativa, senza particolari competenze né costi, si riflette anche nel caso dei *deepfakes* vocali: un semplice video, post o registrazione dell'interessato è più che sufficiente per far sì che l'algoritmo catturi la sua voce e la riutilizzi secondo le istruzioni date dal *prompt*[77].

Un'altra preoccupazione verte sulla credibilità delle registrazioni vocali. L'ampia diffusione dei *deepfakes* vocali, in effetti, può consentire ai criminali di mentire sulla veridicità di un'eventuale registrazione che li incrimina: vi sono autori che, nel descrivere tale fattispecie, hanno usato l'espressione "*liar's dividend*"[78]. La conseguenza è che, se si può sostenere tanto facilmente che un qualsiasi audio o video compromettente è stato manipolato, la registrazione potrebbe perdere l'originaria funzione di incontestabile mezzo di prova[79]. Ad ogni modo, al tema delle criticità sul piano probatorio dei *deepfakes* si dedicherà apposita trattazione più avanti.

In definitiva la voce, così come l'immagine di una persona, è parte del suo essere e disporne senza il suo consenso equivale ad appropriarsi del suo nome, cognome o volto. Emerge la nuova necessità di non assimilare, come erroneamente avviene, il «tecnologicamente possibile» all' «umanamente

---

[76] A. de Ruiter, *The Distinct Wrong of Deepfakes*, in Philosophy & Technology, n. 34, 2021, p. 1313

[77] È il caso di un preside di una scuola superiore del Maryland (Stati Uniti), la cui voce è stata estratta e riutilizzata per creare una falsa registrazione di stampo razzista. Si è dinanzi uno dei molti esempi di come un *deepfake* vocale può minare reputazione e status sociale di un individuo. Cfr. B. Finley, *Deepfake of principal's voice is the latest case of AI being used for harm*, AP News, 2024. Reperibile al sito: https://apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e

[78] D.K. Citron, R. Chesney, *Deep fakes: A looming challenge for privacy, democracy, and national security*, in California Law Review, n. 107, 2019, p. 1785

[79] Cfr. R. Rini, *Deepfakes and the Epistemic Backstop*, in Philosophers' Imprint, n. 20, 2020

sostenibile», con la speranza che la società sappia resistere sin d'ora alla tentazione di far «prevalere l'usabilità sull'umanità»[80].

2. Il diritto all'identità nelle carte fondamentali (CEDU, Carta di Nizza, Costituzione italiana)

Fino a questo momento sono state esaminate le nozioni di "identità", "immagine" e "voce". L'approccio proposto, tuttavia, è stato di stampo teorico e filosofico, sicché è ora necessario un inquadramento sul piano giuridico, a cominciare dal diritto all'identità.

Prima di osservare come le Carte fondamentali, compresa la Costituzione italiana, hanno riconosciuto questo concetto, risulta opportuno darne una definizione: una delle più autorevoli in Italia descrive il diritto all'identità personale «il diritto a che la proiezione sociale della propria personalità non subisca travisamenti o distorsioni a causa della attribuzione di idee, opinioni o comportamenti differenti da quelli che quell'individuo ha manifestato nella vita di relazione»[81]. Confrontando tale espressione con quella più volte richiamata del Garante per la protezione dei dati personali (v. *supra*, 1.), risulta chiara la necessità di includere il diritto all'identità personale in una trattazione che ha ad oggetto i *deepfakes*.

[80] G. Scorza, *Giù le mani da quelle voci: sono frammenti della nostra identità personale*, StartupItalia, 2025. Reperibile al sito: https://startupitalia.eu/tech/giu-le-mani-da-quelle-voci-sono-frammenti-della-nostra-identita-personale/

[81] G. Pino, *Il diritto all'identità personale. Interpretazione costituzionale e creatività giurisprudenziale*, Il Mulino, Bologna, 2003, p. 9

Il diritto in analisi è stato inizialmente frutto di una produzione giurisprudenziale e dottrinale; una «certificazione»[82] legislativa in Italia avverrà solo con l'art. 1 della Legge n. 675 del 31 dicembre 1996[83].

Analogo discorso può farsi per l'Unione Europea, ed è proprio da questo ambito che si intende partire: nell'ordinamento europeo, infatti, il diritto all'identità personale è stato frutto di elaborazione sia dalla Corte di Giustizia dell'Unione Europea (d'ora in poi CGUE), sia dalla Corte europea dei diritti dell'uomo (d'ora in poi Corte EDU). Di quest'ultima è il merito di aver esteso l'ambito di applicazione dell'art. 8 della Convenzione Europea dei Diritti dell'Uomo (CEDU), che tutela il diritto al rispetto della vita privata, includendovi anche il diritto all'identità personale come «parte integrante della vita dell'interessato»[84]. Detto orientamento è ravvisabile già dalla sentenza *Bughartz*[85] in cui il diritto all'identità personale, ritenuto essenziale per identificare l'individuo nella sfera sociale e professionale, trova per la prima volta tutela mediante combinato disposto degli articoli 8 e 14 della CEDU, riconoscendo dunque un legame tra tale diritto e il divieto di discriminazione. È soltanto con la sentenza *Stjerna*[86] che la Corte EDU riterrà sufficiente, per la tutela dell'identità personale, il solo art. 8[87].

La Convenzione EDU non è l'unica a rinviare, seppur indirettamente, al diritto all'identità personale: il contenuto del suo art. 8 è ripreso dall'art. 7 della più recente Carta dei diritti fondamentali dell'Unione Europea (o Carta di Nizza), rubricato «Rispetto della vita privata e della vita familiare». Poiché l'art. 52,

---

[82] Ivi, p. 10

[83] E.C. Raffiotta, *Appunti in materia di diritto all'identità personale*, Forum di Quaderni Costituzionali, 2010. Reperibile al sito: https://www.forumcostituzionale.it/wordpress/images/stories/pdf/documenti_forum/paper/0173_raffiotta.pdf

[84] S. Marino, *L'identità personale alla prova delle libertà di circolazione*, in Eurojus, n. 4, 2020, p. 173

[85] Corte EDU, sent. 22 febbraio 1994, ric. 16213/90, *Burghartz c. Svizzera*

[86] Corte EDU, sent. 25 novembre 1994, ric. 18131/91, *Stjerna c. Finlandia*

[87] S. Marino, *L'identità personale alla prova delle libertà di circolazione*, in Eurojus, n. 4, 2020, p. 174

paragrafo 3 della stessa Carta indica che i diritti in essa inclusi e corrispondenti a quelli della CEDU possiedono uguale significato e portata, deve concludersi che anche l'art. 7 CDFUE è comprensivo del diritto all'identità personale.

Ad ogni modo il diritto in esame, all'interno della presente Carta, trova un altro riferimento[88] nell'art. 8, il quale sancisce la protezione dei dati personali. Tale norma, divenuta il fulcro del Regolamento generale sulla protezione dei dati (GDPR), riconosce all'individuo il diritto a esercitare un controllo sui dati che lo riguardano, ed è noto che una compromissione di questi ultimi equivarrebbe a una lesione della propria identità, specialmente in ambito digitale. Considerando poi che alla base della protezione dei dati vi è il principio di dignità umana (art. 1 della CDFUE), si comprende come l'art. 8, attribuendo al titolare dei dati la sovranità sugli stessi, ne tuteli di riflesso la dignità e contribuisca a un sistema di protezione dell'immagine e identità della persona[89].

Nell'ordinamento italiano, deve rammentarsi che il diritto all'identità personale è sorto come frutto di una elaborazione giurisprudenziale (dagli anni Settanta[90]) e dottrinale (almeno dalla fine degli anni Quaranta[91]), sicché non può

[88] In realtà, numerose sono le disposizioni della Carta di Nizza che presentano un nesso, sia pur non esplicito, con il diritto all'identità personale. Fra queste gli articoli 3 (integrità della persona), 10 (libertà di pensiero, coscienza e religione) e 11 (libertà di espressione), posti a tutela di aspetti significativi per la costruzione del sé. In questa sede saranno analizzati specificamente i soli articoli 1, 7 e 8, in quanto caratterizzati da un legame significativo col tema dell'identità personale.

[89] S. Lissens, *The Foundations of EU personal data protection law: Privacy and Human Dignity*, EU Renew, 2024. Reperibile al sito: https://eu-renew.eu/the-foundations-of-eu-personal-data-protection-law-privacy-and-human-dignity/

[90] Il punto di riferimento pacificamente riconosciuto è dato dalla pronuncia della Pretura di Roma del 6 maggio 1974, che riconosce il diritto di ognuno a non vedersi negata la paternità delle proprie azioni, evitando dunque che si assista a una falsificazione della propria identità. Cfr. V. Zeno-Zenkovich, *Identità personale*, in Digesto delle discipline privatistiche – Sez. civile, IX, Utet, Torino, 1993, p. 295

[91] Una delle prime riflessioni dottrinali perviene da Adriano de Cupis, che traduce il diritto all'identità personale nell' «obbligo di rispetto della verità personale», da intendersi non solo come tutela verso segni distintivi quale il nome, ma anche quale strumento contro una falsa rappresentazione dell'individuo (sia *in melius* che *in peius*). Cfr. A. De Cupis, *La verità nel diritto (Osservazioni in margine a un libro recente)*, in Foro italiano, n. 4, 1952, pp. 223 s.

collocarsi fra i diritti di "prima generazione", in quanto comunque successivo alla Carta entrata in vigore nel 1948[92].

I contributi della dottrina e della giurisprudenza alla elaborazione del diritto all'identità personale sono molteplici[93], ma nell'economia di questo scritto vale concentrarsi solo su quelli inerenti il rapporto fra il diritto in esame e la Costituzione italiana.

Fra le molte sul punto, assume particolare rilevanza la sentenza n. 978 del 1996 della Corte di Cassazione, che individua nell'art. 2 della Costituzione il fondamento costituzionale del diritto all'identità personale, «in correlazione anche all'obiettivo primario di tutela del "pieno sviluppo della persona umana", di cui al successivo art. 3»[94].

Tuttavia, non mancano posizioni per le quali fondare il diritto all'identità personale sui soli articoli 2 e 3 Cost. sarebbe insufficiente, dovendo considerarsi anche l'art. 21 relativo alla libertà di manifestazione del pensiero[95]: secondo tale prospettiva, attribuire a un individuo opinioni mai espresse lede il suo diritto ad essere riconosciuto solo per ciò che effettivamente dice e pensa e anche a non esprimere ciò che pensa.

Alla stessa linea di pensiero si aggiungano le posizioni secondo le quali l'identità personale, poiché non si riduce al nome o all'immagine, necessiterebbe di una protezione ancor più ampia, leggendo l'art. 2 Cost. in combinato disposto con numerosi altri articoli della Carta (1, 3, 4, 19, 21, 29)[96].

---

[92] A. Randazzo, *Diritto all'identità personale e valori costituzionali. Le linee di un modello, traendo spunto da Luigi Pirandello*, in Diritti Fondamentali, n. 3, 2021, p. 366

[93] Per un approfondimento sul tema, cfr. G. Pino, *Il diritto all'identità personale. Interpretazione costituzionale e creatività giurisprudenziale*, Il Mulino, Bologna, 2003, pp. 55-125; A. Randazzo, *Diritto all'identità personale e valori costituzionali. Le linee di un modello, traendo spunto da Luigi Pirandello*, in Diritti Fondamentali, n. 3, 2021, pp. 366-376

[94] Corte di Cassazione, sez. I civile, sent. 7 febbraio 1996, n. 978

[95] G. Pino*, L'identità personale*, in AA.VV., *Gli interessi protetti nella responsabilità civile*, vol. II, Utet, Torino, 2005, p. 371

[96] A. Randazzo, *Diritto all'identità personale e valori costituzionali. Le linee di un modello, traendo spunto da Luigi Pirandello*, in Diritti Fondamentali, n. 3, 2021, p. 379

In ogni caso, per la giurisprudenza (ma non per tutta la dottrina) la rilevanza costituzionale del diritto all'identità personale è indiscutibile, poiché esso è parte del patrimonio inviolabile della persona umana (art. 2 Cost.), come «diritto ad essere sé stesso» e comprensivo di tutto ciò che l'individuo ha acquisito nel tempo (come le idee o le convinzioni morali, religiose e sociali)[97].

Questa interpretazione della giurisprudenza è stata recepita dal legislatore il quale, prima col già citato art. 1 della Legge 675/1996, e poi con il Codice in materia di protezione dei dati personali (D.lgs. 196/2003), considera la riservatezza e l'identità personale collegate «all'orizzonte dei diritti e delle libertà fondamentali e quindi costituzionali»[98].

In conclusione, il formante giurisprudenziale e quello legislativo del diritto hanno già posto individuato solide basi per la tutela del diritto in esame, ma l'identità personale si intreccia con molteplici profili dell'individuo ai quali è connessa la concreta attuazione dei principi costituzionali.

In questo quadro si collocano anche le minacce cui l'identità personale è sottoposta a causa delle nuove tecnologie[99], a conferma della sua natura di concetto «sottoposto a continue ridefinizioni e movimenti»[100]: e questa condizione richiede che il giurista sappia, in ogni circostanza, «inquadrarla nei suoi esatti termini giuridici»[101].

[97] Corte Costituzionale, sent. 3 febbraio 1994, n. 13

[98] G. Pino, *L'identità personale*, in AA.VV., *Gli interessi protetti nella responsabilità civile*, vol. II, Utet, Torino, 2005, p. 372

[99] A. Morelli, *Persona e identità personale*, in *BioLaw Journal*, n. 2, 2019, p. 49

[100] A. Randazzo, *Diritto all'identità personale e valori costituzionali. Le linee di un modello, traendo spunto da Luigi Pirandello*, in Diritti Fondamentali, n. 3, 2021, p. 390

[101] A. De Cupis, *Bilancio di un'esperienza*, in AA. VV., La lesione dell'identità personale e il danno non patrimoniale. Atti del seminario promosso dal centro di iniziativa giuridica P. Calamandrei. Messina 16 aprile 1982, Giuffrè, Milano, p. 197

3. Profili giurisprudenziali di identità personale, reputazione e dignità

Fino a questo momento sono state più volte richiamate le espressioni «identità personale», «reputazione» e «dignità», spesso oggetto di un'erronea sovrapposizione. In chiusura di questo capitolo introduttivo, si ritiene che una loro breve analisi, condotta nell'ottica della giurisprudenza italiana, possa essere strumentale a una più precisa comprensione circa il funzionamento dei *deepfakes*, i quali possono ledere singolarmente o congiuntamente questi delicati aspetti dell'essere umano.

Il principale punto di contatto di questi tre concetti è dato dalla disposizione che tutela i c.d. diritti della personalità, intesi come quelli posti a tutela della dimensione relazionale della persona e degli aspetti ad essa connessi, come la conoscenza, la riconoscibilità e l'identificazione all'interno della società[102]. Invero, i diritti della personalità trovano fondamento nell'art. 2 della Costituzione il quale, unitamente ai richiami contenuti nelle carte sovranazionali[103], consente di ricomprendere in tale categoria anche situazioni giuridiche non ancora disciplinate dalla legge: la stessa giurisprudenza conferisce alla disposizione natura di "clausola aperta" in quanto, valorizzando la «persona umana in tutti i suoi aspetti», legittima l'interprete «a costruire tutte le posizioni soggettive idonee a dare garanzia, sul terreno dell'ordinamento positivo, ad ogni proiezione della persona nella realtà sociale»[104]. Sulla base dell'art. 2 e di altre disposizioni costituzionali e ordinarie, la giurisprudenza ha poi elaborato

[102] G. Pino, *Teorie e dottrine dei diritti della personalità. Uno studio di meta-giurisprudenza analitica*, in Materiali per una storia della cultura giuridica, n. 1, 2003, p. 246

[103] In effetti, le fonti a tutela dei diritti della personalità non si risolvono nel solo art. 2 della Costituzione. A livello interno vi sono il Codice civile (artt. 5-10) nonché leggi ordinarie quali il GDPR. A queste si aggiungano le numerose fonti sovranazionali, tra le quali si citano la Dichiarazione universale dei diritti umani delle Nazioni Unite del 1948 o le già menzionate Convenzione EDU e Carta di Nizza.

[104] Corte di Cassazione, sez. III Civile, sent. 10 maggio 2001, n. 6507

progressivamente nuovi diritti[105] quali, appunto, quelli alla riservatezza e all'identità personale[106].

Pur condividendo la tesi che il loro fondamento risiede nell'art. 2 Cost., si deve rammentare che identità personale, reputazione e dignità non possono considerarsi sinonimi e, pur riconoscendo che discutere dell'approccio giurisprudenziale a questi tre concetti richiederebbe una più ampia e precisa trattazione (che tenga conto anche del contesto sovranazionale), per l'oggetto del presente elaborato saranno esaminate soltanto alcune delle pronunce nazionali utili a inquadrarli in via generale.

Dell'identità personale si è già trattato prima[107], occorrendo ora concentrarsi sulle altre due nozioni. La Corte di Cassazione definisce compiutamente la reputazione, statuendo che essa non si risolve in uno stato emotivo individuale e sconnesso dal mondo esterno, ma che al contrario dipende da quest'ultimo, rappresentando «il senso della dignità personale nell'opinione degli altri»[108]. L'indicazione della Cassazione è utile in quanto separa l'interesse alla reputazione, leso dal reato di diffamazione (art. 595 c.p.), da quello all'onore, col quale si intende la percezione soggettiva che l'individuo ha di sè[109], in passato tutelato con l'abrogato art. 594 c.p. (delitto di ingiuria). Ad ogni modo, la reputazione non può ridursi a semplice interesse, costituendo «un vero e proprio

[105] S. Thobani, *Diritti della personalità e contratto: dalle fattispecie più tradizionali al trattamento in massa dei dati personali*, Ledizioni, Milano, 2018, p. 14

[106] In tempi più recenti anche la dignità, principio fondante dell'ordinamento, è stata riconosciuta come diritto inviolabile della persona: così facendo, la giurisprudenza segna ulteriormente il processo di «depatrimonializzazione del diritto privato», che vede quest'ultimo dall'occuparsi unicamente dei rapporti economici a disciplinare anche situazioni che non lo sono. Cfr. S. Thobani, *Diritti della personalità e contratto: dalle fattispecie più tradizionali al trattamento in massa dei dati personali*, Ledizioni, Milano, 2018, p. 14

[107] Per una definizione giurisprudenziale sull'identità personale, utile per un confronto con quelle di reputazione e dignità, si riprenda quanto detto dalla Corte Costituzionale con sent. n. 13/1994 (v. *supra*, 2.)

[108] Corte di Cassazione, Sez. V Penale, sent. 28 febbraio 1995, n. 3247

[109] A. Cataudella, *Dignità, decoro e identità personale*, in Il Diritto dell'informazione e dell'informatica, n. 2, 1985, p. 575

diritto soggettivo perfetto»[110]: la sua rilevanza è tale da porre limiti ad altri diritti costituzionalmente garantiti quali il diritto di stampa o di cronaca, declinazioni della libertà di manifestazione del pensiero *ex.* art. 21 Cost[111].

Altro dato rilevante sul punto è che il concetto di reputazione, negli ultimi anni, ha cominciato ad assumere nuove vesti. Infatti, oggi si considera anche la reputazione digitale (o *web reputation*), espressione che indica la percezione che gli utenti del *web* maturano verso un determinato soggetto[112]. La nozione di reputazione digitale presenta un'importante peculiarità rispetto a quella tradizionale, fondata su un giudizio sociale che il destinatario può controllare solo in modo limitato, poiché si presenta come un «bene che l'interessato può veicolare» o, meglio, uno «strumento di autopromozione commerciale»[113].

In altri termini, la persona fisica o giuridica[114], con l'ausilio degli strumenti digitali, svolge un ruolo attivo nella «creazione della propria sfera reputazionale»[115], ad esempio scegliendo quali immagini o commenti postare sui propri *social network*. Ne deriva che l'art. 595 c.p., pur non distinguendo tra la

---

[110] Corte di Cassazione, Sez. III Civile, sent. 10 maggio 2001, n. 6507

[111] Difatti, l'esercizio di questi diritti è da considerarsi legittimo solamente al ricorrere di tre condizioni: a) utilità sociale dell'informazione; b) verità dei fatti, i quali devono essere riportati oggettivamente, con diligenza e senza omettere altre informazioni che alterino il senso complessivo della notizia; c) continenza, vale a dire un'esposizione dei fatti civile, finalizzata all'informazione e priva di intenti denigratori o lesivi della dignità della persona. Cfr. Corte di Cassazione, sez. I Civile, sent. 18 ottobre 1984, n. 5259

[112] F. De Stefani, *Web reputation: cos'è e come curare la reputazione online*, Agenda Digitale, 2020. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/web-reputation-cose-e-come-curare-la-reputazione-online/

[113] F. Sanzari, *Reputazione, diritti della personalità, internet*, in S. Previti, *Internet e reputazione aziendale*, Utet giuridica, Torino, 2016, p. 24

[114] Il discorso vale in particolare per le imprese, che vedono nella reputazione un importante investimento per aumentare la loro competitività sul mercato, valorizzare la loro *brand identity* e proteggerla da potenziali minacce (come la diffamazione o la contraffazione del marchio). La reputazione digitale, in tal senso, assume i tratti di un vero e proprio cespite aziendale. Cfr. F. Sanzari, *Reputazione, diritti della personalità, internet*, in S. Previti, *Internet e reputazione aziendale*, Utet giuridica, Torino, 2016, pp. 24-30

[115] Ivi, p. 26

reputazione tradizionale e quella digitale, presenta un forte nesso anche con quest'ultima. Tale legame emerge grazie alla giurisprudenza della Cassazione, che oltre vent'anni fa ha ritenuto configurabile il reato di diffamazione «anche quando la condotta dell'agente consista nella immissione di scritti o immagini lesivi dell'altrui reputazione nel sistema "internet"», riconoscendo inoltre la circostanza aggravante di cui all'art. 595 comma 3 c.p.[116]. Nonostante la sentenza non si riferisca direttamente alla sfera "digitale" della reputazione, è evidente che la diffamazione commessa online incida direttamente su tale aspetto[117].

Resta infine da considerare la dignità umana, che il costituzionalismo novecentesco ha elevato a «valore giuridico supremo dell'ordinamento»[118], individuando in essa la chiave per evitare il ripetersi degli orrori del secondo conflitto mondiale[119]. Fra le molte accezioni esistenti, si può definire la dignità come la condizione in cui l'individuo è conscio del proprio valore ed è rispettato dalla società sia per il ruolo che in essa vi svolge che per le qualità proprie[120].

---

[116] Corte di Cassazione, Sez. V Penale, sent. 17 novembre 2000, n. 4741

[117] Altra prova della centralità progressivamente assunta dalla reputazione digitale è data dal c.d. diritto all'oblio (che sarà ripreso più avanti), qualificato dalla giurisprudenza interna «un nuovo profilo del diritto di riservatezza», e che si sostanzia nel «giusto interesse di ogni persona a non restare indeterminatamente esposta ai danni ulteriori che arreca al suo onore e alla sua reputazione la reiterata pubblicazione di una notizia in passato legittimamente divulgata» (Cfr. Corte di Cassazione, sez. III Civile, sent. 9 aprile 1998, n. 3679). La formulazione, risalente a quasi tre decadi fa, non poteva tener conto della reputazione digitale: oggi, al contrario, il diritto all'oblio si applica prevalentemente al contesto digitale, e anche la giurisprudenza sembra essere conscia della sua evoluzione, «sicché nel caso di notizia pubblicata sul "web"» l'interessato potrà «trovare soddisfazione anche nella sola "deindicizzazione" dell'art. dai motori di ricerca» (Cfr. Corte di Cassazione, sez. I Civile, sent. 19 maggio 2020, n. 9147).

[118] F. Fernández Segado, *La dignità della persona come valore supremo dell'ordinamento giuridico spagnolo e come fonte di tutti i diritti*, Forum di Quaderni Costituzionali, 2010. Reperibile al sito: https://www.forumcostituzionale.it/wordpress/images/stories/pdf/documenti_forum/paper/0214_segado.pdf

[119] Emblematico in tal senso è il Preambolo della Dichiarazione universale dei diritti dell'uomo, che cita: «Il riconoscimento della dignità inerente a tutti i membri della famiglia umana e dei loro diritti, uguali ed inalienabili, costituisce il fondamento della libertà, della giustizia e della pace nel mondo».

[120] A. Cataudella, *Dignità, decoro e identità personale*, in Il Diritto dell'informazione e dell'informatica, n. 2, 1985, p. 576

Con una scelta non casuale, l'ordinamento italiano non ha mai definito esaustivamente la dignità. I Costituenti valorizzarono una nozione «sociale della dignità che possa sostenere e legittimare diritti umani di persone concrete, in condizioni storiche, politiche e sociali date»[121]. In altri termini, l'intento fu quello di creare un legame fra dignità dell'uomo e riconoscimento della sua dimensione sociale, obiettivo che sarà raggiunto con l'art. 2 della Costituzione, il quale connette i «diritti inviolabili dell'uomo» con i «doveri inderogabili di solidarietà»[122]. La disposizione costituisce dunque il fondamento costituzionale della dignità umana tant'è che, secondo la Corte Costituzionale, i diritti inviolabili da essa tutelati sono «coessenziali alla dignità della persona»[123]. A perfezionamento di ciò vi è l'art. 3 Cost. [124], che proclama il principio di uguaglianza ma antepone ad esso la «pari dignità sociale». Dall'unione di questi due elementi sorge l'obbligo per la Repubblica, ai sensi del secondo comma, di rimuovere qualsiasi ostacolo che limiti libertà e uguaglianza, così rendendo la dignità non un principio formale, ma effettivo e che tenga conto della persona «nel concreto della sua situazione sociale, con le sue difficoltà e debolezze»[125]. Il quadro è completato dalla giurisprudenza[126], che nel tempo ha ravvisato un collegamento, seppur indiretto, fra la dignità umana e numerose disposizioni costituzionali[127].

---

[121] G. Azzariti, *Contro il revisionismo costituzionale*, Editori Laterza, Roma-Bari, 2016, p. 167

[122] A. Apostoli, *La dignità sociale come orizzonte della uguaglianza nell'ordinamento costituzionale*, in Costituzionalismo.it, n. 3, 2019, pp. 6-7

[123] Corte Costituzionale, sent. 22 ottobre 1999, n. 388

[124] Si tratta dell'unica disposizione nel testo costituzionale, assieme all'art. 41 in tema di iniziativa economica privata, a citare espressamente la dignità umana.

[125] L. Carlassare, *Nel segno della Costituzione. La nostra carta per il futuro*, Feltrinelli, Bergamo, 2012, p. 20

[126] Per approfondimenti sul punto, cfr. G.P. Dolso, *Dignità, eguaglianza e Costituzione*, EUT Edizioni Università di Trieste, Trieste, 2019, pp. 55-73

[127] Si faccia, fra i molti, l'esempio del diritto alla salute: mentre il comma 2 dell'art. 32 Cost. individua i limiti ai trattamenti sanitari «nel rispetto della persona umana», la Consulta specifica che i trattamenti sanitari «trovano un limite non valicabile nel rispetto della dignità della persona che vi può essere sottoposta». Cfr. Corte Costituzionale, sent. 2 giugno 1994, n. 218.

Occorre infine focalizzare l'attenzione sul rapporto tra identità, reputazione e dignità. Dall'analisi condotta sinora è emerso che: la dignità rappresenta quel «principio costituzionale che informa di sé il diritto positivo vigente»[128]; all'art. 2 Cost. si attribuisce il senso di «significato assoluto della dignità umana»[129]; la medesima disposizione ha natura di "clausola aperta" e costituisce la base dell'elaborazione dei diritti della personalità. Sulla scorta di tali osservazioni si può dunque concludere che la dignità costituisce nucleo e fondamento di identità personale e reputazione, ispirando tutti i diritti inviolabili riconosciuti dall'ordinamento.

La considerazione del ruolo che la dignità, l'identità personale e la reputazione assumono nell'ordinamento consente di rivelare la varietà di beni giuridici che un *deepfake* può compromettere e quante forme possa assumere la manipolazione tramite intelligenza artificiale. In effetti, il fenomeno dei *deepfakes* pone questioni inedite alla giurisprudenza italiana e rende necessario interrogarsi sull'effettiva idoneità degli strumenti di tutela oggi presenti.

[128] Corte Costituzionale, sent. 17 luglio 2000 n. 293

[129] P. Becchi, *Dignità umana*, in U. Pomarici, *Filosofia del diritto. Concetti fondamentali*, Giappichelli, Torino, 2007, p. 161

CAPITOLO II

DALL'IA GENERATIVA AI DEEPFAKES: TECNOLOGIE E RISCHI

SOMMARIO: 1. Cos'è l'IA generativa?. – 2. *Deepfakes*, voci sintetiche, avatar: applicazioni e rischi. – 3. *Dataset* e *training* non autorizzato.

1. Cos'è l'IA generativa?

Nel presente capitolo si tratterà della reale natura di un *deepfake*. Tuttavia, una piena comprensione del fenomeno richiede di muovere da solide premesse: anzitutto, gli elementi suscettibili di lesione a seguito del suo utilizzo, già esaminati nel capitolo precedente; in secondo luogo, la tecnologia (con il suo funzionamento) da cui il *deepfake* trae origine, tema che sarà ora approfondito.

A tal fine, occorre cominciare dal concetto di Intelligenza Artificiale (d'ora in avanti, IA) in senso ampio. Una delle definizioni più autorevoli proviene dalla Commissione europea, secondo cui l'IA comprende «sistemi che mostrano un comportamento intelligente analizzando l'ambiente circostante e intraprendendo azioni – con un certo grado di autonomia – per raggiungere obiettivi specifici». Questi sistemi, prosegue la Commissione, possono essere «puramente basati su software, agendo nel mondo virtuale (ad esempio assistenti vocali, software di analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale)»; oppure incorporati «in dispositivi hardware (ad esempio robot avanzati, auto autonome, droni o applicazioni Internet delle cose)» [130].

Tra le sue numerose modalità di applicazione, assume particolare importanza la distinzione tra i modelli «predittivi» e quelli «generativi» i quali,

[130] The European Commission's High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*, 2018. Reperibile al sito: https://ec.europa.eu/futurium/en/system/*files*/ged/ai_hleg_definition_of_ai_18_december_1.pdf

pur non esaurendo tutte le funzioni dell'IA, ne costituiscono i sistemi più utilizzati e rivoluzionari. L'IA predittiva utilizza dati esistenti per prevedere se uno o più eventi si verificheranno e valutarne il possibile impatto. L'IA generativa, diversamente, va oltre la sola predizione: basandosi sui dati di addestramento e sul *prompt* (ossia, la richiesta testuale) dell'utente, è in grado di creare contenuti di vario tipo come un'immagine, un grafico o un brano musicale[131]. Essendo in grado di generare contenuti realistici, si comprende che è proprio questo il sistema grazie al quale si producono i *deepfakes*, come già accennato (v. *supra*, cap. I, 1.), e al quale deve ora dedicarsi particolare attenzione.

Alla base del funzionamento dell'IA generativa vi sono modelli di *machine learning* e, in misura prevalente, di *deep learning*. Per *machine learning* (ML), o apprendimento automatico, si intende quella branca dell'IA che permette alla macchina di «apprendere autonomamente attraverso l'esperienza»[132]: analizzando una grande quantità di dati, essa individua «relazioni precedentemente sconosciute», che si traducono nel concreto in regole o nella facoltà di elaborare previsioni[133]. Il *deep learning* è una sottocategoria più sviluppata del *machine learning*: a differenza di quest'ultima, dotata di un'architettura composta da uno o due livelli di reti neurali, il *deep learning* impiega reti neurali c.d. profonde, ossia con un numero di livelli che può arrivare fino a centinaia o migliaia.[134]. Tali reti sono poi costituite da più strati di «neuroni artificiali», che si possono sinteticamente definire come «strutture matematiche» ispirate alle omonime cellule del cervello umano. Ciascun neurone artificiale

---

[131] F. Pincelli, *AI generativa e analisi predittiva: perché non se ne può più fare a meno*, BizzIT, 2025. Reperibile al sito: https://bizzit.it/approfondimenti/ai-generativa-e-analisi-predittiva-perche-non-se-ne-puo-piu-fare-a-meno/

[132] G. Buonomo, G. Ciacci, A. Costanzo, *Macchine intelligenti e diritto*, Giappichelli, Torino, 2025, p. 10

[133] M. Chen, *What Is Machine Learning?*, Oracle, 2024. Reperibile al sito: https://www.oracle.com/artificial-intelligence/machine-learning/what-is-machine-learning/

[134] J. Holdsworth, M. Scapicchio, *What is deep learning?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/it-it/think/topics/deep-learning

riceve informazioni codificate in cifre da quello precedente, le combina al fine di ottenere nuove correlazioni e trasmette la nuova e più complessa elaborazione al neurone successivo, fino ad arrivare alla decisione finale (come il riconoscimento di un'immagine)[135].

Sebbene l'espressione «IA generativa» sia entrata nel linguaggio comune solo da alcuni anni, le sue radici sono ben più antiche. Tra i primi modelli in tal senso si annovera "ELIZA": sorto negli anni Sessanta, è un programma di elaborazione del linguaggio naturale il quale, mediante un rudimentale algoritmo di *pattern matching*[136], simulava il dialogo con uno psicoterapeuta virtuale allo scopo di studiare le interazioni fra uomo e macchina e i processi cognitivi che ne derivano[137]. È però nell'ultimo decennio che sono stati sviluppati modelli capaci di trasformare l'IA, da strumento riservato agli esperti o a chi ne chiedeva l'accesso, a tecnologia utilizzata e diffusa a livello mondiale[138].

Primo fra questi è l'Autoencoder variazionale (VAE). Questo modello, introdotto nel 2013, rappresenta un'evoluzione dell'autoencoder tradizionale, un sistema di deep learning costituito da due reti neurali connesse: un *encoder*, che

---

[135] C. Negri, *Cos'è il Deep Learning: esempi e applicazioni reali*, Osservatori.net, 2025. Reperibile al sito: https://www.osservatori.net/blog/artificial-intelligence/deep-learning-significato-esempi-applicazioni/

[136] Nel dettaglio, il funzionamento di ELIZA si componeva di tre frasi: acquisizione dell'*input* dell'utente; individuazione di parole chiave; generazione di una risposta predefinita tramite regole basate su strutture sintattiche. Il risultato è che l'*output* si traduceva in una mera riformulazione dell'*input* e, qualora il sistema non riuscisse a rilevare parole chiave, forniva risposte generiche come "Dimmi di più". Cfr. *ELIZA: The First Step in Human-Computer Interaction Through Natural Language Processing*, Geeksforgeeks, 2025. Reperibile al sito: https://www.geeksforgeeks.org/artificial-intelligence/eliza-the-first-step-in-human-computer-interaction-through-natural-language-processing/

[137] Nonostante "ELIZA" sia considerato il primo chatbot della storia, la finalità principale della sua realizzazione è quella sopraindicata di natura psicosociale. Cfr. J. Shrager, *ELIZA Reinterpreted: The world's first chatbot was not intended as a chatbot at all*, arXiv, 2024. Reperibile al sito: https://arxiv.org/pdf/2406.17650

[138] L'esatto momento in cui è esplosa la popolarità dell'IA è individuato nel lancio di GPT-3.5 Turbo, meglio noto come ChatGPT, avvenuto il 30 novembre 2022. Cfr. E. Bezzecchi, *Intelligenza artificiale: Farsi le domande giuste, capire gli scenari futuri e usare in modo smart l'IA generativa*, Vallardi, Milano, 2024

comprime grandi quantità di dati (come immagini con centinaia di *pixel*) riducendoli a pochi numeri essenziali; e un *decoder*, che a partire da tali parametri ricostruisce il contenuto. Dunque, l'autoencoder è maggiormente impiegato per funzioni di riduzione, archiviazione o trasferimento dei dati[139]. La differenza sostanziale con il VAE è che in quest'ultimo l'encoder, nel comprimere i dati, non restituisce singoli valori, ma produce per ciascuno una distribuzione gaussiana definita da una media e una varianza, parlandosi infatti di «metodo probabilistico». A seguito dell'addestramento, le molteplici gaussiane confluiranno poi in una comune distribuzione gaussiana: conseguenza di ciò è che il VAE non solo scompone e ricostruisce dati, ma ne può generare di nuovi mediante un «campionamento casuale», e cioè "pescando" dei punti dall'unica distribuzione ottenuta[140].

Nel 2014 viene introdotto un altro paradigma di IA generativa, la Rete generativa avversaria (GAN). Essa si compone di due reti neurali: un generatore, che crea dati falsi (come immagini o suoni) allo scopo di mettere alla prova il discriminatore; e il discriminatore stesso, che andrà a distinguere e qualificare i dati tra autentici e non. Nella fase di addestramento, si instaura tra le due reti una vera e propria «sfida», col generatore e il discriminatore che migliorano sempre più le loro prestazioni, fino al punto in cui il primo riesce a creare dei contenuti così realistici da non riuscire più ad essere distinti dal secondo[141].

Lo stesso anno vede l'elaborazione dei c.d. modelli di diffusione i quali, partendo da dati di addestramento reali quali immagini, vi aggiungono del «rumore gaussiano»: in altri termini, il modello decostruisce il dato modificando il suo valore numerico di una quantità presa da una distribuzione gaussiana, per poi imparare ad eliminare progressivamente il rumore e a ricostruire il contenuto

---

[139] C. Stryker, M. Scapicchio, *What is generative AI?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/generative-ai

[140] J. Jordan, *Variational autoencoders*, Jeremy Jordan, 2018. Reperibile al sito: https://www.jeremyjordan.me/variational-autoencoders/

[141] N. Bassetti, *GAN (Generative Adversarial Networks): cosa sono, applicazioni e vantaggi*, Agenda Digitale, 2023. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/gan-generative-adversarial-networks-cosa-sono-applicazioni-e-vantaggi/#post-175384-footnote-1

originale. Dopo l'addestramento, il modello parte direttamente dal rumore ed esegue il processo di *de-noising* coerente con l'*output* richiesto[142]. I modelli di diffusione risultano essere più funzionali di architetture precedenti come il GAN[143], al punto da diventare base di strumenti di IA generativa del calibro di DALL-E di OpenAI.

La forma più recente di IA generativa è infine l'architettura Transformer, documentata per la prima volta nel 2017 e motore dei sistemi oggi più all'avanguardia come ChatGPT (GPT è infatti acronimo di *Generative Pre-trained Transformer*) o Microsoft Copilot. La peculiarità di questo sistema di deep learning è nel poter analizzare grandi quantità di dati in parallelo, come avviene per un testo, a differenza dei precedenti modelli che elaborano informazioni una alla volta. Il Transformer rende tale funzione possibile attraverso un'articolata procedura. Anzitutto, il testo è suddiviso in *token*, unità composte da parole o parte di esse e tradotte in cifre. Ciascun token è in seguito convertito in vettore, ossia una rappresentazione numerica delle parole volta a descriverne il significato, arricchita da informazioni circa la loro posizione nella frase grazie la c.d. codifica posizionale. Si attiva poi il meccanismo di attenzione, fulcro dell'intero modello e che consente ad esso non solo di comprendere quali sono i dati più rilevanti nella sequenza, ma anche di individuare relazioni tra le parole, anche a distanza. Il processo si chiude con la fase di decodifica, dove i vettori sono trasformati in testo leggibile dall'uomo[144]. È importante sottolineare che questo processo si verifica più volte nello stesso modello, grazie ai numerosi

---

[142] P. Sandonnini, *Modelli di diffusione, cosa sono e perché offrono nuove possibilità per la creatività generativa*, AI4Business, 2022. Reperibile al sito: https://www.ai4business.it/news/modelli-di-diffusione-cosa-sono-e-perche-offrono-nuove-possibilita-per-la-creativita-generativa/

[143] *Ibidem*. Sebbene le GAN siano state reputate innovative, il meccanismo competitivo tra generatore e discriminatore si è rivelato difficile da gestire: se una delle due reti neurali progredisce più dell'altra, l'addestramento di quest'ultima tende ad interrompersi. Inoltre, è frequente che le GAN incorrano nel c.d. *mode collapse*, uno stato in cui il generatore comprende quali esatti *output* riescono ad ingannare il discriminatore in modo da riproporli costantemente, con la conseguente perdita di varietà e complessiva efficacia della rete neurale.

[144] C. Biagiolini-Jr., *Introduction to Transformers in Generative AI*, Medium, 2024. Reperibile al sito: https://medium.com/%40biagolini/introduction-to-transformers-in-generative-ai-195902f838c3

livelli di deep learning che agiscono in parallelo e dunque in grado di rilevare diverse connessioni semantiche e sintattiche tra i dati[145].

Esaminata la varietà di modelli capaci di generare dati, deve aggiungersi che altrettanto numerose sono le modalità d'uso dell'IA generativa. Tra queste, la più rilevante per il presente elaborato è indubbiamente la creazione di contenuti multimediali come immagini, video[146] e audio[147], che siano realistici o con uno stile in particolare, nonché di opere d'arte o progettazioni grafiche. Da aggiungere vi è, poi, la produzione di contenuti testuali e tutte le attività annesse quali il riassunto, la riformulazione del testo in base al pubblico o alla formalità richiesta, l'estrazione di precise informazioni da testi quali il nome o l'indirizzo o la traduzione nella lingua desiderata[148]. Oltre queste funzionalità generali, è importante osservare che l'IA generativa trova applicazioni pratiche in settori specifici: in ambito sanitario può, ad esempio, sintetizzare immagini mediche come una radiografia o proporre nuove strutture molecolari; nel campo della *customer experience*, grazie l'IA si possono creare pagine web o immagini a scopo di marketing personalizzato; si pensi ancora alla realizzazione di bozze di codici sorgente in grado di orientare i programmatori allo sviluppo di un *software*[149].

In definitiva, l'IA generativa si presenta come uno strumento destinato ad assumere importanza sempre maggiore nella vita dei singoli, sociale e professionale. D'altra parte, se impiegata in assenza di controlli giuridici e senso etico, può portare alla produzione di contenuti fortemente lesivi. È proprio qui

---

[145] Cfr. *Multi-Head Attention Mechanism*, Geeksforgeeks, 2025. Reperibile al sito: https://www.geeksforgeeks.org/nlp/multi-head-attention-mechanism/

[146] Numerosi sistemi di IA generativa, come il già citato DALL-E o il servizio Picasso di NVIDIA, si occupano specificamente di questo e ambiscono a un costante miglioramento della qualità del contenuto. Cfr. D. Shah, *Generative AI 101: Explanation, Use Cases, Impact*, V7 Labs, 2023. Reperibile al sito: https://www.v7labs.com/blog/generative-ai-guide

[147] Si pensi al servizio "Text-To-Speech" di Google, in grado di generare discorsi e vocalità simili a quella umana, potendo selezionare tra 380 voci e più di 50 lingue.

[148] C. Fregly, A. Barth, S. Eigenbrode, *Generative AI on AWS*, O' Reilly Media, Sebastopol, 2025

[149] C. Stryker, M. Scapicchio, *What is generative AI?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/generative-ai

che entrano in gioco i *deepfakes*, visti come «gli esempi più agghiaccianti di come la potenza dell'IA generativa possa essere applicata con intenti malevoli»[150], e che saranno ora oggetto di analisi approfondita.

2. Deepfakes, voci sintetiche e avatar

2.1. I deepfakes: storia, applicazioni e tecnologie

Nell'affrontare il tema dei *deepfakes*, un solido punto di partenza è la più volte richiamata definizione del Garante per la protezione dei dati personali (v. *supra*, cap. I, 1.) la quale, assieme all'etimologia del termine, dato dalla fusione tra «deep learning» e «fake»[151], evidenzia il forte legame tra IA (in particolare quella generativa) e *deepfakes*.

Le origini di questo fenomeno risalgono già agli anni Novanta del secolo scorso, quando grazie la CGI (*Computer Generated Imagery*) fu possibile creare immagini realistiche di esseri umani. In particolare, nel 1997 venne lanciato il progetto «Video Rewrite Program», in grado di modificare i movimenti facciali di una persona esistente e introdurvi un audio da essa mai pronunciato[152]. Una fase di svolta arriva poi nel 2014, quando il ricercatore Ian Goodfellow introduce il concetto di GAN (di cui si è già discusso), tecnologia che consentirà la

[150] *Ibidem*. I *deepfakes* rappresentano solo uno dei rischi, forse quello più dannoso, legati a un'utilizzazione impropria dell'IA generativa. Per completezza si faccia menzione, a titolo di esempio, di pericoli come le c.d. allucinazioni, ossia *output* privi di senso o imprecisi ma dalla forma plausibile; o la generazione di e-mail di *phishing*, un tipo di attacco informatico che, mediante comunicazioni false ma credibili come *e-mail* o SMS, induce la vittima a condividere dati sensibili o scaricare *malware*.

[151] Garante per la protezione dei dati personali, *Deepfake. Il falso che ti «ruba» la faccia (e la privacy)*, 2020. Reperibile al sito: file:///C:/Users/Utente/Downloads/Deepfake%20-%20Vademecum%20(3).pdf

[152] Cfr. C. Begler, M. Covell, M. Slaney, *Video Rewrite: driving visual speech with audio*, in *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, 1997, pp. 353-360

creazione di veri e propri *deepfakes*[153]. L'espressione, così come la conosciamo oggi, compare per la prima volta alla fine del 2017 su Reddit, popolare *social network*, dove un utente chiamato proprio "*deepfakes*" sovrapponeva volti di personaggi pubblici in contesti pornografici grazie un algoritmo di deep learning. Da qui in poi la produzione di immagini e video realistici con l'ausilio dell'IA è esplosa, tant'è che nel 2018 lo stesso utente di Reddit crea l'applicazione FakeApp, con l'obiettivo di rendere accessibile a tutti la creazione di *deepfakes*, senza che siano necessari particolari competenze e servendosi unicamente del proprio *dataset*[154]. Il conseguimento di tale obiettivo diventa sempre più concreto sia con la nascita di software più accurati come FaceSwap e DeepFaceLab, nonostante essi richiedano dispositivi di qualità, con elevata RAM (*Random Access Memory*) e buona scheda grafica; sia con l'arrivo di applicazioni anche per smartphone come Reface, che permettono di scambiare un volto in un video con una qualsiasi delle immagini presenti nella propria galleria[155].

Le finalità perseguite con la produzione dei *deepfake* sono molteplici. Nella meno dannosa delle ipotesi, essi sono utilizzati a scopo di intrattenimento, nonostante comunque sia assente il consenso dell'interessato: si pensi alla sovrapposizione di un volto celebre in una scena di un film[156]. Tuttavia, nella gran parte dei casi le repliche credibili nascono col chiaro intento di riprodurre la vittima in situazioni compromettenti o lesive della sua reputazione, come avviene per i *deepfake* che colpiscono i politici allo scopo di influenzarne

[153] G. Regan, *A Brief History of Deepfakes*, Reality Defender, 2024. Reperibile al sito: https://www.realitydefender.com/insights/history-of-*deepfakes*

[154] S. Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, Vice, 2018. Reperibile al sito: https://www.vice.com/en/article/reddit-fake-porn-app-daisy-ridley/

[155] L. Cocco, *Deepfake: cosa sono e come riconoscerli per smascherare la disinformazione*, Cyber Security 360, 2025. Reperibile al sito: https://www.cybersecurity360.it/nuove-minacce/deep-fake-cosa-sono-e-come-riconoscerli-per-smascherare-la-disinformazione/

[156] Un esempio è dato dall'inserimento del volto di Donald Trump, attuale presidente degli Stati Uniti, nella figura di Biff Tannen nel film Back to the Future (1985). Reperibile al sito: https://www.youtube.com/watch?v=Qr5LwYXyLtw

l'elettorato[157]. È importante però chiarire che il fenomeno dei *deepfake* non colpisce solo le celebrità e, anzi, queste ultime potrebbero non subirne le conseguenze peggiori: a essere maggiormente esposte sono le persone comuni, in quanto forte è la quantità di post e immagini conservate su Internet, in particolare con l'avvento dei social network. A causa di questa «sorta di impronta digitale»[158], tutti gli utenti del web corrono il rischio di «subire una perdita di controllo sulla loro immagine, sulle loro idee e sui loro pensieri»[159].

Tornando alle modalità di applicazione dei *deepfake*, vi è anzitutto il c.d. *deepnude*, sottocategoria dei primi e che ritrae il volto dell'interessato in un corpo nudo, impegnato in atti di natura sessualmente esplicita o inserito in contesti pornografici, parlandosi in quest'ultimo caso di *deepfake porn*. Il reale pericolo sta nel fatto che contenuti di questo tipo possono essere usati per fenomeni come il c.d. *revenge porn*, ovverosia la «condivisione online a scopo di ricatto, denigrazione o vendetta, da parte di ex partner, amanti o spasimanti respinti»; per il sexting, cioè lo scambio e diffusione di immagini intime che può coinvolgere anche minori; o ancora per promuovere pornografia illegale e pedopornografia[160]. Un'altra utilizzazione dei *deepfakes* riguarda il cyberbullismo, dove video falsi sono generati allo scopo di denigrare e offendere il malcapitato, aggiungendosi talvolta richieste di denaro per la cancellazione del contenuto lesivo. A completare il quadro vi è il ricorso ai contenuti falsi per il c.d. *cybercrime*[161], ossia per illeciti informatici quali il *phishing* (v. *supra*, nota

[157] *Reato di deepfake: che cosa prevede il disegno di legge sull'IA?*, Futuro Digitale, 2025. Reperibile al sito: https://futurodigitale.infocert.it/pillole-normative/reato-di-*deepfake*/

[158] A. Banks, *Op-Ed. Deepfakes & Why the Future of Porn is Terrifying*, Highsnobiety, 2018. Reperibile al sito: https://www.highsnobiety.com/p/what-are-*deepfakes*-ai-porn/

[159] Garante per la protezione dei dati personali, *Deepfake. Il falso che ti «ruba» la faccia (e la privacy)*, 2020. Reperibile al sito: file:///C:/Users/Utente/Downloads/Deepfake%20-%20Vademecum%20(3).pdf

[160] *Ibidem*

[161] *Ibidem*

149) o lo *spoofing*, dato da una falsificazione dell'identità allo scopo di ingannare le vittime e acquisire indebitamente loro dati[162].

Una delle questioni più delicate in tema di *deepfakes*, che hanno come punto di forza proprio quello di sfumare il confine tra falso e reale, riguarda il loro riconoscimento. Numerosi esperti hanno nel tempo individuato segnali dinanzi il quale è possibile sciogliere il dubbio: la pelle può apparire troppo liscia o rugosa, o non coerente con altri tratti come i capelli o gli occhi; le ombre sul volto potrebbero non essere realistiche o posizionate in maniera corretta; barba, baffi, nei, o il movimento del corpo e delle labbra risultano innaturali; ancora, la frequenza del battito delle palpebre può essere troppo scarsa o eccessiva[163].

I *deepfakes*, come più volte evidenziato, costituiscono un prodotto dell'IA generativa, ma occorre in questa sede esaminare le concrete tecnologie alla base del loro funzionamento. In primo luogo, vi sono le Reti generative avversarie: richiamate all'inizio del paragrafo e del presente capitolo, il loro ruolo nell'espansione dei *deepfakes* è dovuto alla loro particolare architettura. La costante "sfida" tra generatore e discriminatore permette di individuare difetti nel *deepfake* e migliorarlo progressivamente. Aggiungendo a questo processo il fatto che le GAN analizzano una grande quantità di dati, ne deriva un prodotto finale così realistico da impedire al discriminatore di riconoscerlo come un falso[164]. Un modello alternativo col quale i *deepfakes* possono essere generati è l'autoencoder, la cui struttura, composta da un encoder che decomprime i dati e da un decoder che li ricostruisce, è già stata oggetto di trattazione (v. *supra*, cap. II, 1.). Tra le sue funzioni più innovative vi è proprio lo scambio di volti: se si vuole inserire il volto di una persona A in un video che ritrae una persona B, sarà necessario utilizzare due autoencoder tra loro collegati, con un encoder condiviso e un decoder per ciascuno. In fase di addestramento, il modello comprime e

[162] È il caso della multinazionale britannica ad Hong Kong, truffata per 25 milioni di dollari e di cui si è già discusso (v. *supra*, cap. I, 1.2.)

[163] M. Groh, *Detect Deepfakes: How to counteract misinformation created by AI*, MIT Media Lab, 2025. Reperibile al sito: https://www.media.mit.edu/projects/detect-fakes/overview/

[164] C. Lisi, *Deepfake: cosa sono, come crearli e come riconoscerli*, Dogma. Reperibile al sito: https://www.dogma.it/it/news/*deepfake*--cosa-sono--come-crearli-e-come-riconoscerli

ricostruisce le immagini di entrambi i soggetti ma, essendoci un unico codificatore, esse saranno tradotte in un linguaggio numerico comune che descrive caratteristiche generali come l'ampiezza del sorriso o la direzione dello sguardo. Il risultato è che gli stessi parametri potranno essere decodificati in entrambi i decoder. Lo scambio di volti avviene semplicemente codificando il volto di B e ricostruendolo con il decoder di A: in tal modo, il corpo e le espressioni facciali saranno di B, ma il volto presente sarà quello di A[165].

Le tecnologie finora considerate mostrano come sia possibile falsificare immagini e video, che però non esauriscono il fenomeno dei *deepfakes*: anche la voce può essere oggetto di manipolazione, e su tale aspetto occorrerà approfondire nel paragrafo successivo.

2.2. Voci sintetiche e avatar virtuali

Come illustrato all'inizio dell'elaborato, uno degli elementi integranti dell'identità di un individuo è la voce: logico dunque che se quest'ultima venisse falsificata, anche la prima verrebbe inevitabilmente compromessa. Tale pericolo è oggi reso concreto dai *deepfake* audio. Per comprenderne il funzionamento è necessario partire dalle c.d. voci sintetiche: si tratta di «elementi audio generati da software specifici, concepiti per emulare il parlato umano, conferendo tonalità, accento, espressione ed emozione alle parole»[166]. Le prime voci sintetiche nascono negli anni Sessanta con la tecnologia TTS (*Text-to-speech*), che permette ai computer di convertire l'*input* testuale in linguaggio parlato. Gli *output* dei tradizionali sistemi di TTS si risolvevano in una serie di parole

[165] N. Giansiracusa, *The Simple, Elegant Idea Behind Face-Swap Deepfakes*, Medium, 2021. Reperibile al sito: https://odsc.medium.com/the-simple-elegant-idea-behind-face-swap-*deepfakes*-456a61c73985

[166] *AI per Avatar e Voci Sintetiche: il futuro della comunicazione video*, MM One, 2023. Reperibile al sito: https://www.mm-one.com/intelligenza-artificiale/ai-per-avatar-e-voci-sintetiche-il-futuro-della-comunicazione-video/

predefinite, con voce fredda, robotica e impossibile da modificare[167]. In tempi più recenti, l'integrazione di tale tecnologia con l'IA e i modelli di *deep learning* ha reso possibile regolare precisamente ritmo e tonalità del linguaggio artificiale, così da renderlo sempre più naturale e simile a quello umano[168]. Questo è possibile mediante un processo che combina analisi del linguaggio e sintesi vocale: il sistema, composto da reti neurali addestrate su grandi quantità di dati testuali e audio, riceve il testo come *input* e lo scompone, individuando le componenti essenziali e stabilendo pronuncia, accento e intonazione di ogni parola[169].

Negli ultimi anni il ricorso all'IA ha contribuito non solo a rendere le voci sintetiche più realistiche, ma a creare una sottocategoria di queste ultime, il c.d. *voice cloning*: questa tecnica permette la generazione di una voce digitale identica a quella campionata ed è proprio questo lo strumento col quale un *deepfake* audio viene ad esistenza. A differenza della voce sintetica, generata *ex novo* e solo ispirata a quella umana, con il *voice cloning* si replica una voce di una persona reale, spesso allo scopo di attribuire ad essa dichiarazioni mai pronunciate e rubarne l'identità. Anche in tal caso è necessario un procedimento che si compone di più fasi: dopo aver raccolto una gran quantità di audio della persona la cui voce vuole essere clonata, questi ultimi sono scomposti al fine di trarre le caratteristiche distintive della voce (come avviene per il processo di TTS). Tali componenti saranno poi utili per il *training* del modello, che una volta addestrato sarà in grado di generare un *output* vocale identico a quello originale[170].

---

[167] Si faccia l'esempio di DECtalk, sintetizzatore vocale progettato negli anni Ottanta e noto per essere stato per lungo tempo la voce di Stephen Hawking.

[168] D. Benedetti, *Voce artificiale: gli ultimi sviluppi, anche per il business*, Agenda Digitale, 2021. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/voce-artificiale-ma-con-personalita-progressi-e-nuovi-usi-del-linguaggio-sintetico/#post-117432-footnote-1

[169] G. Brooker, *How does text-to-speech AI (TTS) work?*, LivePerson, 2024. Reperibile al sito: https://www.liveperson.com/blog/text-to-speech-ai/

[170] *What is Voice Cloning?*, ElevenLabs, 2024. Reperibile al sito: https://elevenlabs.io/blog/what-is-voice-cloning#2-what-is-voice-cloning

Per concludere il discorso sui *deepfakes*, è utile infine distinguerli dal concetto di avatar virtuale, col quale si indica «una rappresentazione virtuale di un'identità fisica in un ambiente digitale», che rende possibile «il riconoscimento e l'interazione tra gli utenti che popolano i mondi virtuali»[171]. Presenti sin dagli anni Novanta, gli avatar hanno nel tempo assunto molteplici forme, da assistenti virtuali a NPC (*Not Playable Character*) nel mondo dei videogiochi, ma soltanto negli ultimi anni, e ancora una volta grazie l'apporto fornito dall'IA, stanno evolvendo da uno stile *cartoon* a rappresentazioni fotorealistiche[172].

Il concetto di avatar è spesso assimilato a quello di *deepfake*, nonostante le sostanziali differenze. I *deepfakes* nascono per ingannare, manipolare o sfruttare la fiducia dell'interessato, mentre gli avatar mirano a rafforzarla, promuovendo connessione con l'utente e dichiarando con trasparenza le proprie finalità. La distinzione è netta anche sul piano etico: il *deepfake* è realizzato senza il consenso della persona coinvolta, al contrario gli avatar includono il consenso dell'utente tra i suoi valori fondamentali. Da ultimo, differenti sono le sfere su cui i due elementi operano: intrattenimento, truffe, mondo della politica o pornografia per quanto riguarda i *deepfakes*; marketing, assistenza, sanità o *e-learning* per gli avatar virtuali[173].

3. Dataset e training non autorizzato

Nel presente capitolo sono stati analizzati i *deepfakes* e la tecnologia che ne costituisce la base, ossia l'IA generativa. Per comprendere appieno il funzionamento di entrambi gli elementi occorre aggiungere un ultimo tassello,

[171] F. La Trofa, *Avatar 3D: cos'è e come funziona la nostra identità nel metaverso*, Tech4Future, 2022. Reperibile al sito: https://tech4future.info/avatar-3d-cose-come-funziona/

[172] *AI per Avatar e Voci Sintetiche: il futuro della comunicazione video*, MM One, 2023. Reperibile al sito: https://www.mm-one.com/intelligenza-artificiale/ai-per-avatar-e-voci-sintetiche-il-futuro-della-comunicazione-video/

[173] T. Shahasane, *Deepfakes vs. AI Avatars: A Personal Take on What Sets Them Apart*, Gan.AI, 2024. Reperibile al sito: https://www.gan.ai/blog/posts/*deepfakes*-vs-ai-avatars-a-personal-take-on-what-sets-them-apart

costituito dal *dataset*: esso non è altro che l'insieme di dati, raccolti e organizzati, necessari per addestrare e validare un algoritmo di IA[174]. È possibile distinguere tre tipologie di *dataset*, a seconda del tipo di dati utilizzati: i *dataset* strutturati, con dati salvati in formati predefiniti e standardizzati, ad esempio sotto forma di tabelle con righe e colonne; i *dataset* semistrutturati, flessibili ma con una organizzazione minima, contenenti *file* quali HTML o XML; infine, i *dataset* non strutturati, le cui informazioni sono eterogenee e non disposte in rigidi schemi[175]. Tale categoria è la più rilevante in questa sede poiché i dati che essa include, come immagini, video o audio, sono essenziali per la creazione di un *deepfake* e, se raffiguranti soggetti realmente esistenti, qualificati come dati personali o, in taluni casi, biometrici.

Il Regolamento generale sulla protezione dei dati (GDPR) impone precise condizioni per il trattamento dei dati personali e biometrici (i quali rientrano nei dati c.d. particolari). Ne consegue che se il loro utilizzo, volto all'addestramento dei modelli, non si conforma a detti requisiti, si configura un'ipotesi di *training* non autorizzato: questo è esattamente ciò che avviene con la creazione dei *deepfakes*, che dunque presentano un vizio ancor prima della loro eventuale diffusione.

Nel prossimo capitolo, dedicato all'attuale inquadramento normativo dei *deepfakes*, si riprenderà proprio dal GDPR, e in particolare dalla definizione di dati biometrici e dalle condizioni che ne garantiscono la liceità.

---

[174] R. Sheldon, *What is a data set?*, TechTarget, 2024. Reperibile al sito: https://www.techtarget.com/whatis/definition/data-set

[175] A. Badman, M. Kosinski, *What is a dataset?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/*dataset*

CAPITOLO III

IL QUADRO NORMATIVO ATTUALE E I SUOI LIMITI

SOMMARIO: 1. Unione Europea. – 1.1. Protezione dei dati personali (GDPR e normativa correlata). – 1.2. DSA Act. – 1.3. AI Act. – 1.4. Diritto d'autore e diritto all'immagine nel contesto europeo. – 2. Stati Uniti. – 2.1. *Copyright Act*. – 2.2. Right of publicity. – 2.3. *Take It Down Act*. – 3. Cina. – 4. I deepfakes in Italia e responsabilità. – 5. I deepfakes come elemento di prova e problemi processuali.

1. Unione Europea

I *deepfakes* costituiscono un prodotto relativamente recente e la cui diffusione, negli ultimi anni, è stata rapida e significativa. In questo contesto, gli ordinamenti nazionali hanno dovuto intervenire tempestivamente con gli strumenti già disponibili. Tale dinamica ha però fatto sorgere uno degli interrogativi più spinosi sul tema, e cioè se gli attuali mezzi siano sufficienti a regolare efficacemente il fenomeno, o se nel prossimo futuro dovranno porsi limiti più stringenti.

Nel presente capitolo si esamineranno normative e questioni attuali nell'Unione Europea, negli Stati Uniti e in Cina, per poi considerare anche il contesto nazionale. La scelta di concentrarsi, in particolare, sulle prime tre realtà è volta a cogliere non solo la diversità degli approcci a uno stesso fenomeno, ma anche come, al contempo, le preoccupazioni derivanti dallo stesso siano comuni. Si comincerà dall'ordinamento europeo, il più connesso con quello italiano nonché quello (come si vedrà, solo apparentemente) più avanzato dal punto di vista della regolazione.

1.1. Protezione dei dati personali (GDPR e normativa correlata)

Come anticipato, alla base della creazione di un *deepfake* vi sono immagini, video e audio i quali, se ritraggono soggetti realmente esistenti, costituiscono dati personali[176] (o, in determinate circostanze, biometrici), e rientrano di conseguenza nell'ambito di applicazione del GDPR[177]. Da qui sorge la necessità di considerare come primo riferimento normativo il Regolamento (UE) n. 2016/679, noto anche come Regolamento generale sulla protezione dei dati (GDPR), che rappresenta la disciplina europea di carattere generale in tema di protezione dei dati personali.

Per comprendere come i *deepfakes* siano inquadrati nel GDPR occorre partire dal suo art. 4 il quale include, tra le molte nozioni presenti, quella di dati biometrici, definiti come i dati personali «ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica che ne consentono o confermano l'identificazione univoca, quali l'immagine facciale o i dati dattiloscopici»[178]. I dati biometrici si connotano per la loro universalità (sono presenti in ogni individuo), unicità (rendono possibile distinguere una persona dalle altre) e permanenza (non possono essere rimossi nel tempo)[179].

È però fondamentale distinguere il dato biometrico dalla sua fonte. Una fotografia che ritrae un soggetto, ad esempio, rappresenta un semplice dato

[176] È discusso se, nel caso di *deepfakes* non riferiti a persone reali, i dati utilizzati per generarli possano comunque ritenersi personali. Secondo parte della dottrina, ciò sarebbe possibile considerando i dati di *input*, ossia i tratti di individui esistenti utilizzati per addestrare il sistema, e le cui tracce nel prodotto finale potrebbero consentirne una loro re-identificazione. Per approfondimenti sul tema, cfr. F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024, p. 11

[177] F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024, p. 11

[178] Regolamento (UE) 2016/679, art. 4, par. 1, n. 14

[179] G. Ciacci, *Firme grafometriche e tutela dei dati personali*, in Rivista elettronica di Diritto, Economia, Management, n. 1, 2014, pp. 50-63

personale[180]: solo applicando ad essa un particolare trattamento tecnico, e che sia finalizzato all'identificazione o autenticazione della persona (come quello che sfrutta immagini per ricavarne un'impronta del volto), si potrà ottenere un dato biometrico[181]. La distinzione è utile per comprendere che le informazioni utilizzate per generare un *deepfake* sono dati personali, ma non necessariamente anche biometrici[182]: affinchè questo avvenga, dovranno ricorrere i presupposti indicati.

Nel GDPR ricoprono un ruolo centrale le c.d. basi giuridiche del trattamento, ossia le condizioni che garantiscono la liceità del trattamento dei dati e che sono previste in due disposizioni. La prima è l'art. 6, paragrafo 1, riferito ai comuni dati personali, che richiede la sussistenza di almeno un requisito fra i seguenti: a) consenso dell'interessato; b) esecuzione di un contratto cui l'interessato è parte o di misure precontrattuali adottate su sua richiesta; c) obbligo legale cui è soggetto il titolare del trattamento; d) esecuzione di un compito di interesse pubblico o connesso all'esercizio di pubblici poteri; e) interesse vitale dell'interessato o di terzi; f) legittimo interesse del titolare o di terzi[183]. Ne consegue che il trattamento dei dati volto a realizzare un *deepfake* si considera legittimo solo se rientrante in una delle previste basi giuridiche: tra

[180] L'art. 4, par. 1, n.1 del GDPR definisce il dato personale come «qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»)», aggiungendo che «si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale».

[181] B. Saetta, *Dati biometrici*, Protezione dati personali, 2024. Reperibile al sito: https://protezionedatipersonali.it/dati-biometrici

[182] A conferma di ciò, il Considerando 51 del GDPR dispone che «Il trattamento di fotografie non dovrebbe costituire sistematicamente un trattamento di categorie particolari di dati personali, poiché esse rientrano nella definizione di dati biometrici soltanto quando saranno trattate attraverso un dispositivo tecnico specifico che consente l'identificazione univoca o l'autenticazione di una persona fisica».

[183] Regolamento (UE) 2016/679, art. 6, par. 1, lett. a-f

queste, si sostiene che quelle utilizzabili in un simile contesto siano il consenso dell'interessato e il legittimo interesse del titolare[184].

Diversamente, se i dati impiegati per un *deepfake* soddisfano la definizione di dati biometrici, oltre alla base giuridica di cui all'art. 6 si dovrà osservare il disposto dell'art. 9 del GDPR[185], che individua le basi giuridiche per il trattamento dei dati c.d. "particolari" e include in tale categoria proprio i dati biometrici[186]. L'art. 9, paragrafo 1, del GDPR vieta espressamente il trattamento dei dati particolari, salvo il verificarsi di uno dei casi previsti alle lettere a-j del paragrafo 2, corrispondenti appunto alle basi giuridiche: tra queste, ad esempio, il consenso esplicito dell'interessato per una o più finalità specifiche, o la necessità di trattamento per tutelare un interesse vitale dell'interessato (o di altra persona fisica in caso di impossibilità dell'interessato ad acconsentire).

Tuttavia, nel contesto dei *deepfake* le prescrizioni imposte dagli articoli 6 e 9 del GDPR risultano meramente teoriche[187]. Nella prassi, i creatori di *deepfake* non individuano una valida base giuridica, né hanno interesse a farlo, trattandosi, nella maggior parte dei casi, di condotte per natura estranee ai vincoli imposti dalla normativa: si pensi alla necessità di un consenso esplicito dell'interessato per un contenuto che sarebbe certamente lesivo della sua immagine. Lo scenario

---

[184] Cfr. G. Proietti, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del deep fake*, in Media Laws, n. speciale I, 2024. Nell'ipotesi del consenso, l'interessato deve acconsentire attivamente sia al contenuto originario che a quello manipolato. Quanto all'interesse legittimo, occorre invece che il titolare ponga un bilanciamento tra i potenziali vantaggi derivanti dalla creazione di un *deepfake*, e gli eventuali danni ai diritti e libertà degli interessati.

[185] A. Tegho, C. Williams, *The dilemma of the deepfake: Deepfake Inc. and data protection*, Mills & Reeve, 2025. Reperibile al sito: https://www.mills-reeve.com/publications/the-dilemma-of-the-*deepfake*-*deepfake*-inc-and-data-protection/

[186] L'art. 9, paragrafo 1 del GDPR precisa che i dati biometrici rientrano nel novero di quelli particolari solo se «intesi a identificare in modo univoco una persona fisica».

[187] Lo stesso ragionamento può estendersi alle numerose disposizioni che il GDPR ha dedicato agli obblighi del titolare, alle misure di sicurezza da adottare e ai principi previsti agli articoli 5 e 25 (tra i quali spicca quello di *accountability*): difatti, tali adempimenti presuppongono un trattamento che abbia quantomeno l'obiettivo di conformarsi alla normativa, mentre questa logica è assente nella creazione di un *deepfake*. Pertanto questi aspetti, seppur fondamentali nella disciplina del GDPR, non saranno oggetto di approfondimento nel presente scritto.

che si verifica è quello, già anticipato, del *training* non autorizzato (v. *supra*, cap. II, 3.): poiché la sfera giuridica dell'interessato subisce una lesione dal trattamento illecito dei suoi dati personali, quest'ultimo potrà intervenire esercitando i diritti riconosciuti dal GDPR.

In effetti, il Regolamento 2016/679 prevede, agli articoli 13-22, numerosi diritti esercitabili dall'interessato. Tra questi, è opportuno menzionarne alcuni particolarmente rilevanti nel contesto dei *deepfakes*, consentendo una tutela attiva da parte dell'interessato: il diritto di accesso a dati e informazioni (come la finalità o i destinatari del trattamento) che lo riguardano, e di ottenerne rettifica e integrazione dal titolare (articoli 15 e 16); il diritto all'oblio, e cioè alla cancellazione dei dati personali, ammesso che ricorrano i presupposti indicati[188] (art. 17); il diritto di limitazione al trattamento, anch'esso esercitabile solo in presenza di una delle ipotesi previste[189] (art. 18); il diritto di opposizione, azionabile dall'interessato per motivi legati alla sua situazione particolare, e verso trattamenti effettuati ai sensi dell'art. 6, lettere e) o f) (art. 21); infine, il diritto per l'interessato a non essere sottoposto a decisioni basate esclusivamente su trattamenti automatizzati[190], compresa la profilazione, che incidano sulla sua sfera giuridica (art. 22)[191].

Dopo aver esaminato, con riferimento ai *deepfakes*, le basi giuridiche del trattamento e i diritti riconosciuti all'interessato, occorre infine concentrarsi sul

[188] Nel dettaglio, l'art. 17, paragrafo 1, lettere a-f, subordina il diritto alla cancellazione dei dati personali al raggiungimento della finalità, revoca del consenso, opposizione al trattamento, illiceità del trattamento, cancellazione per obbligo di legge o raccolta nell'ambito di servizi della società dell'informazione rivolti a minori.

[189] L'art. 18, paragrafo 1, lettere a-d, riconosce all'interessato il diritto ad ottenere la limitazione del trattamento se: a) ne contesta l'esattezza; b) il trattamento è illecito e, invece della cancellazione, si chiede che l'uso dei dati sia limitato; c) i dati non sono più necessari al titolare ma servono ancora all'interessato per far valere un diritto in sede giudiziaria; d) opposizione al trattamento.

[190] Tuttavia, l'art. 22, paragrafo 2, lettere a-c, indica che il diritto di cui sopra non sussiste se la decisione è necessaria per l'esecuzione di un contratto tra interessato e titolare, autorizzata dalla legge o basata sul consenso dell'interessato.

[191] G. Ciacci, G. Buonomo, *Profili di informatica giuridica*, Wolters Kluwer, Milano, 2021, pp. 198-199

tema della responsabilità, che si articola in relazione a tre tipologie. Anzitutto, la realizzazione di un *deepfake* che comporti un trattamento dei dati personali in violazione del GDPR comporta responsabilità amministrativa. Le relative sanzioni pecuniarie, irrogate dal Garante privacy o dal giudice ordinario, saranno commisurate in relazione al caso concreto, e tenendo conto dei parametri stabiliti dall'art. 83, paragrafo 2, lett. a-k del Regolamento come la natura, gravità e durata della violazione, il suo carattere doloso o colposo, la presenza di misure di sicurezza o di violazioni precedenti. Individuati i parametri, si applicherà una delle due sanzioni previste dai paragrafi 4 e 5 dell'art. 83, effettuando la scelta in base al tipo di violazione commessa: i *deepfakes*, spesso lesivi dei principi del trattamento, come quello di liceità, e dei diritti degli interessati, rientrerebbero nel novero delle violazioni più gravi[192] di cui al paragrafo 5, che prevede una sanzione «fino a 20.000.000 di euro, o per le imprese, fino al 4 % del fatturato mondiale totale annuo dell'esercizio precedente, se superiore»[193]. A queste sanzioni si aggiungono quelle eventualmente previste dagli Stati membri ai sensi dell'art. 84 del GDPR, nonché quelle che il Garante può comminare in ossequio all'art. 58, paragrafo 2, anche in via alternativa[194].

Altra forma di responsabilità è quella civile, che in presenza di un *deepfake* si verifica quando quest'ultimo, generato trattando dati personali illecitamente, cagioni un danno all'interessato. L'art. 82, paragrafo 1 del Regolamento precisa questo concetto, specificando non solo che il danno può essere «materiale o immateriale», ma che «il diritto di ottenere il risarcimento del danno» dal titolare o responsabile del trattamento è in capo a «chiunque», dunque non solo

[192] Invero, la sanzione di cui al paragrafo 4 dell'art. 83 corrisponde a violazioni più che altro di carattere organizzativo, come quelle inerenti gli obblighi del titolare, del responsabile del trattamento o dell'organismo di controllo. Ne deriva un importo decisamente minore: 10 milioni di euro o, per le imprese, fino al 2% del fatturato mondiale totale annuo dell'esercizio precedente, se superiore.

[193] Ai sensi dell'art. 83, paragrafo 5 del GDPR, tale sanzione più elevata si applica se risultano violati: a) i principi di base del trattamento; b) i diritti degli interessati; c) trasferimenti di dati verso paesi terzi o organizzazioni internazionali; d) obblighi previsti dalle legislazioni nazionali; e) ordini o provvedimenti dell'autorità di controllo.

[194] G. Ciacci, G. Buonomo, *Profili di informatica giuridica*, Wolters Kluwer, Milano, 2021, p. 239

all'interessato[195]. La disposizione stabilisce poi un onere di inversione della prova: il titolare o responsabile devono provare che l'evento dannoso non era loro imputabile, mentre il danneggiato è tenuto a dimostrare l'esistenza del danno, la violazione del GDPR e il nesso causale tra i primi due elementi[196]. L'azione di responsabilità ai sensi dell'art. 82 del Regolamento, tuttavia, non è da ritenersi esclusiva in quanto «non pregiudica le azioni di risarcimento di danni derivanti dalla violazione di altre norme del diritto dell'Unione o degli Stati membri»[197].

Da ultimo, vi è la responsabilità penale: il GDPR non presenta una disciplina sul punto, rimettendo agli Stati membri la facoltà di introdurre apposite disposizioni nei propri ordinamenti. Nell'ordinamento italiano il D.lgs. 196/2003, modificato dal D.lgs. 101/2018, elenca in numerosi articoli quali fattispecie danno luogo a sanzioni penali. Tra queste, le più facilmente applicabili in caso di *deepfake* sono: trattamento illecito di dati personali (art. 167); comunicazione e diffusione illecita di dati su larga scala (art. 167-bis); acquisizione fraudolenta di dati su larga scala (art. 167-ter)[198].

In conclusione, è possibile affermare che il Regolamento 2016/679, seppur non concepito per disciplinare i *deepfakes*, costituisce uno strumento utile per contrastarli, permettendo di subordinare il trattamento dei dati personali al rispetto di determinate basi giuridiche, e garantendo strumenti all'interessato per

---

[195] Ivi, p. 241

[196] Tale onere di inversione della prova si giustifica, secondo parte della dottrina, qualificando il trattamento dei dati come esercizio di attività pericolosa. L'art. 15 del D.lgs. 196/2003 richiamava espressamente l'art. 2050 c.c., qualora dal trattamento dei dati personali derivasse un danno. Tuttavia, col D.lgs. 101/2018, la disposizione è stata abrogata, lasciando come unico riferimento in tema di responsabilità civile l'art. 82 del GDPR. Cfr. D. Di Leo, *La responsabilità civile alla luce del GDPR e del decreto 101/2018*, Agenda Digitale, 2019. Reperibile al sito: https://www.agendadigitale.eu/sicurezza/privacy/la-responsabilita-civile-alla-luce-del-gdpr-e-del-decreto-101-2018/

[197] Regolamento (UE) 2016/679, Considerando 146

[198] G. Ciacci, G. Buonomo, *Profili di informatica giuridica*, pp. 239-240. Il D.lgs. 196/2003 include altre tipologie di reati, quali la falsità nelle dichiarazioni e notificazioni al Garante (art. 168); inosservanza dei provvedimenti del Garante (art. 170); e violazione delle disposizioni in materia di controlli a distanza e indagini sulle opinioni dei lavoratori (art. 171).

tutelarsi contro i *deepfakes* illegali. Nonostante ciò, la protezione che il GDPR offre si rivela insufficiente, e questo per più ordini di motivi. Anzitutto, l'anonimato spesso presente in questi contenuti può rendere difficoltoso per l'interessato esercitare i propri diritti. Inoltre, l'inapplicabilità del Regolamento a trattamenti di uso esclusivamente personale consente che *deepfakes* potenzialmente dannosi sfuggano inizialmente a ogni controllo, ma con il rischio che una loro rapida condivisione, attraverso l'uso dei social media, possa produrre danni significativi prima ancora che sia possibile intervenire[199].

Alla luce di quanto detto finora, deve introdursi un concetto che sarà più volte ripreso, fondamentale per il presente elaborato: i *deepfakes*, e più in generale tutte le tecnologie di IA, si evolvono così rapidamente da superare la capacità del legislatore di porre una regolamentazione adeguata.

Nel seguito, dedicato alla regolazione dei *deepfakes* posta dal Digital Services Act (DSA) e AI Act, si andrà meglio a comprendere che sebbene l'Unione Europea, mediante la loro introduzione, abbia compiuto passi in avanti verso una regolamentazione più stringente, l'efficacia del quadro normativo attuale è ancora limitata poiché la riposta delle piattaforme risulta ancora poco reattiva[200].

1.2. Digital Services Act (DSA)

La presenza dei *deepfakes* incide non solo sulla protezione dei dati personali, ma anche sul piano della loro circolazione in rete. Lo strumento normativo che, nell'Unione Europea, costituisce il punto di riferimento in materia di responsabilità e obblighi per le piattaforme online e i fornitori di servizi

[199] M. Nandal, *Mitigating Deepfake Threats to Privacy: Legal Frameworks and Technological Safeguards*, in Proceedings of the National Seminar on Enhancing Privacy Protection in the Digital Age: Legal Challenges & Innovations, 2025, pp. 233-234

[200] L. Di Giacomo, *Deepfake e diritto: l'UE è pronta ad affrontare la minaccia dell'IA?*, Diritto.it, 2024. Reperibile al sito: https://www.diritto.it/*deepfake*-diritto-unione-europea-pronta-minaccia-ia/

digitali è il Regolamento (UE) 2022/2065, noto come Digital Services Act (DSA), che necessita dunque di essere analizzato.

La *ratio* del DSA è di regolamentare il mercato interno dei servizi intermediari[201] e arginare la diffusione di contenuti illegali, al fine di garantire un contesto digitale sicuro, affidabile, innovativo e nel quale siano salvaguardati i diritti fondamentali riconosciuti dalla Carta di Nizza[202]. Per raggiungere tale obiettivo, il Regolamento individua tre elementi essenziali: responsabilità, obblighi di diligenza per i prestatori di servizi digitali (in particolare nei riguardi delle piattaforme online) e cooperazione tra autorità[203].

I *deepfakes*, pur non espressamente nominati nel DSA, sono da esso presi fortemente in considerazione: questa scelta si giustifica non solo per la poca influenza che il termine «*deepfake*» assumeva al momento della sua adozione, ma anche per la volontà del legislatore di non sottoporre le piattaforme a obblighi eccessivamente specifici[204]. A prova di ciò, il Regolamento presenta numerose disposizioni applicabili al fenomeno.

È da considerarsi, in primo luogo, la nozione di «contenuto illegale», indicato nel DSA come «qualsiasi informazione che, di per sé o in relazione a un'attività, tra cui la vendita di prodotti o la prestazione di servizi, non è conforme al diritto dell'Unione o di qualunque Stato membro»[205]. La definizione consente di inquadrare i *deepfakes*, che non sono illeciti automaticamente, in tale

[201] L'art. 3, lettera g del DSA definisce «servizio intermediario» come una ramificazione dei servizi della società dell'informazione, distinguendo tre categorie: i) servizio di semplice trasporto (c.d. mere conduit), che trasmette dati o fornisce l'accesso a reti di comunicazione; ii) servizio di memorizzazione temporanea (c.d. caching), il quale memorizza in modo automatico e temporaneo i dati trasmessi, solo per rendere più efficiente il loro successivo inoltro; iii) servizio di memorizzazione di informazioni (c.d. hosting), che memorizza dati forniti da un utente su richiesta dello stesso.

[202] G. Proietti, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del deep fake*, in Media Laws, n. speciale I, 2024

[203] L. Torchia, *Lo stato digitale. Una introduzione*, Il Mulino, Bologna, 2023, p. 77

[204] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[205] Regolamento (UE) 2022/2065, art. 3, lett. h)

più ampia categoria qualora dalla loro diffusione risulti violato il diritto europeo o nazionale.

Stabilito se un *deepfake* rientri o meno nel novero dei contenuti illegali, occorre poi considerare in quale misura le piattaforme che lo rendono accessibile sono responsabili. Regola generale è che il prestatore del servizio non è responsabile per i contenuti caricati dagli utenti, purché non sia effettivamente a conoscenza della loro illiceità e, nel caso in cui ne diventi consapevole, provveda tempestivamente a rimuovere le informazioni o disabilitare l'accesso[206]. Il quadro è poi completato dalla previsione che solleva i *provider* da obblighi generali di sorveglianza rispetto ai contenuti pubblicati dagli utenti, anche in «presenza di attività illegali»[207].

Negli articoli successivi, il DSA impone ai prestatori di servizi intermediari una serie di obblighi di diligenza i quali, rappresentando al contempo strumenti di tutela per i soggetti lesi dalla circolazione di contenuti potenzialmente dannosi, risultano applicabili anche ai *deepfakes*. Anzitutto, i prestatori devono indicare, nelle condizioni generali, le restrizioni sull'uso del loro servizio, specificando politiche e misure di moderazione dei contenuti (incluso processo decisionale algoritmico e verifica umana), nonché le procedure del sistema interno di gestione dei reclami[208]. I *provider* hanno poi l'obbligo di predisporre i c.d. meccanismi di *notice and action*, che consentono a terzi di segnalare la presenza di contenuti ritenuti illegali[209], e di fornire una motivazione chiara agli utenti per restrizioni dovute a materiali illeciti o non conformi alle condizioni generali[210]. Riguardo a quest'ultima decisione, all'utente deve essere garantito di contestarla tramite la messa a disposizione di un sistema interno di

[206] Regolamento (UE) 2022/2065, art. 6, par. 1. Lo stesso art., al paragrafo 6, fa comunque salva la possibilità, per le autorità giudiziarie o amministrative dei singoli Stati membri, di ordinare al prestatore del servizio di impedire o porre fine alla violazione.

[207] Regolamento (UE) 2022/2065, art. 8

[208] Regolamento (UE) 2022/2065, art. 14

[209] Regolamento (UE) 2022/2065, art. 16

[210] Regolamento (UE) 2022/2065, art. 17

gestione dei reclami[211] o, in alternativa, rivolgendosi a un organismo di risoluzione extragiudiziale delle controversie[212]. Ancora, tra le segnalazioni ricevute, i fornitori di piattaforme online devono dare priorità a quelle presentate dai c.d. segnalatori attendibili[213].

Finora, sono stati esaminati doveri di diligenza che si applicano a tutti gli intermediari. Tuttavia, il DSA adotta una regolazione a più livelli, differenziando gli oneri in base alla tipologia di prestatore coinvolto[214]. Difatti, obblighi ulteriori, anch'essi riferibili ai *deepfakes*, sono imposti ai fornitori di piattaforme *online* e motori di ricerca di dimensioni molto grandi (rispettivamente, VLOP e VLOSE), qualificate come tali dalla Commissione europea, e che devono presentare almeno 45 milioni di utenti mensili[215]. Detti soggetti devono valutare, almeno una volta l'anno, eventuali «rischi sistemici» derivanti dal funzionamento e uso dei loro servizi[216], adottando di conseguenza le relative misure di

[211] Regolamento (UE) 2022/2065, art. 20

[212] Regolamento (UE) 2022/2065, art. 21. La disposizione specifica che è possibile ricorrere a detti organismi anche se il sistema interno di gestione dei reclami ha avuto esito negativo, e che resta in ogni caso impregiudicato il ricorso al giudice.

[213] Regolamento (UE) 2022/2065, art. 22. La figura del segnalatore attendibile, riconosciuta dal coordinatore dei servizi digitali dello Stato membro in cui è presente il richiedente, deve essere in possesso dei requisiti indicati dal paragrafo 2 dell'art. stesso: i) competenza nell'individuare e notificare contenuti illegali; ii) indipendenza da fornitori di piattaforme online; iii) presentazione delle segnalazioni in modo diligente, accurato e obiettivo.

[214] G. Monga, *Responsabilità degli intermediari. Il Digital Services Act*, in M. Maggiore (a cura di), *Il commercio elettronico. Digital markets act, digital services act e altre dimensioni giuridiche*, Giappichelli, Torino, 2024, p. 214

[215] Ad oggi, si registra la presenza di diciassette piattaforme online di dimensioni molto grandi (ad esempio Instagram, Youtube, Amazon o TikTok) e di due motori di ricerca (Google Search e Bing). Per aggiornamenti sul punto, cfr. https://digital-strategy.ec.europa.eu/it/policies/list-designated-vlops-and-vloses

[216] Regolamento (UE) 2022/2065, art. 34. La valutazione di cui al paragrafo 1 deve tener conto dei seguenti rischi: a) diffusione di contenuti illegali; b) effetti negativi per l'esercizio dei diritti fondamentali tutelati dalla Carta di Nizza; c) effetti negativi sul dibattito civico, processi elettorali e sicurezza pubblica; d) effetti negativi su violenza di genere, protezione della salute pubblica, tutela dei minori e benessere fisico e mentale della persona.

attenuazione[217]. Fra queste ultime, la previsione include «il ricorso a un contrassegno ben visibile per fare in modo che un elemento di un'informazione, sia esso un'immagine, un contenuto audio o video, generati o manipolati, che assomigli notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che a una persona appaia falsamente autentico o veritiero, sia distinguibile quando è presentato sulle loro interfacce online»[218]: pur mancando una menzione esplicita, si tratta del riferimento più evidente ai *deepfakes* nel DSA. Un'altra disposizione utile, sempre rivolta ai fornitori di VLOPs e VLOSEs, li incentiva a elaborare codici di condotta volontari contro contenuti illegali e rischi sistemici[219].

Dall'analisi degli obblighi a carico dei prestatori, emerge che il DSA introduce misure utili per contrastare il fenomeno dei *deepfakes*: tuttavia, come visto per il GDPR, il presente Regolamento non è esente da criticità. In primo luogo, il DSA è rivolto ai soli servizi intermediari, con la conseguenza che *deepfakes* prodotti da strumenti di IA generativa (come Chat-GPT) o condivisi in servizi di messaggistica privata (come WhatsApp) restano esclusi dal suo ambito di applicazione, a meno che non siano diffusi all'interno di piattaforme rientranti nella disciplina del Regolamento[220]. Ancora, per quanto riguarda i *deepfake porn*, il DSA impone ai fornitori di VLOPs di intervenire rapidamente e rimuovere senza ritardo i materiali «che costituiscono violenza online», come quelli pornografici illegali o i contenuti manipolati[221]: il semplice contenimento *ex post* non è però sufficiente a tutelare la dignità della vittima, dovendosi invece introdurre strumenti di natura preventiva[222]. Infine, nonostante l'assenza di un obbligo generale di sorveglianza per i *provider*, essi devono intervenire sui

[217] Regolamento (UE) 2022/2065, art. 35

[218] Regolamento (UE) 2022/2065, art. 35, par. 1, lett. k)

[219] Regolamento (UE) 2022/2065, art. 40

[220] *Deepfake Regulation: Chaos That Matters Now More Than Ever*, Whisperly, 2025. Reperibile al sito: https://whisperly.ai/*deepfake*-regulation/

[221] Regolamento (UE) 2022/2065, Considerando n. 87

[222] S. Puglisi, C. Vergari, *Il fenomeno del deepfake nell'attuale ordinamento giuridico*, Trento Biolaw Selected Student Papers, T4F Series, n. 1, 2023, p. 12

contenuti segnalati come illeciti tramite lo strumento di *notice and action*: questa situazione genera una zona grigia che permette ai creatori di *deepfakes* di approfittarsene senza che vi sia un controllo immediato, eludendo le misure di moderazione previste dal Regolamento[223].

In definitiva, la disciplina del DSA, in relazione ai *deepfakes*, si rivela utile ma limitata alla loro gestione e segnalazione sulle piattaforme. È però possibile ottenere una forma di regolazione più completa leggendo il DSA in combinazione con l'AI Act, il quale pone invece l'attenzione «prima di tutto sullo strumento tecnologico, oltreché sul loro utilizzo»[224], mostrando come i due strumenti normativi operino in sinergia nel mitigare i rischi connessi ai *deepfakes*. Posta tale premessa, è necessario proseguire con l'AI Act, la cui portata innovativa sarà approfondita nel paragrafo a seguire.

1.3. AI Act

Fino a questo punto, i *deepfakes* sono stati inquadrati nell'ottica di GDPR e DSA, fonti normative le quali, oltre a non farne riferimento diretto, incidono su di essi solo indirettamente: la prima tramite la protezione dei dati personali, la seconda mediante obblighi di moderazione e trasparenza imposti alle piattaforme. In questo quadro si inserisce il più recente Regolamento (UE) 2024/1689 (noto come AI Act), che per la prima volta introduce una regolamentazione diretta dei *deepfakes*.

Entrato in vigore il 1° agosto 2024, l'AI Act assumerà piena operatività nel 2027. Il suo obiettivo è istituire un quadro giuridico armonizzato, all'interno dell'Unione Europea, in materia di intelligenza artificiale, disciplinando

[223] L. Di Giacomo, *Deepfake e diritto: l'UE è pronta ad affrontare la minaccia dell'IA?*, Diritto.it, 2024. Reperibile al sito: https://www.diritto.it/*deepfake*-diritto-unione-europea-pronta-minaccia-ia/

[224] G. Proietti, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del deep fake*, in Media Laws, n. speciale I, 2024

l'importazione, la produzione e l'uso dei relativi sistemi[225]. Il Regolamento si applica ai numerosi soggetti appartenenti alla c.d. «catena del valore» dell'IA, tra cui fornitori (o *provider*), utilizzatori (o *deployer*), importatori, distributori, produttori e rappresentanti autorizzati. La disciplina si applica anche ai *provider* e *deployer* stabiliti fuori dall'UE, se l'*output* del sistema di IA è utilizzato all'interno dell'Unione[226].

Una delle peculiarità del Regolamento è quella di classificare i sistemi di IA secondo diversi livelli di rischio: "inaccettabile", quando comportano pratiche vietate in ogni caso[227], "alto"[228], "limitato" ad alcune applicazioni specifiche, oppure "minimo"[229]. Ciò posto, occorre comprendere se, e in quale fra queste categorie, siano inclusi i *deepfakes*. L'AI Act li definisce come «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona»[230]. Tuttavia, nelle disposizioni dedicate alle pratiche vietate e ai sistemi ad alto rischio, i *deepfakes* non risultano essere inclusi, sebbene possano esservi ricondotti implicitamente[231].

---

[225] G. Buonomo, G. Ciacci, A. Costanzo, *Macchine intelligenti e diritto*, Giappichelli, Torino, 2025, p. 401

[226] M. Kosinski, M. Scapicchio, *What is the EU AI Act?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/eu-ai-act. La disciplina si applica anche ai *provider* e *deployer* stabiliti fuori dall'UE, se l'*output* del sistema di IA è utilizzato all'interno dell'Unione.

[227] L'art. 5 del Regolamento elenca le pratiche di IA vietate, come la rilevazione delle emozioni delle persone nell'ambiente di lavoro, o il c.d. *social scoring*, ossia la classificazione delle persone fisiche o di gruppi di persone per un determinato periodo di tempo sulla base del loro comportamento sociale.

[228] Ai sensi dell'art. 6, par. 1 dell'AI Act, un sistema è considerato ad alto rischio se è utilizzato come componente di sicurezza di un prodotto, ed è soggetto a valutazione di conformità da terzi per essere inserito nel mercato. Il secondo paragrafo della disposizione rimanda all'Allegato III per l'individuazione di ulteriori sistemi ad alto rischio, ad esempio i sistemi di biometria o quelli finalizzati a valutare l'accesso a servizi essenziali.

[229] G. Buonomo, G. Ciacci, A. Costanzo, *Macchine intelligenti e diritto*, Giappichelli, Torino, 2025, p. 401

[230] Regolamento (UE) 2024/1689, art. 3, n. 60

[231] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

In effetti, l'art. 5 dell'AI Act, elencando le pratiche di IA vietate, ne rileva due nelle quali i *deepfakes* possono rientrare: da un lato, le tecniche subliminali, manipolative o ingannevoli che operano senza consapevolezza della persona, finalizzate a indurre uno o più soggetti ad assumere decisioni che non avrebbero altrimenti preso[232]; dall'altro, lo sfruttamento delle vulnerabilità di individui legati a età, disabilità o situazioni sociali ed economiche, in modo da distorcere il loro comportamento[233]. In entrambe le previsioni, si specifica che il divieto sussiste solo in caso di un «danno significativo» per le persone[234]. A livello pratico, alcuni autori ritengono che i *deepfakes*, capaci di estrarre dati da *social media* per generare contenuti falsi di individui ignari, possano rientrare nelle pratiche di IA vietate di cui all'art. 5, par. 1, lett. a) dell'AI Act[235].

Per quanto riguarda i sistemi ad alto rischio, si è già indicato (v. *supra*, nota 227) che l'art. 6, par. 2 del Regolamento rinvia la loro individuazione all'Allegato III. I *deepfakes*, pur non figurando espressamente in tale elenco, potrebbero rientrarvi se utilizzati in «settori sensibili» come l'istruzione, i servizi essenziali, l'attività di contrasto ai crimini o i processi democratici[236]: è quest'ultimo a offrire la formulazione più interessante, qualificando ad alto rischio i sistemi volti a influenzare l'esito di elezioni, di un referendum o il comportamento di voto degli elettori[237]. Affinchè un *deepfake* ricada nella disciplina dei sistemi ad alto rischio, vi è però un altro requisito da soddisfare: fermo restando il rispetto delle condizioni di cui all'art. 6, par. 3, deve esservi presenza di «un rischio significativo di danno per la salute, la sicurezza o i diritti

[232] Regolamento (UE) 2024/1689, art. 5, par. 1, lett. a)

[233] Regolamento (UE) 2024/1689, art. 5, par. 1, lett. b)

[234] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[235] F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024, p. 7

[236] *Ibidem*

[237] Regolamento (UE) 2024/1689, Allegato III, n. 8, lett. b)

fondamentali delle persone fisiche», anche nel senso di «influenzare materialmente il risultato del processo decisionale»[238].

A questo punto, è necessario comprendere quale sia la precisa collocazione dei *deepfakes* all'interno dell'AI Act. Parte della dottrina sostiene che il legislatore abbia voluto distinguerli dai sistemi a basso rischio, per i quali nessun obbligo è previsto, scegliendo di attribuire loro una posizione intermedia tra il rischio alto e quello minimo[239] attraverso l'art. 50. La disposizione impone obblighi di trasparenza a fornitori e *deployer* «di determinati sistemi di IA», comprendendovi i sistemi «destinati a interagire direttamente con le persone fisiche»[240] e «di riconoscimento delle emozioni» o «di categorizzazione biometrica»[241].

Quanto ai *deepfakes*, la rilevanza dell'art. 50 emerge in due passaggi. Anzitutto, si prescrive ai fornitori di sistemi di IA che producono «contenuti audio, immagine, video o testuali sintetici» di garantire che gli *output* siano contrassegnati e identificabili come «generati o manipolati artificialmente», mediante soluzioni tecniche efficaci, solide e affidabili[242]. Più avanti, rivolgendosi ai *deployer* di sistemi che generano o manipolano «immagini o contenuti audio o video che costituiscono un *deepfake*», il Regolamento pone un altro riferimento esplicito, stabilendo che gli utilizzatori devono renderne nota la generazione o manipolazione artificiale, salvo che il loro utilizzo non sia destinato per legge all'accertamento, prevenzione e repressione di reati[243]. Inoltre, se i *deepfakes* sono inseriti in opere o programmi «manifestamente

---

[238] Regolamento (UE) 2024/1689, art. 6, par. 3

[239] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[240] Regolamento (UE) 2024/1689, art. 50, par. 1

[241] Regolamento (UE) 2024/1689, art. 50, par. 3

[242] Regolamento (UE) 2024/1689, art. 50, par. 2. Il presente obbligo tiene conto «dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto», non applicandosi se il sistema di IA è limitato a funzioni di *editing* standard, non altera in modo sostanziale i dati di *input* forniti dal *deployer* o se autorizzato per legge «ad accertare, prevenire, indagare o perseguire reati».

[243] Regolamento (UE) 2024/1689, art. 50, par. 4

artistici, creativi, satirici o fittizi»[244], l'obbligo di trasparenza si riduce a una mera segnalazione della loro esistenza, da effettuare «in modo adeguato» e senza compromettere la fruizione dell'opera[245].

Alla luce di quanto detto, risulta evidente che la difficoltà nel collocare i *deepfakes* in un preciso livello di rischio, provata anche dalle divergenze dottrinali[246], mette in luce la scarsa chiarezza della disciplina[247]: una simile problematica consente di passare al tema ulteriore delle criticità del Regolamento.

In primo luogo, l'incertezza dell'AI Act si riflette sull'applicazione delle sanzioni: accogliendo la teoria che ingloba i *deepfakes* nelle pratiche vietate ex. art. 5, la violazione dei doveri di trasparenza comporterebbe l'applicazione delle sanzioni più gravi, pari a 35 milioni di euro o, per le imprese, fino al 7 % del fatturato mondiale totale annuo dell'esercizio precedente se superiore[248]. Diversamente, un *deepfake* ritenuto non conforme a una qualunque disposizione diversa dall'art. 5 sarà soggetto a sanzioni più lievi, fino a 15 milioni di euro o,

[244] Parte della dottrina ritiene che, utilizzando questi aggettivi, il legislatore contempli la possibilità di includere nella disposizione anche i contenuti pubblicati da utenti sui *social media*. Da una parte, l'aggettivo «fittizie» potrebbe avvalorare la tesi, e addirittura consentire un forte ampliamento delle eccezioni. Dall'altra, l'avverbio «manifestamente» potrebbe limitare tale espansione. Cfr. A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[245] Regolamento (UE) 2024/1689, art. 50, par. 4. Tale limitazione all'obbligo di trasparenza presenta forti ambiguità, poiché non appare chiara né l'effettiva distinzione con l'obbligo generale previsto per i *deepfakes*, né secondo quali criteri la segnalazione possa considerarsi "adeguata". Cfr. A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[246] Sul punto, cfr. M. Łabuz, *Regulating Deep Fakes in the Artificial Intelligence Act*, in Applied Cybersecurity & Internet Governance, vol. 2, n. 1, 2023, p. 265, nel quale è messo in luce, fra le molte teorie dottrinali, l'originale approccio alla valutazione del rischio, che non dovrebbe tener conto della tecnologia in sé, ma del contesto e utilizzo pratico, talvolta anche suscettibili di effetti positivi.

[247] Ciò è provato dalla presenza di contributi che già invocano interventi correttivi, nonostante la recente entrata in vigore dell'AI Act. Cfr. A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024.

[248] Regolamento (UE) 2024/1689, art. 99, par. 3

se l'autore del reato è un'impresa, fino al 3 % del fatturato mondiale totale annuo dell'esercizio precedente se superiore[249].

Oltre il problema della corretta individuazione della sanzione, si pone anche quello della loro applicazione in ambito extraterritoriale. Si è già detto che l'AI Act può applicarsi anche per fornitori e *deployers* stabiliti fuori dall'UE, se i sistemi sono comunque messi in servizio al suo interno: tuttavia, risulta complesso applicare concretamente le sanzioni ai casi transfrontalieri, tant'è che si auspica la presenza di un «diritto penale armonizzato», da realizzare mediante la «cooperazione con le autorità dei paesi terzi e i trattati di mutua assistenza giudiziaria»[250].

Ulteriore considerazione riguarda l'art. 53, che integra il regime di trasparenza prevedendo molteplici obblighi per i fornitori di «modelli di IA per finalità generali»[251], tra cui la redazione e aggiornamento della documentazione tecnica del modello (inclusi processo di addestramento e relative valutazioni); la messa a disposizione delle informazioni per i fornitori di IA che intendono integrare tali modelli nei loro sistemi; e la trasmissione al pubblico di una sintesi dei contenuti utilizzati per l'addestramento del modello[252]. La disposizione è controversa in quanto, nel secondo paragrafo, esonera dai doveri appena citati i «fornitori di modelli di IA rilasciati con licenza libera e *open source*»[253]: in tal modo, alcune delle piattaforme che permettono la creazione di *deepfakes* non solo

[249] Regolamento (UE) 2024/1689, art. 99, par. 4

[250] M. Fragale, V. Grilli, *Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation*, Columbia Journal of European Law, 2024. Reperibile al sito: https://cjel.law.columbia.edu/preliminary-reference/2024/*deepfake*-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/?cn-reloaded=1

[251] L'art. 3, n. 63 del Regolamento definisce il «modello di IA per finalità generali» come quel modello di IA capace di «svolgere con competenza un'ampia gamma di compiti distinti», a prescindere dalle modalità con cui è immesso nel mercato, e che «può essere integrato in una varietà di sistemi o applicazioni a valle». Si escludono da tale definizione i modelli utilizzati per attività di ricerca, sviluppo o prototipazione.

[252] Regolamento (UE) 2024/1689, art. 50, par. 1, lett. a), b), d)

[253] Regolamento (UE) 2024/1689, art. 50, par. 2

operano liberamente[254], ma possono anche godere di un vantaggio competitivo ingiustificato rispetto a quelle imprese che osservano la disciplina[255].

Infine, persistono dubbi sul fatto che l'attuale disciplina dell'AI Act possa garantire una tutela effettiva. A ben vedere, l'attuale disciplina non chiarisce la responsabilità di *provider* e *deployer*, privilegiando doveri di trasparenza i quali, pur potendo giocare un ruolo nella limitazione dei *deepfakes*, non riescono a dissuadere i soggetti dal diffonderli[256].

In definitiva l'AI Act, nonostante il merito di aver introdotto per la prima volta una diretta disciplina dei *deepfakes*, rivela ancora profonde lacune nella capacità di mitigare pienamente il fenomeno: a tal proposito, sono già state avanzate delle proposte per rendere il Regolamento più funzionale. Anzitutto, come già accennato, la disciplina necessita di integrazioni dal punto di vista degli strumenti di protezione *ex-post*, volti a semplificare la rimozione dei *deepfakes* dannosi[257]. Si è poi ipotizzato di intervenire sul Codice Etico dell'Intelligenza Artificiale del 2019, rendendo obbligatori meccanismi di rilevamento e segnalazione dei contenuti falsi, il che richiederebbe un costante aggiornamento dei relativi sistemi e una cooperazione stabile tra governi e imprese

---

[254] Tuttavia, lo stesso paragrafo chiarisce che l'esenzione non si applica per i «modelli di IA per finalità generali con rischi sistemici». Inoltre, il Considerando 103 del Regolamento stabilisce che i componenti di IA liberi e *open source*, i quali in realtà siano «forniti a pagamento o altrimenti monetizzati», non beneficiano delle eccezioni previste dall'art. 50, par. 2 dell'AI Act.

[255] F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024, p. 21

[256] Cfr. M. Łabuz, *Regulating Deep Fakes in the Artificial Intelligence Act*, in Applied Cybersecurity & Internet Governance, vol. 2, n. 1, 2023, p. 265; C. Vanberghen, *The AI Act vs. deepfakes: A step forward, but is it enough?*, Euractiv, 2024. Reperibile al sito: https://www.euractiv.com/opinion/the-ai-act-vs-*deepfakes*-a-step-forward-but-is-it-enough/, nel quale si ribadisce che l'AI Act privilegia strumenti preventivi piuttosto che sanzionatori; F. R. Moreno, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024, p. 24, in cui emerge la necessità di garanzie più solide sia per utenti che per fornitori, poiché solo in tal modo il Regolamento potrà «regolamentare efficacemente i *deepfakes* senza diventare un'arma contro le stesse libertà che cerca di salvaguardare».

[257] E. Candotto, *Deepfake pornografici: le tutele previste dall'AI Act*, LexTech Hub, 2025. Reperibile al sito: https://www.lextech-hub.com/post/*deepfake*-pornografici-le-tutele-previste-dall-ai-act

tecnologiche[258]. Ancora, sul presupposto che il Regolamento, attraverso i soli obblighi di trasparenza, consideri i *deepfakes* non una minaccia a sé stante ma un mero «effetto collaterale» dell'IA, si suggerisce di riclassificarli nella categoria dei sistemi ad alto rischio, rimettendo la disciplina della loro dimensione mediatica ai singoli Stati membri[259].

Esaminati GDPR, DSA e AI Act, si può pervenire a un'importante osservazione per il presente elaborato. Se è vero che le tre fonti normative, da un lato, colmano le proprie insufficienze reciprocamente, dall'altro non riescono comunque ad assicurare una protezione totalmente efficace: appare dunque pacifico che, in futuro, saranno necessari interventi correttivi (o persino nuovi strumenti normativi) in grado di regolare il fenomeno.

1.4. Diritto d'autore e diritto all'immagine nel contesto europeo

A completamento del discorso sulla regolamentazione dei *deepfakes* a livello europeo, è doveroso inserire un ulteriore tassello. È già stato ribadito che GDPR, DSA e AI Act offrono una tutela, seppur parziale, di aspetti che tali tecnologie rischiano di compromettere, come la protezione dei dati personali, la corretta circolazione all'interno delle piattaforme e la trasparenza nell'utilizzo dei sistemi di IA. Tuttavia, nessuna di queste fonti copre le ipotesi in cui un *deepfake* faccia utilizzo di opere preesistenti o falsifichi l'identità visiva o vocale di una persona senza il suo consenso.

Alla luce di tali considerazioni, si procederà a introdurre solo alcuni dei riferimenti normativi che l'Unione Europea dedica al diritto d'autore e all'immagine, selezionati tra quelli maggiormente riconducibili ai *deepfakes*.

Occorre cominciare dal diritto d'autore, la cui disciplina, in ambito comunitario, si fonda su due principali obiettivi: l'armonizzazione delle

[258] L. Di Giacomo, *Deepfake e diritto: l'UE è pronta ad affrontare la minaccia dell'IA?*, Diritto.it, 2024. Reperibile al sito: https://www.diritto.it/*deepfake*-diritto-unione-europea-pronta-minaccia-ia/

[259] M. Łabuz, *Regulating Deep Fakes in the Artificial Intelligence Act*, in Applied Cybersecurity & Internet Governance, vol. 2, n. 1, 2023, p. 267

disposizioni degli Stati membri, volta a rimuovere gli ostacoli alla libera circolazione di contenuti e servizi protetti dal *copyright*; e l'ammodernamento delle leggi nazionali e dell'Unione, orientato a migliorarne chiarezza e qualità[260]. In materia di diritto d'autore, la prima direttiva a perseguire tali finalità è stata la 91/250/CEE, relativa alla tutela dei programmi per elaboratore, mentre oggi il quadro legislativo europeo si compone di 13 direttive e due regolamenti[261].

Fra queste, particolare importanza assume la Direttiva 2001/29/CE (nota anche come Direttiva "Infosoc") sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione: nel dettaglio, il suo obiettivo è quello di creare un mercato unico dei contenuti digitali[262] ed estendere la tutela di opere e prestazioni protette ai nuovi strumenti tecnologici[263].

Considerando l'anno della sua adozione, è evidente che la presente Direttiva mai avrebbe potuto contemplare un fenomeno come i *deepfakes*. Nonostante ciò, la sua formulazione ampia e tecnologicamente neutra consente di estenderne l'applicazione anche a simili forme di manipolazione digitale: si pensi alle disposizioni che attribuiscono agli autori il diritto esclusivo di autorizzare o vietare la riproduzione[264], la comunicazione al pubblico[265] o la distribuzione[266] delle loro opere.

In relazione ai *deepfakes*, sono poi rilevanti alcune delle eccezioni e limitazioni che gli Stati membri possono disporre a tali diritti esclusivi, come nel

---

[260] M. van Eechoud, P. B. Hugenholtz, S. van Gompel, L. Guibault, N. Helberger, *Harmonizing European Copyright Law: the Challenges of Better Lawmaking*, Wolters Kluwer, Alphen aan den Rijn, 2009, p. 23

[261] E. Rosati, *Copyright in the Digital Single Market: a taster*, WIPO, 2021. Reperibile al sito: https://www.wipo.int/web/wipo-magazine/articles/copyright-in-the-digital-single-market-a-taster-42399

[262] L'obiettivo è ribadito dal Considerando 2 della Direttiva, che subordina «lo sviluppo della società dell'informazione in Europa» alla presenza di un «mercato interno dei nuovi prodotti e servizi».

[263] A prova di ciò, il Considerando 25 ritiene necessario prevedere una tutela uniforme anche nel caso in cui un'opera protetta sia trasmessa «su rete».

[264] Direttiva 2001/29/CE, art. 2. La disposizione specifica che la riproduzione può essere «temporanea o permanente», «diretta o indiretta», «in qualunque modo o forma» e «totale o parziale».

[265] Direttiva 2001/29/CE, art. 3

[266] Direttiva 2001/29/CE, art. 4

caso in cui l'utilizzo di un'opera avvenga a fini didattici o di ricerca scientifica[267], di critica o rassegna (purché «relativo a un'opera o altri materiali protetti già messi legalmente a disposizione del pubblico»)[268], o ancora a scopo di «caricatura, parodia o pastiche»[269]. In ogni caso, dette eccezioni e limitazioni si applicano solo in «determinati casi speciali», che non contrastino con «lo sfruttamento normale dell'opera» né con gli «interessi legittimi del titolare»[270].

A seguito della forte evoluzione tecnologica, successiva alla Direttiva 2001/29/CE, si è reso necessario intervenire con la Direttiva (UE) 2019/790 (nota come Direttiva "Copyright"), che non abroga le precedenti direttive, ma le integra ed aggiorna. La sua finalità principale è «armonizzare ulteriormente il quadro giuridico dell'Unione applicabile al diritto d'autore e ai diritti connessi nell'ambito del mercato interno», con particolare attenzione agli «utilizzi digitali e transfrontalieri» delle opere protette[271].

Anche in tal caso, pur non menzionando mai i *deepfakes*, la Direttiva include alcune disposizioni ad essi potenzialmente applicabili.

Anzitutto, è imposto agli Stati membri di introdurre un'eccezione ai diritti di riproduzione e di estrazione, consentendo a organismi di ricerca e istituti di tutela del patrimonio culturale di effettuare operazioni di estrazione di testo e dati (c.d. "*text and data mining*"[272]) da opere protette, purché a fini di ricerca scientifica e a condizione di avere legalmente accesso ai materiali[273]. L'eccezione

[267] Direttiva 2001/29/CE, art. 5, par. 3, lett. a)

[268] Direttiva 2001/29/CE, art. 5, par. 3, lett. d)

[269] Direttiva 2001/29/CE, art. 5, par. 3, lett. k)

[270] Direttiva 2001/29/CE, art. 5, par. 5

[271] Direttiva (UE) 2019/790, art. 1. La disposizione conclude affermando di tenere conto anche delle eccezioni e limitazioni al diritto d'autore, ottenimento delle licenze e buon funzionamento del mercato interno.

[272] L'art. 2, punto 2 della Direttiva definisce l'estrazione di testo e dati come «qualsiasi tecnica di analisi automatizzata volta ad analizzare testi e dati in formato digitale avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni».

[273] Direttiva (UE) 2019/790, art. 3

può estendersi anche a soggetti diversi da quelli menzionati, a meno che l'utilizzo delle opere non sia stato espressamente riservato dal titolare dei diritti[274].

Un altro aspetto interessante, disciplinato dalla Direttiva 2019/790, è l'utilizzo di contenuti protetti da parte di prestatori di servizi di condivisione di contenuti online[275], come i celebri YouTube, Instagram o TikTok. Quando tali soggetti concedono l'accesso pubblico a contenuti protetti, devono fornire allo stesso un atto di comunicazione e ottenere un'autorizzazione dai titolari dei diritti esclusivi, ad esempio tramite la conclusione di un accordo di licenza[276]. In assenza di autorizzazione, i prestatori sono responsabili per gli atti di comunicazione al pubblico non autorizzati, salvo che non dimostrino di aver adottato ogni possibile misura per ottenere un'autorizzazione, assicurare che non siano disponibili opere o materiali segnalati come protetti dai titolari e, in caso di segnalazione, aver agito tempestivamente per la loro rimozione e disabilitazione dell'accesso[277].

Passando ora al diritto all'immagine, deve rilevarsi che, nell'ordinamento europeo, non sono attualmente presenti fonti o disposizioni che ne riconoscano espressamente la tutela.

Un importante contributo in tal senso è però offerto dalla giurisprudenza della Corte Europea dei Diritti dell'Uomo, che ha progressivamente incluso il diritto all'immagine nella più ampia sfera di tutela del diritto al rispetto della vita privata e familiare, di cui all'art. 8 CEDU. In particolare, la Corte di Strasburgo ha affermato che il concetto di vita privata «si estende agli aspetti relativi

[274] Direttiva (UE) 2019/790, art. 4

[275] L'art. 2, punto 6 della Direttiva qualifica il «prestatore di servizi di condivisione di contenuti online» un «prestatore di servizi della società dell'informazione il cui scopo principale o uno dei principali scopi è quello di memorizzare e dare accesso al pubblico a grandi quantità di opere protette dal diritto d'autore o altri materiali protetti caricati dai suoi utenti, che il servizio organizza e promuove a scopo di lucro», esentando da tale categoria numerosi soggetti, ad esempio le enciclopedie online senza scopo di lucro, i mercati online o i servizi *cloud* da impresa a impresa.

[276] Direttiva (UE) 2019/790, art. 17, par. 1. Il successivo art. 18 garantisce il diritto, per autori e artisti che concedono i loro diritti esclusivi in licenza, di ricevere una «remunerazione adeguata e proporzionata».

[277] Direttiva (UE) 2019/790, art. 17, par. 4

all'identità personale», inclusi «l'integrità fisica e psicologica della persona»[278], aggiungendo poi che l'art. 8 CEDU non si limita a imporre allo Stato un obbligo di non ingerenza nei confronti dell'individuo, ma richiede anche «obblighi positivi» tra cui «misure volte a garantire il rispetto della vita privata anche nell'ambito delle relazioni degli individui tra loro»[279]. Sulla base di tali premesse, la Corte EDU conclude che, nel caso di specie, «la pubblicazione da parte di diverse riviste tedesche di foto della ricorrente nella sua vita quotidiana, sia da sola che in compagnia di altre persone, rientri nell'ambito della sua vita privata»[280].

Nella successiva sentenza del 2012, anch'essa originata dalla pubblicazione di ulteriori fotografie della vita privata della ricorrente, la Corte di Strasburgo ribadisce che la pubblicazione di una foto equivale a intromissione nella vita privata di una persona[281], poiché «l'immagine di una persona costituisce uno degli attributi principali della sua personalità»[282]. Più avanti nella decisione, sono introdotte due considerazioni innovative. Da una parte, il confine tra gli obblighi positivi e negativi dello Stato ai sensi dell'art. 8 CEDU non è preciso, dovendosi raggiungere il «giusto equilibrio» tra gli «interessi concorrenti rilevanti»[283]. Dall'altra, la Corte EDU rende necessario un bilanciamento tra il rispetto della vita privata e il diritto alla libertà di espressione di cui all'art. 10 CEDU[284], individuando i relativi criteri da considerare: contributo a un dibattito di interesse generale; notorietà della persona interessata e oggetto della segnalazione; condotta dell'interessato nei confronti dei media; contenuto, forma

[278] Corte EDU, sent. 24 giugno 2004, ric. 59320/00, *Von Hannover v. Germany*, par. 50

[279] Corte EDU, sent. 24 giugno 2004, ric. 59320/00, *Von Hannover v. Germany*, par. 57

[280] Corte EDU, sent. 24 giugno 2004, ric. 59320/00, *Von Hannover v. Germany*, par. 53

[281] Corte EDU, sent. 7 febbraio 2012, ric. 40660/08 e 60641/08, *Von Hannover v. Germany (No. 2)*, par. 95

[282] Corte EDU, sent. 7 febbraio 2012, ric. 40660/08 e 60641/08, *Von Hannover v. Germany (No. 2)*, par. 96

[283] Corte EDU, sent. 7 febbraio 2012, ric. 40660/08 e 60641/08, *Von Hannover v. Germany (No. 2)*, par. 99

[284] Corte EDU, sent. 7 febbraio 2012, ric. 40660/08 e 60641/08, *Von Hannover v. Germany (No. 2)*, par. 100

e conseguenze della segnalazione; circostanze in cui sono state ottenute le fotografie[285].

In definitiva, le sentenze esaminate mostrano un nesso evidente con i *deepfakes* i quali, costituendo manipolazioni digitali dell'immagine, alterano in modo ancor più incisivo l'identità e, di riflesso, la vita privata del soggetto raffigurato. Tuttavia, diversamente dal diritto d'autore, è ancora assente a livello comunitario una disciplina *ad hoc* sul diritto all'immagine, che ad oggi può considerarsi più una proiezione dei diritti umani che un diritto autonomo e distinto.

In altre parole, nel quadro giuridico dell'UE è auspicabile un istituto specifico in materia, come avviene negli Stati Uniti con il c.d. *right of publicity*. Concluso il discorso sulla regolamentazione dei *deepfakes* in Unione Europea, si può ora spostare l'attenzione proprio al contesto statunitense.

2. Stati Uniti

La regolamentazione dell'IA nel sistema statunitense si differenzia nettamente da quella dell'Unione Europea. Ciò è dovuto non solo all'assenza, negli Stati Uniti, di una normativa uniforme in materia, ma anche alla loro struttura federale, che consente ai singoli stati di intervenire autonomamente per colmare i vuoti legislativi[286].

Questo scenario si riflette inevitabilmente sui *deepfakes* per i quali, nonostante il grande passo in avanti compiuto con il *Take It Down Act* (di cui si parlerà a breve), manca ancora una disciplina capace di affrontare globalmente il fenomeno, regolando tutte le sue implicazioni.

[285] Corte EDU, sent. 7 febbraio 2012, ric. 40660/08 e 60641/08, *Von Hannover v. Germany (No. 2)*, par. 109-113

[286] S. Parinardi, J. Crosson, K. Peterson, S. Nadarevic, *Investigating the politics and content of US State artificial intelligence legislation*, in Business and Politics, n. 2, 2024, pp. 240-262

Tuttavia, come nell'ordinamento dell'UE, anche per gli Stati Uniti è possibile estendere ai *deepfakes* l'applicazione di disposizioni e istituti preesistenti, che si andranno ora ad approfondire.

2.1. Copyright Act

Il primo riferimento normativo da considerare è il *Copyright Act* del 1976: incluso nel Titolo 17 dello United States Code, esso costituisce, assieme ai successivi emendamenti come il *Semiconductor Chip Protection Act* e il *Vessel Hull Design Protection Act*, la principale fonte statunitense in materia di diritto d'autore[287]. Negli Stati Uniti, il quadro sul diritto d'autore è poi completato dallo U.S. Copyright Office, ente governativo che amministra le leggi nazionali sul *copyright*, svolge funzioni di consulenza[288] e redige il *Compendium of U.S. Copyright Office Practices*, manuale interno che illustra come l'Ufficio interpreta e applica il *Copyright Act*, in particolare nelle procedure di registrazione, deposito e annotazione delle opere[289].

In questa sede, l'operato dell'Ufficio interessa poiché offre numerose indicazioni applicabili ai *deepfakes*.

Anzitutto, tramite il manuale interno, è richiamato il principio-cardine della paternità umana dell'opera: il diritto d'autore statunitense riconosce come opere protette solo i «frutti del lavoro intellettuale» umano, precisando che, diversamente, l'Ufficio ne rifiuterà la registrazione[290]. Conseguenza logica è che le opere generate autonomamente da sistemi di IA, come avviene per i *deepfakes*,

[287] *Copyright Law of the United States (Title 17) and Related Laws Contained in Title 17 of the United States Code*, U.S. Copyright Office, 2025. Reperibile al sito: https://www.copyright.gov/title17/

[288] *Overview of the Copyright Office*, U.S. Copyright Office, 2025. Reperibile al sito: https://www.copyright.gov/about/

[289] *Introduction to the Third Edition of the Compendium Of U.S. Copyright Office Practices*, U.S. Copyright Office, 2021. Reperibile al sito: https://www.copyright.gov/comp3/docs/introduction.pdf

[290] Compendium of U.S. Copyright Office Practices, Third Edition, sez. 306

saranno escluse dalla tutela prevista dal *Copyright Act* proprio in quanto carenti della creatività umana.

In secondo luogo, il *Copyright Office*, in risposta ai notevoli progressi delle tecnologie di intelligenza artificiale, ha promosso un'iniziativa volta ad affrontare le questioni che esse sollevano nel campo del diritto d'autore, pubblicando contestualmente nuove linee-guida per la registrazione di opere che includono contenuti generati dall'IA[291]. Queste ultime, oltre a ribadire che il diritto d'autore «esclude i non umani»[292], forniscono istruzioni in caso di opere "ibride", ossia quelle caratterizzate da apporti sia umani che artificiali. I richiedenti devono compilare un'apposita domanda di registrazione in cui occorre descrivere la paternità umana, distinguendo gli elementi frutto della propria attività creativa da quelli generati dall'IA[293]. Ne consegue che l'autore può rivendicare il *copyright* esclusivamente per le parti dell'opera alle quali egli ha effettivamente contribuito, con esclusione di quelle create dall'IA che eccedano la soglia minima di creatività umana[294].

Appare evidente che le restrizioni imposte dalle linee-guida si estendono anche ai *deepfakes*, i quali non sono altro che un prodotto dell'IA, ma vi è un'ulteriore considerazione. Il divieto, per il richiedente, di indicare l'IA come «autore o coautore» dell'opera la riduce a mero supporto creativo: sebbene tale funzione ausiliaria, in alcune ipotesi, «non solleverà dubbi sulla paternità umana»[295] (ad esempio, nel caso del semplice miglioramento dell'audio), diverso è il discorso per i *deepfakes*, che danno origine a un contenuto nuovo e falsificato.

---

[291] *Copyright Office Launches New Artificial Intelligence Initiative*, U.S. Copyright Office, 2023. Reperibile al sito: https://www.copyright.gov/newsnet/2023/1004.html

[292] U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 2023, sez. 2

[293] U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 2023, sez. 4, lett. A)

[294] *Ibidem*. Si specifica poi che, nell'ipotesi in cui il richiedente organizzi in modo creativo contenuti umani e generati da IA, esso può rivendicare il copyright per la selezione e disposizione dei contenuti.

[295] *Ibidem*

Le disposizioni che, nel diritto d'autore statunitense, escludono dalle opere protette i contenuti manipolati artificialmente, presentano delle eccezioni e delle limitazioni. In particolare, *la fair use doctrine*, pur legittimamente riconosciuta nell'ordinamento in esame, può presentare profili di criticità nel contesto dei *deepfakes*, in quanto suscettibile di essere invocata per legittimarne la creazione e diffusione.

La *fair use doctrine* permette l'utilizzo di opere protette senza il consenso del titolare dei diritti, per finalità quali critica, commento, cronaca, insegnamento o ricerca, e previa valutazione di quattro fattori: lo scopo e la natura dell'uso; la natura dell'opera originale; la quantità e importanza della porzione utilizzata; e l'effetto sul mercato o sul valore dell'opera protetta[296].

L'utilizzazione di un'opera a scopo didattico e non lucrativo potrebbe essere riconosciuta come "*fair*": tuttavia, la presenza di una finalità commerciale non esclude la liceità dell'uso se questo è "trasformativo", cioè se aggiunge un nuovo significato all'opera originale.

Il concetto è stato introdotto e sviluppato dalla giurisprudenza statunitense, la quale ha affermato che, per definirsi trasformativa, la nuova opera non può limitarsi «a sostituire gli obiettivi dell'opera originaria», ma deve modificarla «attraverso una nuova espressione, un nuovo significato o un nuovo messaggio»[297].

Negli anni, numerose sono state le controversie in cui la *fair use doctrine* è stata invocata per giustificare l'utilizzo di opere protette. Tra queste, di particolare interesse è il caso *The New York Times v. OpenAI*. Nel 2023, il Times ha citato in giudizio OpenAI e Microsoft, accusandole di aver utilizzato, senza autorizzazione, milioni di articoli protetti da *copyright* per l'addestramento dei modelli di IA, come ChatGPT e Copilot. Inoltre, il quotidiano ha contestato che gli *output* dei modelli si risolvessero in riassunti dettagliati o addirittura copie degli articoli protetti, permettendo così ai lettori di accedere ai contenuti

[296] *Copyright Act* of 1976, 17 U.S.C., sez. 107

[297] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994)

eludendo il sistema di *paywall*[298]. OpenAI si è difesa richiamando la *fair use doctrine*, sostenendo che l'utilizzo degli articoli fosse trasformativo in quanto finalizzato non a una loro mera riscrittura, ma alla creazione di un sistema capace di comprendere e generare il linguaggio naturale. L'azienda ha poi precisato che il suo obiettivo sarà quello di ridurre al minimo le ipotesi di *regurgitation*, cioè di riproduzione di interi pezzi di *dataset*[299]. Al punto di svolta decisivo, per la controversia, si è giunti il 26 marzo 2025, quando il giudice federale Sidney Stein, della Corte distrettuale di Manhattan, ha respinto le richieste di archiviazione presentate da OpenAI e Microsoft. Le accuse sollevate dal Times sono state ritenute sufficientemente fondate, e il giudice ha riconosciuto la novità e complessità della questione nel rapporto tra diritto d'autore e IA. Il procedimento, ancora in corso, potrebbe definire per la prima volta quali sono i confini del *fair use* in relazione all'IA generativa, indicando se l'utilizzo di opere protette per l'addestramento di tali modelli possa qualificarsi come trasformativo o, al contrario, integrare una violazione del *copyright*[300].

Il caso *The New York Times v. OpenAI* assume rilevanza, in relazione ai *deepfakes*, perché i creatori di tali contenuti potrebbero giustificarne la generazione ricorrendo alla stessa argomentazione adottata da OpenAI. In effetti, anche per i *deepfakes* si potrebbe parlare di uso trasformativo, essendo i materiali protetti convertiti in nuovi contenuti sintetici. Ne consegue che, in entrambe le circostanze, l'intervento dell'IA potrebbe qualificarsi "*fair*" ai sensi del diritto d'autore statunitense, e dunque lecito.

Tuttavia, la presenza dell'uso trasformativo non esclude potenziali lesioni a diversi soggetti: nella controversia prima illustrata, si verificherebbe un danno

---

[298] A. Pope, *NYT v. OpenAI: The Times's About-Face*, Harvard Law Review, 2024. Reperibile al sito: https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/

[299] V. Tiani, *Il New York Times vincerà la causa contro OpenAI? Un'analisi*, la Repubblica, 2024. Reperibile al sito: https://www.repubblica.it/tecnologia/blog/digital-europe/2024/01/19/news/il_new_york_times_vincera_la_causa_contro_openai_unanalisi-421918307/

[300] P. Feliciani, *NYT contro OpenAI: la battaglia del copyright nell'era dell'IA generativa*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/mercati-digitali/nyt-contro-openai-il-futuro-del-copyright-nellera-dellia-generativa/

economico per i titolari dei diritti d'autore; per i *deepfakes*, invece, il pregiudizio riguarderebbe la sfera personale dei soggetti rappresentati, incidendo sulla loro immagine e reputazione.

In definitiva, sarà importante osservare gli sviluppi del caso *The New York Times v. OpenAI* giacché le sue conseguenze potrebbero riflettersi anche sul modo di concepire i deepfakes: qualora la Corte dovesse concludere che l'utilizzo di opere protette per l'addestramento dei modelli di IA sia illecito, ci si troverebbe dinanzi un precedente in grado di porre severi limiti anche alla generazione di contenuti manipolati artificialmente.

2.2. Right of publicity

Sebbene il *Copyright Act* includa disposizioni potenzialmente applicabili ai *deepfakes*, la protezione offerta contro il fenomeno risulta parziale poiché circoscritta all'opera in quanto tale, non considerando l'appropriazione indebita (spesso accompagnata da lesioni) dell'identità dei soggetti raffigurati. Nell'ordinamento statunitense, la lacuna è però colmata grazie al *right of publicity*, istituto di origine giurisprudenziale che tutela l'individuo contro l'uso da parte di terzi, non autorizzato e a scopo di lucro, del proprio nome, immagine o somiglianza[301].

Inizialmente concepito quale derivazione del diritto alla *privacy*, dagli anni Cinquanta del secolo scorso il *right of publicity* è stato riconosciuto come un diritto autonomo, grazie al quale l'interessato può «concedere in licenza la propria immagine come sua proprietà»[302]. Nei decenni seguenti, la portata di questo diritto è stata espansa progressivamente, fino a comprendere, in alcuni

[301] M. S. Lee, *Entertainment and Intellectual Property Law*, Thomson Reuters, Toronto, 2023, § 3:62

[302] Il punto di svolta, in tal senso, è dato dal caso *Haelan Laboratories, Inc. v. Topps Chewing Gum, Inc.* del 1953. Per approfondimenti sul tema, cfr. T. Splagounias, *Rethinking the right of publicity in the era of generative AI*, in IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property, vol. 65, n. 1, 2024, pp. 43-44

Stati federati, «tutti gli elementi che contribuiscono a identificare una persona», come il nome, la voce e, più in generale, l'identità individuale[303].

Ad oggi, non esiste una normativa che riconosca il *right of publicity* a livello federale: di conseguenza, i singoli Stati regolano autonomamente la materia, dando luogo a un quadro normativo disomogeneo e composto da approcci tra loro divergenti[304]. A titolo di esempio, alcuni stati disciplinano il *right of publicity* tramite statuti *ad hoc*, mentre altri ricorrono alla sola *common law*[305]. Ancora, a variare è il numero di anni per il riconoscimento del *right of publicity post mortem*: nello Stato della California la durata è fissata a settant'anni «dalla morte della personalità deceduta»[306], in Virginia è ridotta a vent'anni, in Florida arriva a quaranta, nel Tennessee è solo decennale, addirittura in Indiana raggiunge i cento anni, estendendosi retroattivamente ai soggetti deceduti nei cent'anni precedenti[307].

Non potendosi, in questa sede, esaminare tutte le discipline adottate dai singoli stati, può essere utile soffermarsi su quella della California, che rappresenta l'esempio più rilevante e avanzato, fornendo una regolamentazione dettagliata in tema di *right of publicity* all'interno del *California Civil Code*. In particolare, la Section 3344 stabilisce che «chiunque utilizzi consapevolmente il nome, la voce, la firma, la fotografia o l'immagine di un'altra persona», a fini pubblicitari o di vendita e senza il suo consenso, è responsabile per i danni ad

---

[303] J. T. McCarthy, R. E. Schechter, *The Rights of Publicity & Privacy (2d ed.)*, Thomson Reuters, Toronto 2024, § 3:20

[304] D. Ervin, J.B. Steinberg, *AI and the Right of Publicity: A Patchwork of State Laws the Only Guidance, For Now*, Crowell, 2023. Reperibile al sito: https://www.crowell.com/en/insights/client-alerts/ai-and-the-right-of-publicity-a-patchwork-of-state-laws-the-only-guidance-for-now

[305] J. Faber, *A Brief History of Right of Publicity (NIL)*, Right Of Publicity, 2025. Reperibile al sito: https://rightofpublicity.com/brief-history-of-rop. Attualmente, gli Stati che hanno adottato statuti *ad hoc* sul *right of publicity* sono venticinque, mentre quelli che lo riconoscono in generale, anche grazie la sola *common law*, sono trentotto. Per una mappa sugli Stati che, ad oggi, possiedono uno statuto sul *right of publicity*, cfr. https://rightofpublicity.com/statutes

[306] California Civil Code, Section 3344.1, lett. g)

[307] J. Faber, *A Brief History of Right of Publicity (NIL)*, Right Of Publicity, 2025. Reperibile al sito: https://rightofpublicity.com/brief-history-of-rop

essa cagionati, dovendo corrispondere un importo di 750 dollari o pari ai «danni effettivamente subiti», se superiore, oltre a «qualsiasi profitto derivante»[308]. La disposizione, precisando che tale utilizzo può avvenire «in qualsiasi modo», presenta un'ampia portata applicativa[309], potendo dunque essere invocata anche in caso di *deepfakes*, purché ricorrano le predette condizioni. Ciò nonostante, il *right of publicity* non è esente da eccezioni: anzitutto, non è richiesto il consenso dell'interessato se l'uso riguarda «qualsiasi notizia, evento pubblico o trasmissione o resoconto sportivo, o qualsiasi campagna politica»[310]; inoltre, la finalità commerciale non è di per sé sufficiente a configurare l'illecito, dovendosi valutare, caso per caso, la presenza di un nesso diretto tra l'uso e lo scopo di lucro[311]; infine, non si considerano responsabili i mezzi di comunicazione che diffondano pubblicità in violazione della stessa Section 3344, a meno che non si dimostri la loro conoscenza circa l'uso non autorizzato dell'immagine della persona[312].

La disciplina fornita dal *California Civil Code* è poi integrata dalla Section 3344.1, che non solo estende il *right of publicity* alle persone decedute, applicando le medesime sanzioni e condizioni di responsabilità[313], ma introduce una responsabilità ulteriore per chi «produce, distribuisce o rende disponibile la replica digitale della voce o dell'immagine di una personalità deceduta» senza il consenso dei soggetti legittimati. Emerge dunque l'attenzione verso le riproduzioni digitali dell'immagine, tanto da prevedere la più grave sanzione di

---

[308] California Civil Code, Section 3344, lett. a)

[309] A prova di ciò, si consideri la lettera b) della Section 3344, che definisce il termine «fotografia» come «qualsiasi fotografia o riproduzione fotografica, fissa o in movimento, o qualsiasi videocassetta o trasmissione televisiva in diretta, di qualsiasi persona, tale che la persona sia facilmente identificabile». Considerando che la disposizione sanziona l'utilizzo dell'immagine in qualsiasi modo, l'elenco può essere interpretato estensivamente, fino ad includere anche i *deepfakes*.

[310] California Civil Code, Section 3344, lett. d)

[311] California Civil Code, Section 3344, lett. e)

[312] California Civil Code, Section 3344, lett. f)

[313] California Civil Code, Section 3344.1, lett. a), 1), A). Il paragrafo successivo prevede delle eccezioni al *right of publicity* "*post mortem*", non applicandosi se la finalità è di intrattenimento o in presenza di opere drammatiche, letterarie o musicali.

diecimila dollari o di un importo pari ai «danni effettivi subiti», se superiore[314], pur non mancando anche in tal caso eccezioni[315]. Più avanti, la Section 3344.1 definisce la «replica digitale» come una «rappresentazione elettronica altamente realistica, generata da computer», della voce o dell'immagine di una persona, e inserita in un contenuto sonoro o audiovisivo «in cui l'individuo reale non ha effettivamente eseguito o non è apparso, oppure l'individuo reale ha eseguito o è apparso, ma il carattere fondamentale dell'esecuzione o dell'apparizione è stato materialmente alterato»[316]. È questo il riferimento più esplicito ai *deepfakes* offerto dal *California Civil Code*.

Uscendo dal contesto californiano, è opportuno ricordare che quella appena considerata è solo una delle discipline presenti, negli Stati Uniti, sul *right of publicity*. Nonostante le differenze tra gli Stati federati, si può rilevare una regola di carattere generale: in caso di presunta violazione del *right of publicity*, l'attore deve provare la validità del proprio diritto e che quest'ultimo sia stato effettivamente leso dal convenuto[317]. Nel dettaglio, l'attore è tenuto a dimostrare che il convenuto abbia utilizzato la sua immagine in modo da renderlo identificabile e che, di conseguenza, ne sia stato danneggiato il valore commerciale, mentre non è necessario che il convenuto abbia agito con l'intenzione di nuocere[318]. Una volta accertata la violazione, si presume un danno alla persona, ma se l'attore richiede un risarcimento pecuniario, dovrà provare e quantificare il danno patrimoniale effettivamente subito: tale onere probatorio risulta agevole quando la violazione si manifesta in un *output* di IA generativa,

[314] California Civil Code, Section 3344.1, lett. a), 2), A), i)

[315] Nel dettaglio, la Section 3344.1, lett. a), 2), A), ii), rende lecita una replica digitale senza consenso se l'utilizzo avviene: a fini informativi, di commento, critica, satira o parodia; in documentari, storici o biografici, dell'individuo rappresentato; in modo fugace o incidentale; o nelle attività promozionali delle opere prima descritte.

[316] California Civil Code, Section 3344.1, lett. a), 2), B), ii)

[317] J. T. McCarthy, R. E. Schechter, *The Rights of Publicity & Privacy (2d ed.)*, Thomson Reuters, Toronto 2024, § 3:2

[318] *Ibidem*

ma difficoltosa nel caso in cui la propria immagine sia stata inclusa in un *dataset* per l'addestramento di un sistema di IA[319].

Un altro riferimento normativo, applicabile a livello federale, è incluso nel Lanham Act del 1946. In particolare, la Section 43, a) prevede che chiunque, in relazione a beni o servizi, utilizzi nel commercio parole, simboli o rappresentazioni false e fuorvianti, è responsabile qualora il pubblico sia tratto in inganno circa l'esistenza di un rapporto di connessione tra due soggetti, oppure sul fatto che i beni, i servizi o le attività commerciali di uno siano originati, sponsorizzati o approvati dall'altro[320]. Ne deriva che la portata di tale disposizione si estende anche ai *deepfakes*, in particolare quelli a fini promozionali e coinvolgenti personaggi noti[321], seppur in maniera più limitata rispetto al *right of publicity*. Difatti, mentre quest'ultimo si limita a vietare l'uso non autorizzato e a fini commerciali dell'immagine altrui, per la Section 43, a) del Lanham Act l'utilizzo deve necessariamente generare confusione o inganno nel pubblico sull'origine, sponsorizzazione o approvazione di un bene o servizio.

Dall'analisi svolta sul *right of publicity*, è possibile trarre un'importante considerazione. Per lungo tempo, il diritto in questione è stato in grado di evolversi, anche grazie al contributo della giurisprudenza, fino al punto da tutelare la persona da ogni forma di appropriazione indebita della sua immagine. Tuttavia, dinanzi al costante miglioramento della tecnologia, la normativa in

---

[319] T. Splagounias, *Rethinking the right of publicity in the era of generative AI*, in IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property, vol. 65, n. 1, 2024, p. 47

[320] Lanham (Trademark) Act (15 U.S.C.), sez. 43, a), 1), A). La disposizione, più avanti, estende la responsabilità anche a chi, nella pubblicità o promozione commerciale, rappresenti falsamente «la natura, le caratteristiche, le qualità o l'origine geografica dei beni, servizi o attività commerciali, propri o di un'altra persona».

[321] Sebbene la lettera della Section 43, a) del Lanham Act si rivolga genericamente a qualsiasi soggetti, nella prassi la tutela è circoscritta alle sole persone note, poiché solo in tali ipotesi l'uso dell'immagine altrui può generare confusione o inganno nel pubblico. Addirittura, vi è chi ritiene che il Congresso abbia concepito tale disposizione per «proteggere i consumatori dall'erronea convinzione che una celebrità abbia sponsorizzato un particolare prodotto». Sul punto, cfr. K. Rubin, *The Key to Being a Good Referee: The Call the Ninth Circuit Missed When Evaluating False Endorsement Claims*, in Emory Law Journal, vol. 64, 2015.

materia di *right of publicity* non si è mai aggiornata, e i pochi e frammentati strumenti vigenti non riescono più a fronteggiare le crescenti minacce dell'IA generativa[322], in particolare dei *deepfakes*, che rendono la manipolazione dell'immagine altrui estremamente accessibile.

È dunque necessaria la presenza di nuovi interventi, capaci di adeguare il *right of publicity* alle sfide dell'epoca digitale: a tal proposito, non sono mancate proposte normative avanzate dagli Stati Uniti, come il *No AI FRAUD Act* e il *No FAKES Act*, ai quali sarà dedicata attenzione più avanti.

2.3. Take It Down Act

La regolamentazione statunitense dei *deepfakes* ha raggiunto un punto di svolta significativo in tempi assai recenti: il 28 aprile del 2025 il Congresso ha approvato il *Take It Down Act* (acronimo di "Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act"), legge federale volta a criminalizzare il fenomeno del c.d. *revenge porn*, ossia la diffusione non consensuale di immagini intime, comprese quelle generate dall'IA tramite i *deepfakes*[323].

Sebbene il *Take It Down Act* non menzioni espressamente il termine «*deepfake*», il richiamo al fenomeno è evidente: la Section 2, che a sua volta modifica la Section 223 del Communications Act del 1934 (47 U.S.C.), introduce la rilevante nozione di «*digital forgery*» (o falsificazione digitale), con la quale si intende «qualsiasi rappresentazione visiva intima di un individuo identificabile» generata tramite «*machine learning*, intelligenza artificiale o qualsiasi altro mezzo generato dal computer o tecnologico», nonché la manipolazione di «una rappresentazione visiva autentica, che, se vista nel suo

[322] T. Splagounias, *Rethinking the right of publicity in the era of generative AI*, in IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property, vol. 65, n. 1, 2024, pp. 47-48

[323] *The Take It Down Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, Congress.gov, 2025. Reperibile al sito: https://www.congress.gov/crs-product/LSB11314

insieme da una persona ragionevole, è indistinguibile da una rappresentazione visiva autentica dell'individuo»[324]. La stessa Section stabilisce poi che è illegale «pubblicare consapevolmente una falsificazione digitale di un individuo identificabile», sia esso adulto[325] o minorenne[326], a meno che detta divulgazione non rientri in una «attività investigativa, protettiva o di *intelligence* legalmente autorizzata», o sia effettuata in buona fede e con ragionevolezza[327]. La disposizione, oltre a incriminare tali condotte[328], sanziona altresì chi minaccia di realizzarle.

L'innovatività del *Take It Down Act* emerge non solo nella previsione di nuove fattispecie incriminatrici, ma soprattutto nell'istituzione di una procedura di *notice and takedown* (notifica e rimozione). Entro il 16 maggio 2026, ogni «*covered platform*»[329] dovrà predisporre un procedimento col quale ogni individuo identificabile possa segnalare la presenza di una propria rappresentazione visiva intima, pubblicata senza il suo consenso, e richiederne la

[324] Communications Act of 1934 (47 U.S.C.), sez. 223, h), 1), A)

[325] La Section 223, h), 3), A) del Communications Act del 1934, così come modificata dalla Section 2 del *Take It Down Act*, prevede che, qualora la falsificazione ritragga un soggetto adulto, la condotta è vietata se: i) manca il consenso dell'interessato; ii) quanto pubblicato non è stato volontariamente esposto dall'individuo; iii) la raffigurazione non ha finalità di interesse pubblico; iv) la pubblicazione è destinata a causare o provocare danni psicologici, finanziari o reputazionali.

[326] Communications Act of 1934 (47 U.S.C.), sez. 223, h), 3), B). In caso di minori, la pubblicazione di una falsificazione digitale è illecita se ha lo scopo di «abusare, molestare o degradare il minore», oppure di «suscitare o gratificare il desiderio sessuale di una persona».

[327] Communications Act of 1934 (47 U.S.C.), sez. 223, h), 2), C)

[328] Per le pubblicazioni di falsificazioni digitali, si prevede una multa ai sensi del Titolo 18 dello U.S.C., l'incarcerazione per non più di due anni (tre se la parte lesa è un minore), o entrambe le pene. In caso di sola minaccia la pena è la medesima, ma l'incarcerazione, in questo caso, è ridotta a non più di diciotto mesi (trenta mesi se la parte lesa è un minore).

[329] La Section 4 del *Take It Down Act* approfondisce la nozione di «*covered platform*» (o piattaforma coperta), con la quale si intende «un sito web, un servizio online, un'applicazione online o un'applicazione mobile» che fornisce uno spazio per contenuti generati dagli utenti, come messaggi, video, immagini e audio, oppure che, nel normale corso della propria attività, ospita o pubblica rappresentazioni visive intime non consensuali.

rimozione[330]. La piattaforma deve rendere accessibili i processi di avviso e rimozione, indicando le proprie responsabilità e le modalità tramite cui un individuo può segnalare e chiedere la cancellazione di un contenuto[331]. In caso di richiesta di rimozione, la piattaforma dovrà agire tempestivamente, e comunque non oltre le 48 ore, adottando «ogni ragionevole sforzo» per eliminare anche eventuali copie identiche del medesimo contenuto[332].

La violazione degli obblighi di *notice and takedown* è considerata un atto o una pratica sleale ingannevole, con conseguente applicazione della disciplina di cui alla Section 18 del *Federal Trade Commission Act*: a tal fine, si attribuiscono alla Federal Trade Commission gli stessi poteri, doveri e procedure previste per tali violazioni dal FTC Act[333]. Tuttavia, la piattaforma non è responsabile nel caso in cui, in buona fede, disabiliti l'accesso o rimuova un contenuto che ritiene essere una rappresentazione visiva intima non consensuale, anche qualora in seguito se ne accerti la liceità[334].

Nonostante l'approvazione quasi unanime del Congresso e il sostegno di colossi della tecnologia come Meta, Google o TikTok, non sono mancate forti preoccupazioni verso il nuovo *Take It Down Act*. Anzitutto, si teme che la legge possa essere utilizzata per eliminare contenuti sessuali legittimi, limitare la libertà di espressione[335] o addirittura censurare materiale non gradito dal governo o dai suoi alleati politici[336]. Ulteriore criticità riguarda la limitazione dell'obbligo di rimozione alle sole *covered platforms*, con la conseguenza che un *deepfake porn* può circolare liberamente tramite la messaggistica privata: il paradosso è

---

[330] *Take It Down Act* (Public Law 119-12), sez. 3, a), 1), A)

[331] *Take It Down Act* (Public Law 119-12), sez. 3, a), 2)

[332] *Take It Down Act* (Public Law 119-12), sez. 3, a), 3)

[333] *Take It Down Act* (Public Law 119-12), sez. 3, b)

[334] *Take It Down Act* (Public Law 119-12), sez. 3, a), 4)

[335] A. Martínez, *Take It Down Act, la ley que criminaliza los deepfakes y la 'pornovenganza'*, El País, 2025. Reperibile al sito: https://elpais.com/us/2025-05-20/take-it-down-act-la-ley-que-criminaliza-los-deepfakes-y-la-pornovenganza.html

[336] L. Felner, *Take It Down Act heads to Trump's desk*, The Verge, 2025. Reperibile al sito: https://www.theverge.com/news/657632/take-it-down-act-passes-house-deepfakes

che, per garantire l'effettivo funzionamento della legge, le autorità potrebbero esser costrette a indebolire l'attuale sistema di crittografia dei canali di comunicazione riservata[337]. Ancora, il *Take It Down Act* sanziona chi diffonde contenuti illeciti, ma non interviene nei confronti di coloro che ne permettono la generazione, lasciando così un importante vuoto di tutela[338].

Concludendo, la rilevanza del *Take It Down Act* è indiscussa, in quanto rappresenta la prima fonte statunitense che ha riconosciuto le rappresentazioni intime generate dall'IA, introducendo dei relativi strumenti di tutela. Nonostante ciò, l'attuale disciplina, oltre a mostrare lacune evidenti, regola la sola dimensione pornografica dei *deepfakes*, trascurando tutte le altre sfumature del fenomeno, come le manipolazioni vocali o quelle usate a fini politici.

Tramite correttivi o nuove fonti che estendano il campo applicativo, risolvano le criticità prima evidenziate e predispongano anche strumenti di repressione *ex ante*, il quadro normativo statunitense potrebbe risultare uno degli strumenti normativi più avanzati in tema di *deepfakes*.

3. Cina

Dopo aver esaminato la regolamentazione dei *deepfakes* nell'Unione Europea e negli Stati Uniti, è opportuno uscire dal contesto occidentale per comprendere come il fenomeno sia disciplinato in altre realtà del globo. In particolare, si andrà a considerare l'approccio della Repubblica Popolare Cinese, profondamente diverso da quelli visti finora a causa del peculiare rapporto tra tecnologia e controllo statale dell'informazione.

Difatti, in Occidente la *governance* di Internet non è rimessa ai singoli Stati, ma è fondata su un approccio c.d. *multistakeholder*, nel quale concorrono i governi, la società civile e i fornitori di servizi online. Diversamente, la Cina ha voluto discostarsi da questo modello, temendo che «entità estere» potessero

[337] A. Mantzarlis, *The real problem with the Take It Down Act is not free speech*, Faked Up, 2025. Reperibile al sito: https://fakedup.org/the-real-problem-with-the-take-it-down-act-is-not-free-speech/

[338] *Ibidem*

compromettere la sicurezza nazionale influenzando i rapporti giuridici legati al *web* e rendendo le infrastrutture tecnologiche vulnerabili[339]. Un simile scenario spiega perché, nell'ordinamento cinese, lo sviluppo di Internet sia stato costantemente affiancato da un sistema di filtraggio volto ad oscurare qualsiasi contenuto contrario alla prospettiva del governo cinese[340].

In altre parole, nel territorio cinese si applica la c.d. «*cyber-sovranity*» (*wǎngluò zhǔquán*), principio secondo il quale lo Stato esercita una sovranità esclusiva non solo sul «*cyberspace*», ma anche sulle tecnologie emergenti: quest'ultimo aspetto interessa particolarmente per il presente scritto, poiché negli ultimi anni la Cina ha regolato non solo l'uso e lo sviluppo dell'IA, ma anche le relative applicazioni, tra cui i *deepfakes*[341].

Si ritiene che il governo cinese abbia iniziato a mostrare attenzione verso questi contenuti nel 2019, con il lancio dell'applicazione ZAO[342], basata sulla tecnica dello "scambio di volti" già vista per programmi occidentali come FakeApp (v. *supra*, cap. II, 2.1.). Di fronte a simili violazioni della *privacy*, le autorità cinesi hanno tempestivamente provveduto a regolamentare i *deepfakes*, dopo aver addirittura considerato l'ipotesi di un loro totale divieto. A pochi mesi dall'uscita di ZAO, la Cyberspace Administration of China ha approvato, con entrata in vigore dal 1° gennaio 2020, l'*Administrative Regulations on Online Audio and Video Information Services*[343]. Il Regolamento, finalizzato a «promuovere lo sviluppo sano e ordinato» dei servizi online e a «salvaguardare

---

[339] G. Santoni, *La Cina e lo spazio digitale. Questioni di governance nello spazio digitale globale*, in OrizzonteCina, vol. 11, 2020, pp. 70-71

[340] Ivi, p. 72. Si faccia l'esempio di testate giornalistiche internazionali come *The New York Times* o *Amnesty International* le quali, dopo aver pubblicato articoli contro il governo cinese, sono state messe al bando dal Paese.

[341] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[342] Y. Geng, *Comparing "Deepfake" Regulatory Regimes in The United States, the European Union, and China*, in Georgetown Law Technology Review, vol. 7, 2023, p. 167

[343] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

la sicurezza nazionale e gli interessi pubblici»[344], pone una serie di obblighi ai fornitori di servizi online, alcuni dei quali assumono particolare rilevanza in relazione ai *deepfakes*. È infatti previsto che, se i fornitori lanciano servizi basati su tecnologie come il *deep learning* o la realtà virtuale, devono condurre valutazioni di sicurezza conformi alle normative nazionali[345]. Inoltre, i fornitori devono rafforzare la gestione dei contenuti pubblicati dagli utenti, provvedendo a bloccarli, cancellarli e segnalarli in caso di contrasto con la legge[346], e istituendo appositi «meccanismi di smentita» delle informazioni falsamente diffuse[347]. Rilevante è infine la disposizione che introduce un obbligo generale di trasparenza, stabilendo che i fornitori di servizi online e gli utenti che «creano, pubblicano o diffondono informazioni audiovisive non autentiche», mediante strumenti di IA, devono «chiaramente etichettare tali informazioni»[348].

Sebbene introduca per la prima volta riferimenti normativi applicabili ai *deepfakes*, il Regolamento del 2020 è incentrato principalmente sulla gestione dei contenuti audiovisivi online. Nel 2022, considerando la rapida evoluzione assunta dall'IA generativa, la *Cyberspace Administration of China* ha introdotto un nuovo Regolamento, intitolato "*Provisions on the Administration of Deep Synthesis Internet Information Services*". La normativa si applica anche, ma non solo, ai *deepfakes*, i quali sono inclusi nella più ampia categoria delle c.d. «*deep synthesis*», ossia quelle tecnologie, come il *deep learning*, che «utilizzano algoritmi di sequenziamento generativo per creare testo, immagini, audio, video,

[344] Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 1

[345] Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 10

[346] Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 12

[347] Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 13

[348] Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 11

scene virtuali o altre informazioni»[349]. Obiettivo del Regolamento è proprio regolare l'applicazione di tali tecnologie all'interno dei servizi di informazione su Internet.

Anche in questo caso è possibile individuare disposizioni rilevanti nell'ambito dei *deepfakes*. Anzitutto, si vieta ad organizzazioni e individui di utilizzare servizi di *deep synthesis* per generare o diffondere contenuti illegali o dannosi, come quelli che minacciano la sicurezza nazionale o «danneggiano l'immagine della nazione», nonché per produrre *fake news*[350]. È poi previsto che i fornitori di servizi di *deep synthesis* garantiscano la sicurezza dei dati di *training* e, in caso di strumenti che consentono la modifica di dati biometrici, impongano agli utenti di informare i soggetti coinvolti e ottenerne il consenso[351]. Ancora, in relazione a sistemi che generano o modificano dati biometrici, i fornitori devono effettuare valutazioni di sicurezza[352], predisporre canali di ricorso per gli utenti[353], autenticare l'identità di questi ultimi[354] e apporre un'etichettatura visibile sul contenuto, in modo da avvisare il pubblico sulla loro natura sintetica[355].

È indubbio che il Regolamento del 2022 garantisca e rafforzi la protezione degli utenti contro la diffusione incontrollata dei *deepfakes*. Tuttavia, parte della dottrina ritiene che gli obblighi imposti ai fornitori potrebbero comunque rivelarsi insufficienti. In effetti, il testo non specifica quale etichettatura debba

[349] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 23. La disposizione riporta, a scopo esemplificativo, le tecnologie per la generazione o la modifica di contenuti testuali, vocali, musicali, caratteristiche biometriche di immagini e contenuti video o personaggi virtuali. Appaiono evidenti, in alcuni di questi esempi, i riferimenti ai *deepfakes*.

[350] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 6. Pur in assenza di un divieto generalizzato, non manca chi ritiene che la vaghezza della formulazione, tipica nell'ordinamento cinese, renda potenzialmente sanzionabile qualsiasi *deepfake*. Cfr. A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[351] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 14

[352] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 15

[353] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 12

[354] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 9

[355] Provisions on the Administration of Deep Synthesis Internet Information Services, art. 17

concretamente applicarsi, la quale, in ogni caso, può essere rimossa tramite altre tecnologie, lasciando i fornitori comunque responsabili per la circolazione di tali contenuti. Oltre a questo, l'obbligo di rimozione dei contenuti dannosi di cui all'art. 10 presenta evidenti vuoti di tutela, poiché un contenuto appena generato può essere estratto dal servizio originario, mediante tecniche quali lo *screenshot*, e condiviso altrove[356].

Per reagire ad alcune di queste lacune, la Cyberspace Administration of China ha pubblicato, con entrata in vigore dal 1° settembre 2025, le *Measures for Labeling of AI-Generated Synthetic Content*, prevedendo l'obbligo di etichettatura dei contenuti sintetici generati dall'IA[357]. Tali misure, redatte anche sulla base del Regolamento del 2022, distinguono tra etichettature «esplicite», chiaramente individuabili dagli utenti e che appaiono in forme quali testo o immagini, e «implicite», non immediatamente percepibili ma integrate nei dati del *file*[358].

Centrale è l'art. 4, che prevede l'obbligo, per i fornitori di servizi, di apporre «etichette esplicite» sugli *output* generati tramite *deep synthesis* e potenzialmente ingannevoli, precisandone i requisiti in base alla tipologia di contenuto: ad esempio, un *file* audio dovrà contenere una notifica vocale all'inizio, al centro o alla fine della traccia, mentre un'immagine o un video dovranno riportare un'etichetta di notifica evidente e collocata in posizioni appropriate[359]. Gli stessi fornitori devono aggiungere anche delle etichette

[356] A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

[357] L'art. 3 definisce il «contenuto sintetico generato dall'intelligenza artificiale» come quel «testo, immagini, audio, video, scene virtuali o altre informazioni generate o sintetizzate utilizzando la tecnologia dell'intelligenza artificiale».

[358] Measures for Labeling of AI-Generated Synthetic Content, art. 3

[359] Measures for Labeling of AI-Generated Synthetic Content, art. 4. Sono previste istruzioni anche per i *file* testuali, per la realtà virtuale e per i casi che non rientrino in quelli prima menzionati, le quali fanno tutte leva sul posizionamento dell'etichetta in posizioni appropriate dell'output finale, così da rendere palese al pubblico la loro natura sintetica.

implicite «ai metadati dei *file* di contenuto sintetico»[360]. Fermo restando il rispetto delle disposizioni precedenti, è eccezionalmente consentito che a un utente siano forniti contenuti sintetici senza etichetta, purché esso lo richieda e accetti un «contratto di servizio» con le relative responsabilità d'uso. In ogni caso, i fornitori devono conservare i registri delle attività per almeno sei mesi[361].

Tali misure, oltre che ai fornitori, si estendono alle «piattaforme di distribuzione di applicazioni Internet», le quali sono tenute a chiedere ai fornitori se le app offrono servizi di IA generativa. In caso di risposta affermativa, la piattaforma deve verificare la conformità dell'etichettatura[362].

Da ultimo, sono presenti disposizioni che coinvolgono anche gli utenti. È infatti stabilito che chiunque pubblichi contenuti sintetici deve dichiararlo «in modo proattivo» e utilizzare le etichettature di cui dispone la piattaforma. L'etichetta così creata non può essere rimossa, modificata o occultata, né essere adoperata per ledere i diritti e gli interessi legittimi altrui[363].

La disciplina inclusa nelle misure si chiude prevedendo che, in caso di violazione delle precedenti disposizioni, intervengano «i dipartimenti di regolazione competenti», come quelli per le telecomunicazioni o la sicurezza pubblica, in conformità con la normativa vigente[364].

Concludendo sulla regolamentazione dei *deepfakes* in Cina, può senza dubbio rilevarsi la presenza di un sistema attualmente tra i più completi e operativi. Al contrario, negli Stati Uniti è ancora assente una legge federale sui *deepfakes* intesi globalmente. In Europa, invece, l'approccio delineato dall'AI Act appare più cauto e selettivo in confronto a quello cinese. Ad esempio, gli obblighi di trasparenza sono rivolti ai soli fornitori e *deployers*, non anche agli

[360] Measures for Labeling of AI-Generated Synthetic Content, art. 5. La disposizione definisce i «metadati dei *file*» come le «informazioni descrittive incorporate nell'intestazione del *file* secondo uno specifico formato di codifica, utilizzato per registrare il contenuto informativo, quali l'origine del *file*, gli attributi e gli utilizzi».

[361] Measures for Labeling of AI-Generated Synthetic Content, art. 9

[362] Measures for Labeling of AI-Generated Synthetic Content, art. 7

[363] Measures for Labeling of AI-Generated Synthetic Content, art. 10

[364] Measures for Labeling of AI-Generated Synthetic Content, art. 13

utenti e alle piattaforme di distribuzione di app. Ancora, l'obbligo di etichettatura è previsto per i contenuti che assomigliano a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbero falsamente autentici o veritieri, mentre in Cina devono essere etichettati tutti i contenuti generati dall'IA[365]. Deve poi considerarsi che, a differenza della normativa cinese, l'AI Act entrerà pienamente in vigore soltanto dal 1° agosto 2026. Tale netta inversione di tendenza rispetto al modello occidentale può trovare giustificazione nella diversa filosofia regolatoria di quest'ultimo, incentrata sulla cooperazione con le parti interessate e l'applicazione di sanzioni e misure proporzionate al rischio[366].

In ogni caso, mentre il resto del mondo dibatte ed è in cerca di soluzioni più rapide e concrete, la Cina ha prodotto un vero e proprio «effetto dimostrazione». Paradossalmente, l'Unione Europea, tradizionale *leader* nel campo della regolazione, si trova oggi in una posizione di «rincorsa» alla disciplina dei *deepfakes* (e più in generale dell'IA): solo osservando le ricadute pratiche della normativa cinese sarà possibile comprendere quale direzione dovrà intraprendere l'Occidente[367].

4. I deepfakes in Italia e responsabilità

Dopo aver esaminato gli approcci normativi adottati in contesti differenti come l'Unione Europea, gli Stati Uniti e la Cina, può valutarsi con maggiore consapevolezza critica la disciplina legislativa italiana.

In Italia, il tema dei *deepfakes* irrompe nel dibattito pubblico e giuridico a partire dal 2020, con la scoperta di un *bot* sull'app Telegram che, tramite il *software* "DeepNude", aveva manipolato immagini di oltre 680mila donne

[365] D. Clementi, *Contenuti IA: la Cina etichetta tutto, l'Europa riflette (e insegue)*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/contenuti-ia-la-cina-etichetta-tutto-leuropa-riflette-e-insegue/

[366] *Ibidem*

[367] *Ibidem*

(incluse giovani influencer italiane) al fine di denudarle[368]. Di conseguenza, il Garante per la protezione dei dati personali ha aperto un'istruttoria nei confronti di Telegram[369], pubblicando in seguito numerosi documenti volti ad approfondire e sensibilizzare i cittadini sul crescente fenomeno dei *deepfakes*[370].

Nonostante la reazione delle autorità, è doveroso precisare che, fino alla recente entrata in vigore della legge n. 132 del 23 settembre 2025, di cui si tratterà a breve, l'ordinamento italiano era rimasto privo di strumenti capaci di contrastare direttamente la diffusione di contenuti manipolati dall'IA e lesivi della reputazione o dell'immagine altrui. Pertanto, i soggetti coinvolti potevano tutelarsi esclusivamente invocando disposizioni preesistenti e applicabili estensivamente al fenomeno dei *deepfakes*, che meritano dunque di essere analizzate.

Anzitutto, qualora un *deepfake* risulti lesivo della reputazione di una persona, è integrato il reato di diffamazione previsto all'art. 595 c.p.: come già visto (v. *supra*, cap. I, 3.), la giurisprudenza della Cassazione ha ritenuto configurabile tale reato anche attraverso l'utilizzo di Internet, riconoscendo in tal caso anche la circostanza aggravante di cui all'art. 595 comma 3 c.p.[371].

Nel caso in cui un contenuto manipolato artificialmente sia generato allo scopo di ingannare il destinatario, la condotta rientrerà, a seconda delle circostanze, nelle fattispecie di sostituzione di persona (art. 494 c.p.), truffa (art.

---

[368] R. Rijtano, *Deepfake su Telegram, c'è un bot che spoglia le donne: quasi 700 mila vittime*, Repubblica, 2020. Reperibile al sito: https://www.repubblica.it/tecnologia/sicurezza/2020/10/20/news/deepfake_un_bot_sveste_le_donne_oltre_100mila_vittime-271131375/

[369] Garante per la Protezione dei Dati Personali, *Deep fake: il Garante privacy apre un'istruttoria nei confronti di Telegram per il software che "spoglia" le donne*, 2020. Reperibile al sito: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9470722

[370] Si pensi, tra i più rilevanti, al vademecum "Deepfake. Il falso che ti «ruba» la faccia (e la privacy)", che introduce il concetto di *deepfake*, ne illustra i collegamenti con il *deepnude*, il *cyberbullismo* e il *cybercrime*, per poi fornire brevi istruzioni al pubblico su come riconoscere e proteggersi dal fenomeno. Cfr. Garante per la Protezione dei Dati Personali, *Deep fake. Il falso che ti ruba la faccia (e la privacy)*, 2020. Reperibile al sito: file:///C:/Users/Utente/Downloads/Deepfake%20-%20Vademecum%20(3).pdf

[371] Corte di Cassazione, Sez. V Penale, sent. 17 novembre 2000, n. 4741

640 c.p.), frode informatica (art. 640-ter), estorsione (art. 629 c.p.) o atti persecutori (art. 612-bis)[372].

La produzione di un *deepfake porn* è invece più complessa da inquadrare. L'art. 10 della legge n. 69/2019, nota come "Codice Rosso" e finalizzata a rafforzare la tutela delle vittime di violenza domestica e di genere, ha introdotto nel codice penale il nuovo art. 612-ter, che punisce la diffusione non consensuale di immagini o video sessualmente espliciti[373]. Ciononostante, la disposizione, la cui *ratio* è quella di contrastare il fenomeno del *revenge porn*, non si estende alle ipotesi in cui i contenuti ritraenti la persona offesa siano generati dall'IA. Ne deriva che, qualora sia diffuso un *deepfake porn* di una persona reale senza il suo consenso, la condotta non risulterebbe punibile ai sensi dell'art. 612-ter c.p.[374].

Diverse sono le ipotesi di produzione o detenzione di materiale pornografico nel quale siano rappresentati soggetti minorenni, cui si applicano rispettivamente gli artt. 600-ter e 600-quater c.p. L'art. 4 della Legge n. 38/2006 ha inserito nel codice penale l'art. 600-quater 1, che estende l'applicabilità dei predetti articoli al caso in cui il contenuto pedopornografico «rappresenta immagini virtuali realizzate utilizzando immagini di minori degli anni diciotto o parti di esse»[375]. Il comma successivo specifica che per «immagini virtuali» si intendono quelle «realizzate con tecniche di elaborazione grafica non associate in tutto o in parte a situazioni reali», ma in grado di far «apparire come vere situazioni non reali»[376]: è evidente come una simile definizione ricomprenda

---

[372] M. Cartisano, *Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi*, Agenda Digitale, 2023. Reperibile al sito: https://www.agendadigitale.eu/sicurezza/privacy/deepfake-porno-con-giorgia-meloni-ecco-tutti-i-reati-commessi/

[373] Nel dettaglio, l'art. 612-ter c.p. stabilisce che «chiunque, dopo averli realizzati o sottratti, invia, consegna, cede, pubblica o diffonde immagini o video a contenuto sessualmente esplicito, destinati a rimanere privati, senza il consenso delle persone rappresentate» è punito con la reclusione da uno a sei anni e con multa da 5000 a 15000 euro.

[374] G. Ranghi, *Violenza digitale: la diffusione illecita di immagini o video sessualmente espliciti*, Salvis Juribus, 2022. Reperibile al sito: https://www.salvisjuribus.it/violenza-digitale-la-diffusione-illecita-di-immagini-o-video-sessualmente-espliciti/

[375] Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 640-quater 1, co. 1

[376] Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 640-quater 1, co. 2

anche i *deepfake porn*, considerati dunque punibili dal codice penale nei soli casi previsti dagli artt. 600-ter e 600-quater c.p.

Oltre al codice penale, occorre considerare quali siano le disposizioni invocabili nell'eventualità che un deepfake risulti lesivo dell'immagine di una persona. L'art. 96 della legge n. 633/1941, in materia di protezione del diritto d'autore e degli altri diritti connessi, vieta a chiunque di riprodurre o pubblicare il ritratto di una persona senza il consenso di quest'ultima[377]. Il successivo art. 97 chiarisce che il consenso dell'interessato non è necessario se la diffusione dell'immagine è giustificata dalla sua notorietà, dall'esercizio di pubbliche funzioni, da esigenze di giustizia o di polizia, oppure da scopi scientifici, didattici o culturali. Tuttavia, l'esposizione dell'immagine altrui resta in ogni caso illecita qualora «rechi pregiudizio all'onore, alla reputazione od anche al decoro della persona ritrattata»[378]. Agli artt. 96 e 97 va poi ricollegato l'art. 10 c.c., il quale ribadisce che se l'immagine di una persona (o dei genitori, coniuge e figli) è pubblicata fuori dai casi di legge o «con pregiudizio al decoro o alla reputazione», l'interessato potrà chiedere al giudice di ordinarne la rimozione, salvo il diritto al risarcimento dei danni[379].

Le disposizioni analizzate finora hanno rappresentato, per molto tempo, gli unici possibili strumenti di tutela contro la produzione e diffusione di un *deepfake*. Il legislatore italiano ha però deciso di intervenire con la legge n. 132/2025, pubblicata in G.U. lo scorso 25 settembre ed entrata in vigore il 10 ottobre: l'Italia risulta così il primo paese nell'UE ad adeguare il proprio ordinamento al quadro fornito dall'AI Act[380]. Obiettivo della presente Legge è disciplinare l'utilizzo dei sistemi di IA nei diversi settori, come quello sanitario, del lavoro o delle pubbliche amministrazioni, avendo inoltre cura di delineare gli indirizzi della strategia nazionale per l'IA e introdurre disposizioni in materia di

[377] Legge 22 aprile 1941, n. 633, art. 96

[378] Legge 22 aprile 1941, n. 633, art. 97

[379] Codice Civile (R.D. 16 marzo 1942, n. 262), art. 10

[380] *Approvata in via definitiva la legge italiana sull'Intelligenza Artificiale*, Dipartimento per la trasformazione digitale, 2025. Reperibile al sito: https://innovazione.gov.it/notizie/articoli/approvata-in-via-definitiva-la-legge-italiana-sull-intelligenza-artificiale/

diritto d'autore e sanzioni penali. È quest'ultimo l'aspetto che, in questa sede, interessa particolarmente.

L'art. 26 della Legge n. 132/2025 introduce, tra le altre modifiche apportate al codice penale, il nuovo art. 612-quater, rubricato «Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale»: si tratta della prima specifica fattispecie di reato per i *deepfakes*. La disposizione punisce, con la reclusione da uno a cinque anni, chiunque arrechi un «danno ingiusto» a una persona «cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità»[381]. Il reato è perseguibile tramite querela della persona offesa, salvo i casi, previsti dal secondo comma dello stesso articolo, in cui è possibile procedere d'ufficio[382].

Dalla lettera della disposizione, si desume che non è sanzionata la mera creazione del contenuto falsificato, ma unicamente la sua diffusione, e a condizione che si concretizzi un danno ingiusto per la persona coinvolta. Pertanto, l'art. 612-quater c.p. configura un reato di danno, e non di pericolo, non essendo penalmente perseguibile chi crea un *deepfake* ma non lo diffonde, chi pubblica un prodotto manifestamente sintetico e inidoneo a ingannare i destinatari nel concreto, oppure chi ha creato un contenuto falso ma non è a conoscenza della successiva diffusione da parte di terzi, i quali diverranno responsabili[383].

Accanto al nuovo illecito sui *deepfakes*, la Legge n. 132/2025 inserisce anche una circostanza aggravante comune all'art. 61 c.p., ossia l'aver commesso il delitto utilizzando un sistema di IA quando lo stesso sia stato impiegato come

[381] Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 612-quater

[382] Specificamente, si procede d'ufficio se il fatto: a) è connesso a un altro delitto perseguibile d'ufficio; b) è commesso contro un soggetto incapace, per età o infermità, o una pubblica autorità a causa delle funzioni esercitate.

[383] L. Turini, *Fermare i deepfake*, Appunti – di Stefano Feltri, 2025. Reperibile al sito: https://appunti.substack.com/p/fermare-i-deepfake

«mezzo insidioso», rendendo più difficile la difesa o aggravando le conseguenze del reato[384].

Con l'introduzione di una fattispecie di reato *ad hoc*, l'ordinamento italiano dimostra di essere all'avanguardia, almeno in Europa, in tema di regolamentazione dei *deepfakes*. Tale considerazione trova conferma nella presenza di un articolato sistema di responsabilità. Difatti, alla già esaminata responsabilità penale, in Italia si aggiungono le disposizioni dei Regolamenti UE, in particolare il Digital Services Act e l'AI Act, che impongono obblighi specifici per piattaforme, *providers* e *deployers*.

Ciononostante, alcune problematiche sono tuttora irrisolte. In primo luogo, nell'applicazione dell'art. 612-quater c.p., la giurisprudenza dovrà delineare il confine tra la tutela della dignità umana e la libertà di espressione, di critica e di satira[385]. Inoltre, la medesima norma contiene l'ampia nozione di «danno ingiusto», il cui contenuto, ambito applicativo e limiti risultano ancora da chiarire onde evitare incertezze interpretative. Ancora, la nuova disposizione del codice penale non ha risolto il problema dell'assenza, nell'ordinamento italiano, di strumenti per la tutela immediata dell'offeso e di misure cautelari volte ad evitare la reiterazione del danno: a tal proposito, parte della dottrina ritiene utile estendere l'applicazione dell'art. 700 c.p.c. in materia di provvedimenti d'urgenza, o introdurre un rito speciale in caso di *deepfakes*[386]. Infine, anche strumenti complementari come il DSA e l'AI Act presentano evidenti criticità, già esaminate (v. *supra*, 1.2, 1.3), le quali si riflettono inevitabilmente sull'ordinamento italiano.

In definitiva, l'analisi della regolazione dei *deepfakes*, in Italia e nelle altre realtà considerate, permette di individuare le risposte normative attuali per contrastarne la diffusione illecita, ma al tempo stesso evidenzia come tali strumenti siano ancora incompleti, necessitando di interventi correttivi nel

[384] Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 61, n. 11-undecies)

[385] M. Martorana, *Deepfake, quando l'AI manipola la realtà: il diritto in affanno*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/deepfake-quando-lai-manipola-la-realta-il-diritto-in-affanno/

[386] *Ibidem*

prossimo futuro. Pertanto, il quarto e ultimo capitolo sarà dedicato alle nuove proposte sviluppate sul tema nei diversi ordinamenti del globo.

5. I deepfakes come elemento di prova e problemi processuali

Prima di esaminare le più recenti proposte normative dei vari ordinamenti sulla regolazione dei *deepfakes*, è essenziale inquadrare il contesto in cui sono inseriti in ambito giudiziario.

Il processo odierno si trova a fronteggiare una sfida inedita: i *deepfakes*, inizialmente riconoscibili a occhio nudo, stanno evolvendo fino a risultare sempre più indistinguibili da un contenuto reale. Ne deriva che, nel caso in cui un materiale sintetico assuma valenza probatoria nel processo, il giudice dovrà necessariamente ricorrere a supporti tecnologici per valutarne l'autenticità, risultando ormai inidonei a tale scopo i sistemi tradizionali[387]. Ad esempio, la perizia tecnica crea il paradosso per cui la veridicità di un prodotto dell'IA può essere accertata soltanto utilizzando un altro sistema di IA, facendo sì che la decisione finale non nasca più dal giudice, ma dalla mera interpretazione di tecnologie avanzate[388]. Si pensi ancora all'intuizione del giudice la quale, pur socialmente accettata come metro di valutazione dell'attendibilità di una testimonianza, presenta evidenti limiti poiché si fonda su apprezzamenti soggettivi difficilmente verificabili[389].

Sulla base di tali premesse, è possibile procedere a una disamina dei nodi problematici che i *deepfakes* sollevano a livello processuale e probatorio.

Anzitutto, in presenza di un contenuto manipolato dall'IA, non è sufficiente qualificarne la natura, occorrendo anche accertarne l'impatto e la corrispondenza a situazioni realmente avvenute. A differenza della

[387] J. Nieva-Fenoll, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024

[388] *Ibidem*

[389] B. Russell, *The Collected Papers of Bertrand Russell, Volume 11. Last Philosophical Testament 1947-68*, Routledge, London, 1997, pp. 542 ss.

rappresentazione satirica di una persona esistente, tutelata dalla libertà di espressione salvo intenti esclusivamente denigratori, il materiale sintetico ha la capacità di rendersi indistinguibile dal reale[390]. Da qui sorge il problema per il diritto penale: la sanzione non si fonda sulla veridicità del fatto rappresentato, ma sulla potenziale lesione di beni giuridici come la reputazione, l'immagine pubblica o l'integrità psicologica della persona coinvolta[391].

Deve poi rilevarsi che, mentre nelle falsificazioni tradizionali, come quella di un documento, la valutazione ha ad oggetto l'autenticità e solo eventualmente il contenuto, nel caso dei *deepfakes* deve essere accertata simultaneamente la loro natura, che può essere reale o fittizia, e la sussistenza degli elementi costitutivi per il verificarsi della fattispecie penale contestata[392]. La contestualità dell'accertamento, tuttavia, rende palesi alcuni profili critici. Da un lato, tali elementi, in particolare il primo e il terzo, non possono essere rimessi alla mera percezione del giudice, in quanto fattori soggettivi come sensibilità o pregiudizi individuali potrebbero compromettere il principio di neutralità. Dall'altro, il ricorso a sistemi di IA, benché orientato a una maggior oggettività, deve garantire la prevenzione di *bias* tecnici che riflettano visioni parziali, specialmente in ambiti eticamente sensibili, come la sessualità[393]. Inoltre, si assiste a un mutamento della funzione della prova, non più limitata a verificare la falsità o meno del contenuto, ma estesa alla sua decostruzione al fine di distinguere le parti giuridicamente rilevanti da quelle che non lo sono[394]. Per esemplificare, si consideri un medesimo soggetto ritratto in due *deepfakes*. Nel primo, palesemente falso, egli è raffigurato in condotte suscettibili di ledere la sua immagine o

---

[390] J. Nieva-Fenoll, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024

[391] Cfr. I. Salvadori, *La controvertida relevancia penal del sexting en el derecho italiano y comparado*, in Revista Electrónica de Ciencia Penal y Criminología, n. 19 (29), 2017

[392] J. Nieva-Fenoll, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024

[393] *Ibidem*

[394] Cfr. K.T. May, S. Bray, T. Davies, L.D. Griffin, *Warning: Humans cannot reliably detect speech deepfakes*, in PLOS One, n. 18 (8), 2023

reputazione; nel secondo, più realistico, il protagonista compie attività quotidiane: è evidente come solo il primo *file*, pur meno accurato del secondo, assuma una concreta rilevanza giuridica.

Ulteriore elemento da considerare è la perizia tecnica che, pur costituendo all'apparenza la prova più affidabile per determinare l'esistenza di un contenuto manipolato dall'IA, presenta ostacoli attualmente difficili da superare. In primo luogo, per il giurista, comprendere pienamente le tecniche di rilevamento dei *deepfakes* è assai difficile, non solo per la loro costante evoluzione[395], ma anche per la necessità di competenze specialistiche che una corretta applicazione richiede. A ciò si aggiunge un forte aumento dei costi del contenzioso, dovuto alla remunerazione degli esperti, nonché l'assenza, verso tali tecniche, di un consenso generale presso la comunità scientifica[396], fattore che ne impedisce l'ammissibilità probatoria in sede giudiziaria[397].

Nonostante questo, almeno nel prossimo futuro, è verosimile che le relazioni peritali sui materiali sintetici continueranno a essere prodotte in giudizio, anche per la possibilità della controparte di difendersi nominando a sua volta un esperto. Tuttavia, la crescente diffusione dei *deepfakes* potrebbe mettere fortemente a rischio la credibilità del documento audiovisivo come elemento di prova, in quanto diverrebbe insostenibile sottoporre ogni immagine o video a un accertamento specialistico.

---

[395] Cfr. M. Westerlund, *The Emergence of Deepfake Technology: A Review*, in Technology Innovation Management Review, vol. 9, n. 11, pp. 40-53. A titolo di esempio, si menzioni "NoiseDF", una tecnica di rilevamento dei deepfakes basata sull'analisi del rumore digitale presente nei video. In particolare, il metodo confronta il rumore presente nei fotogrammi del volto con quello dello sfondo, per verificare che siano coerenti. Una forte discordanza tra i due valori conferma che il file è stato manipolato. Il sistema, tuttavia, funziona solo in presenza di uno sfondo, poiché esaminando il solo volto non sarebbe possibile un confronto.

[396] S.C. Fernández, *El «deepfake» en los documentos audiovisuales: un reto para la valoración de la prueba*, in Actualidad Civil, n. 2, 2024, p. 6

[397] J. Nieva-Fenoll, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024. Si registra dunque un approccio diverso rispetto al passato, in cui mezzi peritali privi di supporto scientifico, come le impronte digitali o del DNA, erano ammesse con maggior elasticità.

Dinanzi a un sistema così incerto, la dottrina ha cominciato a prospettare possibili soluzioni. Tra queste, rileva l'introduzione di una legge che indichi precisi requisiti per l'autenticità di un'immagine: tuttavia, qualora nuove tecnologie riescano ad emulare detti requisiti, anche una proposta legislativa potrebbe rivelarsi fragile[398]. Vi è poi chi propende per una sostituzione della perizia tecnica con le prove tradizionali, sebbene queste ultime presentino forti limiti, come nel caso della testimonianza[399]. Ancora, secondo parte della dottrina sarà necessario modificare le attuali norme processuali, a partire da un innalzamento degli *standard* probatori per provare l'autenticità[400]. Infine, una prospettiva più drastica mette in discussione il contenzioso come mezzo privilegiato di risoluzione delle controversie, suggerendo un'apertura a soluzioni alternative per «i conflitti delle generazioni future»[401].

In ultima analisi, l'ingresso dei *deepfakes* come elemento di prova ha determinato una forte instabilità nell'intero quadro processuale, non limitandosi a coinvolgere l'attendibilità del singolo contenuto manipolato, ma anche di tutte le immagini, video o suoni aventi valore probatorio. Le soluzioni dottrinali fin qui illustrate potrebbero senz'altro apportare maggiore chiarezza in futuro, ma presuppongono un ripensamento dell'attuale disciplina probatoria e processuale, oltre a una costante formazione di giudici e avvocati sulle tecnologie emergenti[402]. Qualora gli ordinamenti non dovessero offrire una risposta adeguata, il rischio principale sarebbe quello di assistere a una sorta di equiparazione tra realtà e finzione, in cui il falso può apparire più convincente del reale, e il vero può essere contestato come artificiale.

---

[398] *Ibidem*

[399] S.C. Fernández, *El «deepfake» en los documentos audiovisuales: un reto para la valoración de la prueba*, in Actualidad Civil, n. 2, 2024, p. 5

[400] J.P. LaMonaga, *A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes*, in American University Law Review, vol. 69, 2020

[401] J. Nieva-Fenoll, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024

[402] Per approfondimenti sulla necessità di formazione dei giuristi in tema di *deepfakes*, cfr. G. Bellas, *Deepfakes in the Courtroom: Problems and Solutions*, in Bench & Bar, vol. 55, n. 5, 2025, p. 5

CAPITOLO IV

VERSO NUOVE FORME DI TUTELA DEI DEEPFAKES

SOMMARIO: 1. Nuove proposte normative dal mondo. – 2. Profili etici.

1. Nuove proposte normative dal mondo

Nel capitolo precedente è stato ricostruito il quadro normativo attuale in materia di *deepfakes*, con riferimento all'ordinamento dell'Unione Europea, degli Stati Uniti e della Cina, richiamando anche il contesto italiano. A questo punto, può essere utile dedicare l'attenzione alle più recenti iniziative normative, avanzate da diverse realtà globali, al fine di comprendere i possibili sviluppi della regolamentazione del fenomeno.

In primo luogo, occorre analizzare nuovamente il contesto statunitense: sebbene il *Take It Down Act* sia entrato in vigore recentemente, la sua portata, come già rilevato (v. *supra*, cap. III, 2.3.), è stata circoscritta ai soli *deepfakes* pornografici. Per tale ragione, è stato rilevato come l'approvazione di una proposta quale il *No Fakes Act* del 2024 costituirebbe «il passo successivo perfetto» della disciplina esistente, poiché capace di offrire rimedi anche «alle vittime di danni invasivi che vanno oltre le immagini intime affrontate da quella legislazione»[403]. Il disegno di legge ha l'obiettivo di attribuire alle persone un nuovo «diritto di replica digitale» federale, volto a impedire che versioni generate dall'IA della propria voce o della propria immagine siano condivise da terzi senza

[403] Queste le dichiarazioni di Mitch Glazier, CEO della *Recording Industry Association of America*. Cfr. S. Parvini, *Industry leaders urge Senate to protect against AI deepfakes with No Fakes Act*, AP News, 2025. Reperibile al sito: https://apnews.com/article/no-fakes-act-ai-artificial-intelligence-ec07483bac26818116b9b5a1713fe250

autorizzazione[404]. Nel dettaglio, tale diritto è esteso fino a settanta anni dopo la morte dell'interessato, potendo quest'ultimo (o i suoi eredi) citare in giudizio non solo chi diffonda contenuti manipolati non autorizzati, ma anche le aziende che ne rendano possibile la creazione. Inoltre, la proposta prevede una procedura di *notice and takedown*, la quale imporrebbe alle piattaforme online, una volta ricevuto un reclamo, di rimuovere o disabilitare tempestivamente l'accesso a qualsiasi presunta replica digitale, al fine di evitare l'irrogazione di sanzioni. Infine, sono presenti eccezioni per il contenuto inserito in una notizia giudicata «in buona fede» o avente ad oggetto «documentari, opere storiche, opere biografiche, commenti, studi accademici, satira o parodia»[405].

Nonostante il *No Fakes Act* abbia goduto di un più ampio sostegno da parte del Senato[406], esso non rappresenta la prima iniziativa statunitense in materia, inserendosi in un più ampio filone di proposte legislative. Tra queste, è opportuno menzionare il precedente *No AI Fraud Act*, volto a istituire un diritto federale di proprietà sull'immagine e la voce di una persona, la cui violazione si configurerebbe qualora una persona, fisica o giuridica, che «in qualsiasi modo influenzi il commercio interstatale o estero» renda disponibile al pubblico un servizio di clonazione personalizzata, diffonda consapevolmente repliche vocali o rappresentazioni digitali non autorizzate, o contribuisca materialmente alla

[404] J. Coleman, *The NO FAKES Act is a real threat to free expression*, FIRE, 2025. Reperibile al sito: https://www.thefire.org/news/no-fakes-act-real-threat-free-expression

[405] *Ibidem*. Le esenzioni previste necessitano di alcune considerazioni. Anzitutto, la presenza di una notizia «in buona fede» implica che la replica digitale in esso inserita sarà coperta dall'eccezione solo qualora costituisca «l'oggetto della notizia» o sia «materialmente rilevante» per essa. Tuttavia, ciò comporterebbe non solo una necessaria sottoposizione degli articoli giornalistici a pareri legali che accertino se il contenuto è in buona fede, ma anche a un incremento delle azioni giudiziarie da parte di politici o personaggi pubblici nei confronti dei giornalisti. In secondo luogo, una rappresentazione digitale inclusa in documentari o in opere storiche o biografiche risulterebbe comunque illegittima nel caso in cui queste ultime creino una falsa percezione della loro autenticità: per l'autore, ciò costituirebbe un vero e proprio paradosso, dal momento in cui la verosimiglianza costituisce caratteristica essenziale di un *deepfake*.

[406] J.E. Rothman, *House Introduces Its Companion Version of NO FAKES Act*, Rothman's Roadmap to the Right of Publicity, 2024. Reperibile al sito: https://rightofpublicityroadmap.com/news_commentary/house-introduces-its-companion-version-of-no-fakes-act/

realizzazione delle condotte indicate[407]. La durata di tale diritto è limitata a dieci anni dopo la morte dell'interessato mentre, sul piano sanzionatorio, è previsto un risarcimento danni di 50.000 dollari per i servizi di clonazione e di 5.000 dollari per le singole rappresentazioni e repliche vocali digitali, in alternativa al risarcimento dei danni effettivi e dei profitti conseguiti[408]. La grande differenza che emerge tra il *No Fakes Act* e il *No AI Fraud Act* è che il primo tutela l'identità in quanto tale, rilevando giuridicamente il semplice utilizzo non autorizzato della replica digitale, mentre il secondo considera illecito l'utilizzo dell'identità altrui soltanto in presenza di una condotta qualificata e inserita in un contesto economicamente rilevante.

Passando oltre il contesto degli Stati Uniti, è necessario fare riferimento al sistema giuridico della Danimarca che, il 26 giugno del 2025, ha annunciato una proposta di modifica della disciplina nazionale sul diritto d'autore, volta al raggiungimento di due obiettivi. Da una parte, è prevista una protezione specifica, rivolta agli artisti interpreti o esecutori, da imitazioni non autorizzate delle loro opere creative o delle performance generate tramite sistemi di IA[409]. Dall'altra, vi è la tutela per il pubblico contro le riproduzioni digitali realistiche senza consenso. Il nuovo art. 73a garantirebbe ai cittadini danesi il diritto d'autore in relazione «al proprio corpo, ai propri lineamenti del viso e alla propria voce», consentendo loro di richiedere alle piattaforme online la rimozione di contenuti pubblicati illecitamente: la riforma introdurrebbe così uno dei profili più innovativi, non essendo ad oggi presente, nel panorama europeo, un'analoga

[407] J.E. Rothman, *House's Draft AI Bill Risks Loss of Control over Our Own Voices and Likenesses*, Rothman's Roadmap to the Right of Publicity, 2024. Reperibile al sito: https://rightofpublicityroadmap.com/news_commentary/houses-draft-ai-bill-risks-loss-of-control-over-our-own-voices-and-likenesses/

[408] C. Mcsherry, *The No AI Fraud Act Creates Far More Problems Than It Solves*, EFF, 2024. Reperibile al sito: https://www.eff.org/deeplinks/2024/01/no-ai-fraud-act-creates-way-more-problems-it-solves

[409] M. Bryant, *Denmark to tackle deepfakes by giving people copyright to their own features*, The Guardian, 2025. Reperibile al sito: https://www.theguardian.com/technology/2025/jun/27/deepfakes-denmark-copyright-law-artificial-intelligence

previsione[410]. Inoltre, la violazione di tale diritto consentirebbe ai cittadini il ricorso ai rimedi risarcitori previsti dalla disciplina danese in materia di responsabilità civile, mentre non sarebbero previste sanzioni penali per gli autori della diffusione. Diversamente, le piattaforme tecnologiche che, a seguito di una segnalazione di contenuti illeciti ai sensi del Digital Services Act, non intervenissero tempestivamente, sarebbero multate dall'Autorità di vigilanza nazionale o dalla Commissione europea[411]. Ad ogni modo, le autorità danesi hanno indicato la prima metà del 2026 come termine indicativo per l'approvazione dell'emendamento all'attuale disciplina sul *copyright*.

Un altro ordinamento giuridico meritevole di menzione è quello spagnolo, il cui governo, nel gennaio del 2026, ha approvato una proposta di legge che vieta qualsiasi rappresentazione artificiale di una persona senza il suo consenso, anche qualora essa persegua finalità commerciali o pubblicitarie, nonché nei casi in cui l'immagine o la voce illecitamente utilizzate siano state «generate *ex novo*» da sistemi di IA[412]. È previsto inoltre un rafforzamento della tutela dei minori, stabilendo a sedici anni l'età minima per prestare il consenso all'uso della propria immagine. Similmente ad alcune delle precedenti iniziative considerate, quella spagnola prevede esenzioni per le rappresentazioni di carattere creativo, satirico o di finzione[413]. La proposta iberica, pur avendo un chiaro intento regolatorio, è in una fase ancora embrionale, essendo necessarie la consultazione pubblica e il

[410] G. Santoni, *Deepfake, la Danimarca fa scuola: tutele a persone e copyright*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/mercati-digitali/deepfake-diritti-della-persona-e-copyright-il-modello-danese/#_ftnref1

[411] N. Jaiswal, *'Your face, your rights': Denmark's tough new Deepfake law could change how AI imitations are handled across Europe*, Indiatimes, 2025. Reperibile al sito: https://www.indiatimes.com/news/new-danish-law-slaps-tech-platforms-with-penalties-as-citizens-gain-full-rights-over-ai-generated-clones-performances-663505.html

[412] G. Laganà, *La Spagna anticipa le regole europee e blocca i deepfake non consensuali*, Byte Legali, 2026. Reperibile al sito: https://www.bytelegali.it/la-spagna-anticipa-le-regole-europee-e-blocca-i-deepfake-non-consensuali/

[413] J. Calero, *Spain moves to curb AI deepfakes, tighten consent rules on images*, Reuters, 2026. Reperibile al sito: https://www.reuters.com/business/media-telecom/spain-moves-curb-ai-deepfakes-tighten-consent-rules-images-2026-01-13/

passaggio in parlamento per la sua eventuale trasformazione in un atto giuridicamente vincolante.

Considerate alcune delle più rilevanti iniziative dei singoli stati, l'analisi può ora estendersi alle soluzioni prospettate a livello sovranazionale per la futura regolazione dei *deepfakes*.

Anzitutto, l'Unione Europea intende regolare il fenomeno non limitandosi all'adozione di strumenti di *hard law* quali l'AI Act, ma anche ricorrendo ad atti di *soft law* con funzione integrativa dei primi. In particolare, nel dicembre del 2025, l'Ufficio per l'IA della Commissione europea ha pubblicato la prima bozza del «Codice di condotta sulla marcatura e l'etichettatura dei contenuti generati dall'intelligenza artificiale», prevedendo di ultimarlo entro il mese di giugno dell'anno successivo[414]. Obiettivo del presente Codice è agevolare il rispetto degli obblighi di trasparenza per fornitori e utilizzatori di sistemi di IA, previsti dall'art. 50 dell'AI Act (v. *supra*, cap. III, 1.3.). L'attuale bozza si compone di due sezioni: la prima ha ad oggetto gli obblighi di marcatura e rilevamento dei «contenuti audio, immagine, video o testuali sintetici», applicabili ai *providers*; la seconda riguarda il dovere di etichettatura per i *deployers* di sistemi che generano o manipolano «immagini o contenuti audio o video che costituiscono un deepfake». In sostanza, il Codice intende offrire a fornitori e utilizzatori uno strumento volontario che consenta sia di rispettare i relativi obblighi di trasparenza, sia di conformarvisi prima del 2 agosto 2026, data della loro entrata in vigore[415].

---

[414] *Commission publishes first draft of Code of Practice on marking and labelling of AI-generated content*, European Commission, 2025. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/news/commission-publishes-first-draft-code-practice-marking-and-labelling-ai-generated-content

[415] *Code of Practice on marking and labelling of AI-generated content*, European Commission, 2026. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content#ecl-inpage-Marking-and-labelling-of-ai-generated-content. Interessante è notare come, nel processo di redazione del Codice, l'Ufficio per l'IA stia coinvolgendo, accanto a organizzazioni del settore pubblico e privato, esperti accademici e altre figure qualificate, proprio fornitori e utilizzatori di sistemi di IA, in seguito a una fase consultiva avviata con gli stessi e conclusasi il 9 ottobre 2025. Cfr. *Commission launches consultation to develop guidelines and Code of Practice on transparent AI systems*, European Commission,

Da ultimo, è pertinente accennare alla Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto. Adottata dal Consiglio d'Europa il 17 maggio 2024, essa volge a garantire che «le attività svolte nell'ambito del ciclo di vita dei sistemi di intelligenza artificiale» siano «pienamente coerenti con i diritti umani, la democrazia e lo Stato di diritto»[416], mediante il rispetto di principi fondamentali quali la dignità umana, l'uguaglianza, la non discriminazione, il diritto alla privacy e la trasparenza, favorendo al contempo il progresso tecnologico e l'innovazione. Peculiarità della Convenzione quadro è l'adozione di un approccio neutrale verso la tecnologia, giustificato dalla volontà di preservarne l'efficacia a fronte della rapida e imprevedibile evoluzione dei sistemi di IA[417].

Tale fonte normativa si applica, in relazione alle attività prima evidenziate, alle «autorità pubbliche o da soggetti privati che agiscono per loro conto»[418]. Inoltre, pur non menzionando espressamente i *deepfakes*, presenta alcune disposizioni che possono applicarsi indirettamente a questi ultimi. Anzitutto, l'art. 8 impone a ciascun contraente di garantire, nel corso del ciclo di vita dei sistemi di IA, requisiti di trasparenza e vigilanza, inclusa «l'identificazione dei contenuti generati dai sistemi di intelligenza artificiale»[419]. Il successivo art. 16 ha ad oggetto i «rischi posti dai sistemi di intelligenza artificiale», rispetto ai quali sorge per i contraenti l'obbligo di adottare misure per la loro identificazione, valutazione, prevenzione e mitigazione, nonché la facoltà di

2025. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-develop-guidelines-and-code-practice-transparent-ai-systems

[416] Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 1(1)

[417] *The Framework Convention on Artificial Intelligence*, The Council of Europe, 2026. Reperibile al sito: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificialintelligence

[418] Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 3(1)(a)

[419] Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 8

introdurre una moratoria, un divieto o altre «misure appropriate»[420]. La disposizione assume rilievo in questa sede poiché, mediante interpretazione sistematica, si consente di ricondurre in tali rischi anche il fenomeno dei *deepfakes*.

La Convenzione quadro, redatta dai 46 Stati membri del Consiglio d'Europa e da tutti gli Stati osservatori, è stata aperta alla firma[421] il 5 settembre 2024. La sua entrata in vigore è subordinata al decorso di un periodo di tre mesi dalla ratifica di almeno cinque Stati firmatari, di cui almeno tre membri del Consiglio d'Europa[422]: ciò avvenuto, si configurerebbe un momento storico, potendo assistere alla nascita del «primo trattato vincolante al mondo sull'intelligenza artificiale»[423].

2. Profili etici

A conclusione del presente elaborato, è possibile sviluppare alcune riflessioni sui *deepfakes* che trascendano il solo piano giuridico.

In particolare, occorre domandarsi se il fenomeno integri un illecito *ab origine*, cioè sin dal momento della sua creazione, o se possa coesistere con un'utilizzazione eticamente corretta. Preferibile la seconda opzione. È indubbio che, nella maggior parte dei casi, un contenuto sintetico sia generato senza

[420] Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 16

[421] Consultando il sito *web* ufficiale del Consiglio d'Europa, è possibile osservare in tempo reale quali siano gli Stati firmatari della Convenzione quadro, ad oggi fermi a diciassette (inclusa l'Unione Europea). Cfr. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

[422] S. De La Feld, *Al Consiglio d'Europa nasce il primo trattato sull'intelligenza artificiale. Von der Leyen: "Legge Ue diventa globale"*, Eunews, 2024. Reperibile al sito: https://www.eunews.it/2024/09/05/europa-trattato-intelligenza-artificiale/

[423] B.I. Marcu, *The World's First Binding Treaty on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: Regulation of AI in Broad Strokes*, The Future of Privacy Forum, 2024. Reperibile al sito: https://fpf.org/blog/the-worlds-first-binding-treaty-on-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-regulation-of-ai-in-broad-strokes/

l'autorizzazione del soggetto rappresentato e, di conseguenza, classificato come illegale. Tuttavia, tale qualificazione può essere superata non solo dal consenso dell'interessato, ma anche guardando alla finalità perseguita di volta in volta[424]: si pensi, a titolo di esempio, all'applicazione di un *deepfake* in ambito cinematografico per la ricostruzione digitale di un attore defunto, o alle ipotesi di liceità, previste in diverse fonti normative già esaminate, qualora l'utilizzo avvenga a scopo di critica, satira, parodia o semplice intrattenimento[425].

Posto che la presenza di un *deepfake* non costituisce di per sé motivo di illiceità, è necessario individuare quali fattori, oltre i già citati consenso e finalità lecita, rendano la sua applicazione eticamente giustificabile. Nel 2019, la Commissione europea ha pubblicato gli «Orientamenti etici per un'intelligenza artificiale affidabile» secondo i quali un'IA deve rispettare i principi di legalità, eticità e robustezza, intesi rispettivamente come conformità alla legislazione vigente, rispetto dei valori etici e coerenza con il «suo ambiente sociale»[426]. Inoltre, affinché un sistema di IA possa definirsi «affidabile», si richiede che questi tre principi siano tradotti in «requisiti concreti», tra cui la sorveglianza umana, la sicurezza, la *privacy* e *governance* dei dati, la trasparenza, la diversità e non discriminazione, il benessere sociale e ambientale, nonché l'*accountability*[427]. Per l'analisi svolta, il documento assume grande rilevanza in quanto, pur richiamando il principio di legalità, evidenzia come la sola

[424] *I "deepfake" possono essere etici?*, Il Post, 2022. Reperibile al sito: https://www.ilpost.it/2022/10/23/deepfake-etici/

[425] *Ibidem*. Si faccia l'esempio di "deeptomcruise", profilo di TikTok nel quale sono presenti *deepfake video* ritraenti il sosia del noto attore Tom Cruise. I contenuti ivi presenti, la cui funzione è puramente ludica, rendono chiara la loro natura non autentica, rappresentando secondo l'autore una rappresentazione etica dei deepfakes. Per approfondimenti, cfr. S. Nover, *Can a deepfake company be ethical?,* Quartz, 2022. Reperibile al sito: https://qz.com/2123102/can-deepfakes-be-ethical

[426] *Ethics guidelines for trustworthy AI*, European Commission, 2019. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[427] *Ibidem*. Il documento specifica che tale elenco di requisiti non è esaustivo, tenendo conto dei più rilevanti «aspetti sistemici, individuali e sociali».

regolamentazione normativa non sia sufficiente a garantire un approccio etico ai *deepfakes*[428].

È tuttavia doverosa una considerazione finale. Sebbene la necessità dei requisiti sopra indicati sia pacifica, è fondamentale che alla loro base si collochi un «nuovo senso di responsabilità globale»[429], da tradurre sia nella sensibilizzazione sui rischi connessi al fenomeno, sia nell'alfabetizzazione tecnologica generalizzata per contrastarne la diffusione e i danni conseguenti.

[428] D. Giribaldi, *Deepfake video: ecco perché serve una nuova coscienza etica sull'intelligenza artificiale*, Agenda Digitale, 2019. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/deepfake-video-ecco-perche-serve-una-nuova-coscienza-etica-sullintelligenza-artificiale/

[429] *Ibidem*

CONCLUSIONI

L'analisi condotta nei capitoli precedenti consente di formulare alcune considerazioni finali. Il principale interrogativo affrontato dal presente elaborato riguarda l'esistenza, allo stato attuale, di strumenti normativi capaci di regolare adeguatamente i *deepfakes*. Il terzo capitolo risulta cruciale poiché emerge come la maggior parte dei sistemi giuridici mondiali, pur avendo introdotto atti normativi direttamente applicabili al fenomeno, presenti una disciplina in materia ancora insufficiente. Si pensi all'ordinamento statunitense, il cui *Take It Down Act*, unico intervento federale riferibile ai *deepfakes*, circoscrive questi ultimi alla sola dimensione pornografica, oltre a presentare numerose lacune. Una simile incertezza rileva nel panorama europeo, nel quale l'*AI Act* non qualifica chiaramente la natura dei contenuti manipolati artificialmente[430], limitandosi ad assoggettarli agli obblighi di trasparenza di cui all'art. 50.

Altro fattore da considerare è che tali atti normativi risultano incompleti sebbene siano entrati in vigore recentemente (addirittura, l'*AI Act* assumerà piena operatività solo nel 2027). Il quadro descritto risulta ancor più complesso in quanto, nel frattempo, i sistemi di IA e le loro applicazioni continuano a evolversi in tempi assai rapidi. È dunque possibile trarre un'altra fondamentale conclusione: la velocità con cui si evolve la tecnologia, e in particolare i *deepfakes*, è superiore a quella necessaria al legislatore per introdurre una regolamentazione efficace.

Ciononostante, il presente scritto ha individuato nella Cina l'ordinamento giuridico capace di distinguersi per rapidità e completezza nella disciplina del fenomeno. Come già osservato (v. *supra*, cap. III, 3.), la scelta del governo cinese di allontanarsi dal modello *multistakeholder*, tipico dei sistemi occidentali, per andare incontro al principio della *cyber-sovranity* (*wǎngluò zhǔquán*) si è tradotta nella concentrazione esclusiva in capo allo Stato della *governance* di Internet e

[430] Si è infatti osservato (v. *supra*, cap. III, 1.3.) che l'AI Act evita di qualificare i *deepfakes* tra i sistemi ad alto o a basso rischio, e che solo una ricostruzione dottrinale ne ha attribuito una collocazione "intermedia". Cfr. A. Orlando, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

delle tecnologie emergenti. Di conseguenza, sin dalle prime fasi di sviluppo dei *deepfakes*, la Cina ha tempestivamente adottato un approccio mirato, attraverso operazioni come: la loro inclusione nella più ampia categoria dei *deep synthesis*, l'imposizione di obblighi stringenti non solo ai fornitori, ma anche agli utenti e alle piattaforme di distribuzione, nonché l'introduzione di un sistema di etichettatura articolato in forma «esplicita» e «implicita».

Tuttavia, l'approccio cinese, pur avendo messo in discussione la tradizionale *leadership* dell'Unione Europea in ambito regolatorio, evidenzia il pericolo che un opprimente controllo statale possa di fatto limitare l'innovazione tecnologica.

Concludendo, l'attuale disciplina dei *deepfakes* ruota attorno a due approcci diametralmente opposti, nessuno dei quali pare in grado di risolvere tutte le complessità del fenomeno, e tantomeno di elevarsi a risposta definitiva. Il futuro della regolazione dei contenuti sintetici dovrà dunque prescindere dal modello occidentale, lento e incompleto, e da quello cinese, rapido ma fortemente invasivo, per andare incontro a nuove soluzioni.

Tra queste, sembra persuasiva una prospettiva "ibrida", capace di resistere all'evoluzione tecnologica mediante norme tecnologicamente neutre, e di legare tempestività e proporzionalità degli interventi attraverso un costante aggiornamento e una graduazione delle misure a seconda del livello di rischio.

A questo punto, non resta che osservare le ricadute pratiche delle normative vigenti per comprendere se nuove proposte regolatorie dovranno essere prese in considerazione e se l'attuale posizione di "rincorsa" dei sistemi occidentali rispetto a quello cinese resterà immutata o condurrà a una ridefinizione degli attuali equilibri.

BIBLIOGRAFIA

D. ABERCROMBIE, *Elements of general phonetics*, Edinburgh University Press, Edinburgh, 1967

D. AGRILLO, E. AMEDEO, A. DI NOBILE, C. TARALLO, *L'immagine nel mondo, il mondo nell'immagine: nuove prospettive per un approccio pluridisciplinare alla rappresentazione testuale ed extra-testuale*, Quaderni della ricerca – 1, UniorPress, Napoli, 2016

G.W. ALLPORT, H. CANTRIL, *Judging Personality from Voice*, in Journal of Social Psychology, n. 5, 1934

A. APOSTOLI, *La dignità sociale come orizzonte della uguaglianza nell'ordinamento costituzionale*, in Costituzionalismo.it, n. 3, 2019

G. AZZARITI, *Contro il revisionismo costituzionale*, Editori Laterza, Roma-Bari, 2016

P. BECCHI, *Dignità umana*, in U. POMARICI, *Filosofia del diritto. Concetti fondamentali*, Giappichelli, Torino, 2007

C. BEGLER, M. COVELL, M. SLANEY, *Video Rewrite: driving visual speech with audio, in SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, 1997

H. BELTING, Antropologia delle immagini, Carocci editore, Roma, 2011

E. BEZZECCHI, *Intelligenza artificiale: Farsi le domande giuste, capire gli scenari futuri e usare in modo smart l'IA generativa*, Vallardi, Milano, 2024

G. BUONOMO, G. CIACCI, A. COSTANZO, *Macchine intelligenti e diritto*, Giappichelli, Torino, 2025

L. CARLASSARE, *Nel segno della Costituzione. La nostra carta per il futuro*, Feltrinelli, Bergamo, 2012

A. CATAUDELLA, *Dignità, decoro e identità personale*, in Il Diritto dell'informazione e dell'informatica, n. 2, 1985

G. CIACCI, *Firme grafometriche e tutela dei dati personali*, in Rivista elettronica di Diritto, Economia, Management, n. 1, 2014

G. CIACCI, G. BUONOMO, *Profili di informatica giuridica*, Wolters Kluwer, Milano, 2021

F. CIOTTI, C. MORABITO, *La narrazione come incontro*, Firenze University Press, Firenze, 2022

D.K. CITRON, R. CHESNEY, *Deep fakes: A looming challenge for privacy, democracy, and national security*, in California Law Review, n. 107, 2019

V. CODELUPPI, *Il ruolo sociale delle immagini digitali*, in Mediascapes Journal, n. 19, 2022

K. COOPER, L.G.E. SMITH, A.J. RUSSELL, *Gender Identity in Autism: Sex Differences in Social Affiliation with Gender Groups*, in Journal of Autism and Developmental Disorders, vol. 48, 2018

A. CRUTTENDEN, *Intonation*, Cambridge University Press, Cambridge, 1997

A. DE CUPIS, *La verità nel diritto (Osservazioni in margine a un libro recente)*, in Foro italiano, n. 4, 1952

A. DE CUPIS, *Bilancio di un'esperienza*, in AA. VV., *La lesione dell'identità personale e il danno non patrimoniale. Atti del seminario promosso dal centro di iniziativa giuridica P. Calamandrei. Messina 16 aprile 1982*, Giuffrè, Milano

A. DE RUITER, *The Distinct Wrong of Deepfakes*, in Philosophy & Technology, n. 34, 2021

R. DEBRAY, *Vita e morte dell'immagine (trad. it. A. Pinotti)*, Editrice Il Castoro, Milano, 1999

P. DEL CORE, *Rapporto tra identità e cultura, una coordinata essenziale per «comprendere l'altro» nella prospettiva dell'intercultura*, in EDUCA, n. 2, 2016

G.P. DOLSO, *Dignità, eguaglianza e Costituzione*, EUT Edizioni Università di Trieste, Trieste, 2019

H. ECKERT, J. LAVER, *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*, Beltz/Psychologie Verlags Union, Weinheim, 1994

E. H, ERIKSON, *Young Man Luther: A Study in Psychoanalysis and History*, Norton & Company, New York, 1993

J. FAY, W.C. MIDDLETON, *Judgement of Kretschmerian Body Types on the Voice Transmitted over a Public Address System*, in Journal of Social Psychology, n. 12, 1940

S.C. FERNÁNDEZ, *El «deepfake» en los documentos audiovisuales: un reto para la valoración de la prueba*, in Actualidad Civil, n. 2, 2024

C. FREGLY, A. BARTH, S. EIGENBRODE, *Generative AI on AWS*, O' Reilly Media, Sebastopol, 2025

Y. GENG, *Comparing "Deepfake" Regulatory Regimes in The United States, the European Union, and China*, in Georgetown Law Technology Review, vol. 7, 2023

C. GENOVA, *Il cerchio nello spazio. Ipotesi e strumenti per un'analisi della ri–significazione dei luoghi*, in Lexia. Rivista di semiotica 9-10, 2011

E. GOFFMAN, *La vita quotidiana come rappresentazione (trad. it. di M. Ciacci)*, Il Mulino, Bologna, 1997

S. HE, Y. LEI, Z. ZHANG, Y. SUN, S. LI, C. ZANG, J. YE, *Identity Deepfake Threats to Biometric Authentication Systems: Public and Expert Perspectives*, arXiv, 2025

M. HEIDEGGER, *Sein und Zeit*, Max Neimeyer, Tübingen, 1967

M. ŁABUZ, *Regulating Deep Fakes in the Artificial Intelligence Act*, in Applied Cybersecurity & Internet Governance, vol. 2, n. 1, 2023

P. LADEFOGED, *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech*, Oxford University Press, London, 1967

J.P. LAMONAGA, *A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes*, in American University Law Review, vol. 69, 2020

J. LAVER, *The Gift of Speech: Papers in the Analysis of Speech and Voice*, Edinburgh University Press, Edinburgh, 1991

M. S. LEE, *Entertainment and Intellectual Property Law*, Thomson Reuters, Toronto, 2023

R. LUSCHINGER, G.E. ARNOLD, *Voice, Speech, Language: Clinical Communicology: Its Physiology and Pathology*, Wadsworth Publishing, Belmont, 1965

E. MACALUSO, *Che cosa sono le immagini? Riflessioni su una questione radicale*, in Rivista interdisciplinare di comunicazione, n. 3, 2021

S. MARINO, *L'identità personale alla prova delle libertà di circolazione*, in Eurojus, n. 4, 2020

K.T. MAY, S. BRAY, T. DAVIES, L.D. GRIFFIN, *Warning: Humans cannot reliably detect speech deepfakes*, in PLOS One, n. 18 (8), 2023

J. T. MCCARTHY, R. E. SCHECHTER, *The Rights of Publicity & Privacy (2d ed.)*, Thomson Reuters, Toronto 2024

N. MENDOZA-DENTON, *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*, Blackwell, Malden, Mass., 2008

G. MONGA, *Responsabilità degli intermediari. Il Digital Services Act, in M. Maggiore (a cura di)*, Il commercio elettronico. Digital markets act, digital services act e altre dimensioni giuridiche, Giappichelli, Torino, 2024

A. MORELLI, *Persona e identità personale*, in BioLaw Journal, n. 2, 2019

F. R. MORENO, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in International Review of Law, Computers & Technology, n. 38, 2024

M. NANDAL, *Mitigating Deepfake Threats to Privacy: Legal Frameworks and Technological Safeguards*, in Proceedings of the National Seminar on Enhancing Privacy Protection in the Digital Age: Legal Challenges & Innovations, 2025

J. NIEVA-FENOLL, *La prueba de los deepfakes pornográficos: I.A. sobre I.A.*, in Diario La Ley, n. 10516, 2024

A. ORLANDO, *La regolamentazione del deepfake in Europa, Stati Uniti e Cina*, in Media Laws, n. speciale I, 2024

S. PARINARDI, J. CROSSON, K. PETERSON, S. NADAREVIC, *Investigating the politics and content of US State artificial intelligence legislation*, in Business and Politics, n. 2, 2024

B. PAYE HERRERO, *Voice and Identity: A contrastive study of identity perception in voice, Universitat de València*, Servei de Publicacions, Valencia, 2009

A. PEROTTI, *La via obbligata dell'interculturalità*, Editrice Missionaria Italiana, Città di Castello (Perugia), 1994, p. 18

R. PETRI, *Identità: parabole politiche di un concetto*, in Memoria e Ricerca, n. 1, 2023

S. PEZZANO, *L'immagine digitale. Una vera-falsa "nuova immagine"*, in Leitmotiv, n. 4, 2004

G. PINO, *Il diritto all'identità personale. Interpretazione costituzionale e creatività giurisprudenziale*, Il Mulino, Bologna, 2003

G. PINO, *L'identità personale*, in AA.VV., *Gli interessi protetti nella responsabilità civile, vol. II*, Utet, Torino, 2005

G. PINO, *Teorie e dottrine dei diritti della personalità. Uno studio di meta-giurisprudenza analitica*, in Materiali per una storia della cultura giuridica, n. 1, 2003

F. PIZZETTI, *Con AI Verso la Società digitale*, in Federalismi, n. 23, 2023

G. PROIETTI, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del deep fake*, in Media Laws, n. speciale I, 2024

S. PUGLISI, C. VERGARI, *Il fenomeno del deepfake nell'attuale ordinamento giuridico*, Trento Biolaw Selected Student Papers, T4F Series, n. 1, 2023

A. RANDAZZO, *Diritto all'identità personale e valori costituzionali. Le linee di un modello, traendo spunto da Luigi Pirandello*, in Diritti Fondamentali, n. 3, 2021

F. REMOTTI, *Contro l'identità*, Editori Laterza, Roma-Bari, 2001

G. RESTA, *Identità personale e identità digitale*, in Diritto dell'Informazione e dell'Informatica, n. 3, 2007

P. RICOEUR, *Sé come un altro (a cura di D. Iannotta)*, Jaca Book, Milano, 2005

R. RINI, *Deepfakes and the Epistemic Backstop*, in Philosophers' Imprint, n. 20, 2020

S. ROCCAS, M. B. BREWER, Social identity complexity, in Personality and Social Psychology Review, n. 6, 2002

K. RUBIN, *The Key to Being a Good Referee: The Call the Ninth Circuit Missed When Evaluating False Endorsement Claims*, in Emory Law Journal, vol. 64, 2015

M. RUBINI, E. CROCETTI, *Una o più identità? Articolando identità personale e sociale*, in Psicologia sociale, n. 1, 2018

B. RUSSELL, *The Collected Papers of Bertrand Russell, Volume 11. Last Philosophical Testament 1947-68*, Routledge, London, 1997

I. SALVADORI, *La controvertida relevancia penal del sexting en el derecho italiano y comparado*, in Revista Electrónica de Ciencia Penal y Criminología, n. 19 (29), 2017

G. SANTONI, *La Cina e lo spazio digitale. Questioni di governance nello spazio digitale globale*, in OrizzonteCina, vol. 11, 2020

F. SANZARI, *Reputazione, diritti della personalità, internet*, in S. Previti, Internet e reputazione aziendale, Utet giuridica, Torino, 2016

M.M. SAZ RUBIO, B. PENNOCK-SPECK, *Male and Female Stereotypes in Spanish and British Commercials*, in M. Muñoz-Calvo, C. Bueso-Gómez, M. Á. Ruiz-Moneva, New Trends in Translation and Cultural Identity

T. SCHECHNER, *Gender identity disorder: a literature review from a developmental perspective*, Israel Journal of Psychiatry and Related Sciences, vol. 47 (2), 2010

A. SCHIRMER, *Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing*, in Social Cognitive and Affective Neuroscience, n. 13, 2018

J. SOLOMON, A. SOLOMON, N. JOSEPH, S. NORTON, *Impression management, myth creation and fabrication in private social and environmental reporting: Insights from Erving Goffman, in in Accounting, Organizations and Society*, vol. 38, n. 3, 2013

T. SPLAGOUNIAS, *Rethinking the right of publicity in the era of generative AI, in IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property*, vol. 65, n. 1, 2024

H. TAJFEL, C. J. TURNER, *An integrative theory of intergroup conflict*, in W. G. AUSTIN & S. WORCHEL, *The social psychology of intergroup relations*, Brooks/Cole Publishing Company, California, 1979

A. TESTA, *Le vie del senso*, Garzanti, Milano, 2021

S. THOBANI, *Diritti della personalità e contratto: dalle fattispecie più tradizionali al trattamento in massa dei dati personali*, Ledizioni, Milano, 2018

L. TORCHIA, *Lo stato digitale. Una introduzione*, Il Mulino, Bologna, 2023

R. L. TRASK, *A dictionary of phonetics and phonology*, Routledge, London, 1996

J. C. TURNER, *Rediscovering the Social Group: A Self-Categorization Theory*, Basil Blackwell, Oxford, 1987

U.S. COPYRIGHT OFFICE, *Compendium of U.S. Copyright Office Practices, Third Edition*, 2014

U.S. COPYRIGHT OFFICE, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 2023

M. WESTERLUND, *The Emergence of Deepfake Technology: A Review*, in Technology Innovation Management Review, vol. 9, n. 11

J.-J. WUNENBURGER, *Filosofia delle immagini (trad. it. S. Arecco)*, Einaudi, Torino, 1999

T. VAN EECHOUD, P. B. HUGENHOLTZ, S. VAN GOMPEL, L. GUIBAULT, N. HELBERGER, *Harmonizing European Copyright Law: the Challenges of Better Lawmaking*, Wolters Kluwer, Alphen aan den Rijn, 2009

T. VAN LEEUWEN, Speech, music, sound, Macmillan, Basingstoke, 1999

V. ZENO-ZENKOVICH, *Identità personale*, in Digesto delle discipline privatistiche – Sez. civile, IX, Utet, Torino, 1993

SITOGRAFIA

AGENZIA PER L'ITALIA DIGITALE (AGID), *Identità digitale*, 2025. Reperibile al sito: https://www.agid.gov.it/it/ambiti-intervento/identita-digitale

AGENZIA PER L'ITALIA DIGITALE (AGID), *SPID - Sistema Pubblico di Identità Digitale*, 2025. Reperibile al sito: https://www.agid.gov.it/it/piattaforme/spid

AGENZIA PER L'ITALIA DIGITALE (AGID), *Carta Nazionale dei Servizi*, 2024. Reperibile al sito: https://www.agid.gov.it/it/piattaforme/carta-nazionale-servizi

A. BADMAN, M. KOSINSKI, *What is a dataset?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/dataset

A. BANKS, *Op-Ed. Deepfakes & Why the Future of Porn is Terrifying*, Highsnobiety, 2018. Reperibile al sito: https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/

N. BASSETTI, *GAN (Generative Adversarial Networks): cosa sono, applicazioni e vantaggi*, Agenda Digitale, 2023. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/gan-generative-adversarial-networks-cosa-sono-applicazioni-e-vantaggi/#post-175384-footnote-1

D. BENEDETTI, *Voce artificiale: gli ultimi sviluppi, anche per il business*, Agenda Digitale, 2021. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/voce-artificiale-ma-con-personalita-progressi-e-nuovi-usi-del-linguaggio-sintetico/#post-117432-footnote-1

C. BIAGIOLINI-JR., *Introduction to Transformers in Generative AI*, Medium, 2024. Reperibile al sito: https://medium.com/%40biagolini/introduction-to-transformers-in-generative-ai-195902f838c3

G. BROOKER, How does text-to-speech AI (TTS) work?, LivePerson, 2024. Reperibile al sito: https://www.liveperson.com/blog/text-to-speech-ai/

M. BRYANT, *Denmark to tackle deepfakes by giving people copyright to their own features*, The Guardian, 2025. Reperibile al sito: https://www.theguardian.com/technology/2025/jun/27/deepfakes-denmark-copyright-law-artificial-intelligence

J. CALERO, *Spain moves to curb AI deepfakes, tighten consent rules on images*, Reuters, 2026. Reperibile al sito: https://www.reuters.com/business/media-telecom/spain-moves-curb-ai-deepfakes-tighten-consent-rules-images-2026-01-13/

E. CANDOTTO, *Deepfake pornografici: le tutele previste dall'AI Act*, LexTech Hub, 2025. Reperibile al sito: https://www.lextech-hub.com/post/deepfake-pornografici-le-tutele-previste-dall-ai-act

M. CARTISANO, *Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi*, Agenda Digitale, 2023. Reperibile al sito: https://www.agendadigitale.eu/sicurezza/privacy/deepfake-porno-con-giorgia-meloni-ecco-tutti-i-reati-commessi/

M. CHEN, *What Is Machine Learning?*, Oracle, 2024. Reperibile al sito: https://www.oracle.com/artificial-intelligence/machine-learning/what-is-machine-learning/

D. CLEMENTI, *Contenuti IA: la Cina etichetta tutto, l'Europa riflette (e insegue)*, Agenda Digitale, 2025. Reperibile al sito:

https://www.agendadigitale.eu/cultura-digitale/contenuti-ia-la-cina-etichetta-tutto-leuropa-riflette-e-insegue/

L. COCCO, *Deepfake: cosa sono e come riconoscerli per smascherare la disinformazione*, Cyber Security 360, 2025. Reperibile al sito: https://www.cybersecurity360.it/nuove-minacce/deep-fake-cosa-sono-e-come-riconoscerli-per-smascherare-la-disinformazione/

S. COLE, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, Vice, 2018. Reperibile al sito: https://www.vice.com/en/article/reddit-fake-porn-app-daisy-ridley/

J. COLEMAN, *The NO FAKES Act is a real threat to free expression*, FIRE, 2025. Reperibile al sito: https://www.thefire.org/news/no-fakes-act-real-threat-free-expression

Congress.gov, *The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, 2025. Reperibile al sito: https://www.congress.gov/crs-product/LSB11314

F. DE STEFANI, *Web reputation: cos'è e come curare la reputazione online*, in Agenda Digitale, 2020. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/web-reputation-cose-e-come-curare-la-reputazione-online/

S. DE LA FELD, *Al Consiglio d'Europa nasce il primo trattato sull'intelligenza artificiale. Von der Leyen: "Legge Ue diventa globale"*, Eunews, 2024. Reperibile al sito: https://www.eunews.it/2024/09/05/europa-trattato-intelligenza-artificiale/

L. DI GIACOMO, *Deepfake e diritto: l'UE è pronta ad affrontare la minaccia dell'IA?*, Diritto.it, 2024. Reperibile al sito: https://www.diritto.it/deepfake-diritto-unione-europea-pronta-minaccia-ia/

D. DI LEO, *La responsabilità civile alla luce del GDPR e del decreto 101/2018*, Agenda Digitale, 2019. Reperibile al sito: https://www.agendadigitale.eu/sicurezza/privacy/la-responsabilita-civile-alla-luce-del-gdpr-e-del-decreto-101-2018/

DIPARTIMENTO PER LA TRASFORMAZIONE DIGITALE, *Approvata in via definitiva la legge italiana sull'Intelligenza Artificiale*, 2025. Reperibile al sito: https://innovazione.gov.it/notizie/articoli/approvata-in-via-definitiva-la-legge-italiana-sull-intelligenza-artificiale/

G. DRAGONI, *Identità Digitale: cos'è, come funziona e come si usa*, Osservatori.net, 2025. Reperibile al sito: https://www.osservatori.net/blog/digital-identity/identita-digitale-significato-funzionamento/

EUROPEAN COMMISSION, *Ethics guidelines for trustworthy AI*, 2019. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

EUROPEAN COMMISSION, *Commission publishes first draft of Code of Practice on marking and labelling of AI-generated content*, 2025. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/news/commission-publishes-first-draft-code-practice-marking-and-labelling-ai-generated-content

EUROPEAN COMMISSION, *Commission launches consultation to develop guidelines and Code of Practice on transparent AI systems*, 2025. Reperibile al sito: https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-develop-guidelines-and-code-practice-transparent-ai-systems

EUROPEAN COMMISSION, *Code of Practice on marking and labelling of AI-generated content*, 2026. Reperibile al sito: https://digital-

strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content#ecl-inpage-Marking-and-labelling-of-ai-generated-content

ELEVENLABS, *What is Voice Cloning?*, 2024. Reperibile al sito: https://elevenlabs.io/blog/what-is-voice-cloning#2-what-is-voice-cloning

D. ERVIN, J.B. STEINBERG, *AI and the Right of Publicity: A Patchwork of State Laws the Only Guidance, For Now*, Crowell, 2023. Reperibile al sito: https://www.crowell.com/en/insights/client-alerts/ai-and-the-right-of-publicity-a-patchwork-of-state-laws-the-only-guidance-for-now

J. FABER, *A Brief History of Right of Publicity (NIL)*, Right Of Publicity, 2025. Reperibile al sito: https://rightofpublicity.com/brief-history-of-rop

P. FELICIANI, *NYT contro OpenAI: la battaglia del copyright nell'era dell'IA generativa*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/mercati-digitali/nyt-contro-openai-il-futuro-del-copyright-nellera-dellia-generativa/

L. FELNER, *Take It Down Act heads to Trump's desk*, The Verge, 2025. Reperibile al sito: https://www.theverge.com/news/657632/take-it-down-act-passes-house-deepfakes

F. FERNÁNDEZ SEGADO, *La dignità della persona come valore supremo dell'ordinamento giuridico spagnolo e come fonte di tutti i diritti*, Forum di Quaderni Costituzionali, 2010. Reperibile al sito: https://www.forumcostituzionale.it/wordpress/images/stories/pdf/documenti_forum/paper/0214_segado.pdf

B. FINLEY, *Deepfake of principal's voice is the latest case of AI being used for harm*, AP News, 2024. Reperibile al sito: https://apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e

M. FRAGALE, V. GRILLI, *Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation*, Columbia Journal of European Law, 2024. Reperibile al sito: https://cjel.law.columbia.edu/preliminary-reference/2024/deepfake-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/?cn-reloaded=1

FUTURO DIGITALE, *Reato di deepfake: che cosa prevede il disegno di legge sull'IA?*, 2025. Reperibile al sito: https://futurodigitale.infocert.it/pillole-normative/reato-di-deepfake/

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*, 2020. Reperibile al sito: https://garanteprivacy.it/home/docweb/-/docweb-display/docweb/9512278

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Deepfake. Il falso che ti «ruba» la faccia (e la privacy), 2020. Reperibile al sito: file:///C:/Users/Asus/Downloads/Deepfake%20-%20Vademecum%20(3).pdf

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Deep fake: il Garante privacy apre un'istruttoria nei confronti di Telegram per il software che "spoglia" le donne, 2020. Reperibile al sito: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9470722

GEEKSFORGEEKS, *ELIZA: The First Step in Human-Computer Interaction Through Natural Language Processing*, 2025. Reperibile al sito: https://www.geeksforgeeks.org/artificial-intelligence/eliza-the-first-step-in-human-computer-interaction-through-natural-language-processing/

GEEKSFORGEEKS, *Multi-Head Attention Mechanism*, 2025. Reperibile al sito: https://www.geeksforgeeks.org/nlp/multi-head-attention-mechanism/

N. GIANSIRACUSA, *The Simple, Elegant Idea Behind Face-Swap Deepfakes*, Medium, 2021. Reperibile al sito: https://odsc.medium.com/the-simple-elegant-idea-behind-face-swap-deepfakes-456a61c73985

GIORNALETTISMO, *Differenze tra immagini digitali e analogiche: una guida completa*, 2025. Reperibile al sito: https://www.giornalettismo.com/immagine-digitale-e-immagine-analogica-differenze/

D. GIRIBALDI, *Deepfake video: ecco perché serve una nuova coscienza etica sull'intelligenza artificiale*, Agenda Digitale, 2019. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/deepfake-video-ecco-perche-serve-una-nuova-coscienza-etica-sullintelligenza-artificiale/

G. HOLDSWORTH, M. SCAPICCHIO, *What is deep learning?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/it-it/think/topics/deep-learning

IDENTITY MANAGEMENT INSTITUTE, *Deepfake Deception in Digital Identity*, 2025. Reperibile al sito: https://identitymanagementinstitute.org/deepfake-deception-in-digital-identity

IL POST, *I "deepfake" possono essere etici?*, 2022. Reperibile al sito: https://www.ilpost.it/2022/10/23/deepfake-etici/

N. JAISWAL, *'Your face, your rights': Denmark's tough new Deepfake law could change how AI imitations are handled across Europe*, Indiatimes, 2025. Reperibile al sito: https://www.indiatimes.com/news/new-danish-law-slaps-tech-platforms-with-penalties-as-citizens-gain-full-rights-over-ai-generated-clones-performances-663505.html

J. JORDAN, *Variational autoencoders*, 2018. Reperibile al sito: https://www.jeremyjordan.me/variational-autoencoders/

G. KOSINSKI, *Cos'è l'autenticazione?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/it-it/think/topics/authentication

M. KOSINSKI, M. SCAPICCHIO, *What is the EU AI Act?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/eu-ai-act

M. KOSINSKI, *What is Voice Cloning?*, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/voice-cloning

F. LA TROFA, *Avatar 3D: cos'è e come funziona la nostra identità nel metaverso*, Tech4Future, 2022. Reperibile al sito: https://tech4future.info/avatar-3d-cose-come-funziona/

G. LAGANÀ, *La Spagna anticipa le regole europee e blocca i deepfake non consensuali*, Byte Legali, 2026. Reperibile al sito: https://www.bytelegali.it/la-spagna-anticipa-le-regole-europee-e-blocca-i-deepfake-non-consensuali/

C. LISI, *Deepfake: cosa sono, come crearli e come riconoscerli*, Dogma. Reperibile al sito: https://www.dogma.it/it/news/deepfake--cosa-sono--come-crearli-e-come-riconoscerli

S. LISSENS, *The Foundations of EU personal data protection law: Privacy and Human Dignity*, in EU Renew, 2024. Reperibile al sito: https://eu-renew.eu/the-foundations-of-eu-personal-data-protection-law-privacy-and-human-dignity/

K. MAGRAMO, *British engineering giant Arup revealed as $25 million deepfake scam victim*, CNN, 2024. Reperibile al sito: https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html

A. MANTZARLIS, *The real problem with the Take It Down Act is not free speech*, Faked Up, 2025. Reperibile al sito: https://fakedup.org/the-real-problem-with-the-take-it-down-act-is-not-free-speech/

B.I. MARCU, *The World's First Binding Treaty on Artificial Intelligence*, Human Rights, Democracy, and the Rule of Law: Regulation of AI in Broad Strokes, The Future of Privacy Forum, 2024. Reperibile al sito: https://fpf.org/blog/the-worlds-first-binding-treaty-on-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-regulation-of-ai-in-broad-strokes/

A. MARTÍNEZ, *Take it Down Act, la ley que criminaliza los deepfakes y la 'pornovenganza'*, El País, 2025. Reperibile al sito: https://elpais.com/us/2025-05-20/take-it-down-act-la-ley-que-criminaliza-los-deepfakes-y-la-pornovenganza.html

M. MARTORANA, *Deepfake, quando l'AI manipola la realtà: il diritto in affanno*, Agenda Digitale, 2025. Reperibile al sito: https://www.agendadigitale.eu/cultura-digitale/deepfake-quando-lai-manipola-la-realta-il-diritto-in-affanno/

C. MCSHERRY, *The No AI Fraud Act Creates Far More Problems Than It Solves*, EFF, 2024. Reperibile al sito: https://www.eff.org/deeplinks/2024/01/no-ai-fraud-act-creates-way-more-problems-it-solves

MINISTERO DELL'INTERNO, *Carta di Identità Elettronica*, 2025. Reperibile al sito: https://www.cartaidentita.interno.gov.it/la-carta/

MM ONE, *AI per Avatar e Voci Sintetiche: il futuro della comunicazione video*, 2023. Reperibile al sito: https://www.mm-one.com/intelligenza-artificiale/ai-per-avatar-e-voci-sintetiche-il-futuro-della-comunicazione-video/

C. NASS, *Machine voices*, 2005. Reperibile al sito: https://www.pbs.org/speak/ahead/technology/voiceinterface/#

C. NEGRI, *Cos'è il Deep Learning: esempi e applicazioni reali*, Osservatori.net, 2025. Reperibile al sito: https://www.osservatori.net/blog/artificial-intelligence/deep-learning-significato-esempi-applicazioni/

S. NOVER, *Can a deepfake company be ethical?*, Quartz, 2022. Reperibile al sito: https://qz.com/2123102/can-deepfakes-be-ethical

S. PARVINI, *Industry leaders urge Senate to protect against AI deepfakes with No Fakes Act*, AP News, 2025. Reperibile al sito: https://apnews.com/article/no-fakes-act-ai-artificial-intelligence-ec07483bac26818116b9b5a1713fe250

F. PINCELLI, AI generativa e analisi predittiva: perché non se ne può più fare a meno, BizzIT, 2025. Reperibile al sito: https://bizzit.it/approfondimenti/ai-generativa-e-analisi-predittiva-perche-non-se-ne-puo-piu-fare-a-meno/

A. POPE, *NYT v. OpenAI: The Times's About-Face*, Harvard Law Review, 2024. Reperibile al sito: https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/

E.C. RAFFIOTTA, *Appunti in materia di diritto all'identità personale*, Forum di Quaderni Costituzionali, 2010. Reperibile al sito: https://www.forumcostituzionale.it/wordpress/images/stories/pdf/documenti_forum/paper/0173_raffiotta.pdf

G. RANGHI, *Violenza digitale: la diffusione illecita di immagini o video sessualmente espliciti*, Salvis Juribus, 2022. Reperibile al sito: https://www.salvisjuribus.it/violenza-digitale-la-diffusione-illecita-di-immagini-o-video-sessualmente-espliciti/

G. REGAN, *A Brief History of Deepfakes*, Reality Defender, 2024. Reperibile al sito: https://www.realitydefender.com/insights/history-of-deepfakes

R. RIJTANO, *Deepfake su Telegram, c'è un bot che spoglia le donne: quasi 700 mila vittime*, Repubblica, 2020. Reperibile al sito: https://www.repubblica.it/tecnologia/sicurezza/2020/10/20/news/deepfake_un_bot_sveste_le_donne_oltre_100mila_vittime-271131375/

E. ROSATI, *Copyright in the Digital Single Market: a taster*, WIPO, 2021. Reperibile al sito: https://www.wipo.int/web/wipo-magazine/articles/copyright-in-the-digital-single-market-a-taster-42399

J.E. ROTHMAN, *House Introduces Its Companion Version of NO FAKES Act*, Rothman's Roadmap to the Right of Publicity, 2024. Reperibile al sito: https://rightofpublicityroadmap.com/news_commentary/house-introduces-its-companion-version-of-no-fakes-act/

J.E. ROTHMAN, *House's Draft AI Bill Risks Loss of Control over Our Own Voices and Likenesses*, Rothman's Roadmap to the Right of Publicity, 2024. Reperibile al sito: https://rightofpublicityroadmap.com/news_commentary/houses-draft-ai-bill-risks-loss-of-control-over-our-own-voices-and-likenesses/

B. SAETTA, *Dati biometrici*, Protezione dati personali, 2024. Reperibile al sito: https://protezionedatipersonali.it/dati-biometrici

P. SANDONNINI, *Modelli di diffusione, cosa sono e perché offrono nuove possibilità per la creatività generativa*, AI4Business, 2022. Reperibile al sito: https://www.ai4business.it/news/modelli-di-diffusione-cosa-sono-e-perche-offrono-nuove-possibilita-per-la-creativita-generativa/

G. SANTONI, *Deepfake, la Danimarca fa scuola: tutele a persone e copyright*, Agenda Digitale, 2025. Reperibile al sito:

https://www.agendadigitale.eu/mercati-digitali/deepfake-diritti-della-persona-e-copyright-il-modello-danese/#_ftnref1

G. SCORZA, *Giù le mani da quelle voci: sono frammenti della nostra identità personale*, StartupItalia, 2025. Reperibile al sito: https://startupitalia.eu/tech/giu-le-mani-da-quelle-voci-sono-frammenti-della-nostra-identita-personale/

F. B. SCHNEIDER, *Authentication of People*, Cornell University, 2005. Reperibile al sito: https://www.cs.cornell.edu/courses/cs513/2005fa/NNLauthPeople.html

D. SHAH, *Generative AI 101: Explanation, Use Cases, Impact*, V7 Labs, 2023. Reperibile al sito: https://www.v7labs.com/blog/generative-ai-guide

T. SHAHASANE, *Deepfakes vs. AI Avatars: A Personal Take on What Sets Them Apart*, Gan.AI, 2024. Reperibile al sito: https://www.gan.ai/blog/posts/deepfakes-vs-ai-avatars-a-personal-take-on-what-sets-them-apart

R. SHELDON, *What is a data set?*, TechTarget, 2024. Reperibile al sito: https://www.techtarget.com/whatis/definition/data-set

J. SHRAGER, *ELIZA Reinterpreted: The world's first chatbot was not intended as a chatbot at all*, arXiv, 2024. Reperibile al sito: https://arxiv.org/pdf/2406.17650

I. SIBELLA, *Identità narrativa: racconto o significato?*, in Psicologia Fenomenica, 2019. Reperibile al sito: https://www.psicologiafenomenologica.it/identita-narrativa-raccontosignificato/

C. STRYKER, M. SCAPICCHIO, What is generative AI?, IBM, 2024. Reperibile al sito: https://www.ibm.com/think/topics/generative-ai

A. TEGHO, C. WILLIAMS, *The dilemma of the deepfake: DeepFake Inc. and data protection*, Mills & Reeve, 2025. Reperibile al sito: https://www.mills-reeve.com/publications/the-dilemma-of-the-deepfake-deepfake-inc-and-data-protection/

THE COUNCIL OF EUROPE, *The Framework Convention on Artificial Intelligence*, 2026. Reperibile al sito: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificialintelligence

THE EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: Main capabilities and scientific disciplines*, 2018. Reperibile al sito: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

V. TIANI, *Il New York Times vincerà la causa contro OpenAI? Un'analisi*, la Repubblica, 2024. Reperibile al sito: https://www.repubblica.it/tecnologia/blog/digitaleurope/2024/01/19/news/il_new_york_times_vincera_la_causa_conro_openai_unanalisi-421918307/

L. TURINI, *Fermare i deepfake*, Appunti – di Stefano Feltri, 2025. Reperibile al sito: https://appunti.substack.com/p/fermare-i-deepfake

U.S. COPYRIGHT OFFICE, *Copyright Law of the United States (Title 17) and Related Laws Contained in Title 17 of the United States Code*, 2025. Reperibile al sito: https://www.copyright.gov/title17/

U.S. COPYRIGHT OFFICE, *Overview of the Copyright Office*, 2025. Reperibile al sito: https://www.copyright.gov/about/

U.S. COPYRIGHT OFFICE, *Introduction to the Third Edition of the Compendium Of U.S. Copyright Office Practices*, 2021. Reperibile al sito: https://www.copyright.gov/comp3/docs/introduction.pdf

U.S. COPYRIGHT OFFICE, *Copyright Office Launches New Artificial Intelligence Initiative*, 2023. Reperibile al sito: https://www.copyright.gov/newsnet/2023/1004.html

U.S. GOVERNMENT ACCOUNTABILITY OFFICE, *Science & Tech Spotlight: Deepfakes*, 2020. Reperibile al sito: https://www.gao.gov/products/gao-20-379sp

D. USERA, *Communicating to Connect: Interpersonal Communication for Today*, LibreTexts, 2025. Reperibile al sito: https://socialsci.libretexts.org/Bookshelves/Communication/Introduction_to_Communication/Communicating_to_Connect__Interpersonal_Communication_for_Today_(Usera)/03%3A_Identity/3.02%3A_Self-Image

C. VANBERGHEN, *The AI Act vs. deepfakes: A step forward, but is it enough?*, Euractiv, 2024. Reperibile al sito: https://www.euractiv.com/opinion/the-ai-act-vs-deepfakes-a-step-forward-but-is-it-enough/

M. VIGARANI, *"Giù la maschera": fenomenologia e rischi del Self Impression Management nel processo di Recruitment*, Meliusform, 2020. Reperibile al sito: https://www.meliusform.it/giu-la-maschera-fenomenologia-e-rischi-del-self-impression-management-nel-processo-di-recruitment.html

WHISPERLY, *Deepfake Regulation: Chaos That Matters Now More Than Ever*, 2025. Reperibile al sito: https://whisperly.ai/deepfake-regulation/

WORKINVOICE, *Identità Digitale (Digital Identity): Cos'è, Come Funziona e gli Strumenti in Italia*, 2024. Reperibile al sito: https://www.workinvoice.it/identita-digitale

GIURISPRUDENZA

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)

CORTE COSTITUZIONALE, sent. 2 giugno 1994, n. 218

CORTE COSTITUZIONALE, sent. 3 febbraio 1994, n. 13

CORTE COSTITUZIONALE, sent. 22 ottobre 1999, n. 388

CORTE COSTITUZIONALE, sent. 17 luglio 2000 n. 293

CORTE DI CASSAZIONE, sez. I Civile, sent. 18 ottobre 1984, n. 5259

CORTE DI CASSAZIONE, sez. V Penale, sent. 28 febbraio 1995, n. 3247

CORTE DI CASSAZIONE, Sez. I civile, sent. 7 febbraio 1996, n. 978

CORTE DI CASSAZIONE, sez. III Civile, sent. 9 aprile 1998, n. 3679

CORTE DI CASSAZIONE, sez. V Penale, sent. 17 novembre 2000, n. 4741

CORTE DI CASSAZIONE, sez. III Civile, sent. 10 maggio 2001, n. 6507

CORTE DI CASSAZIONE, sez. I Civile, sent. 19 maggio 2020, n. 9147

CORTE EDU, sent. 22 febbraio 1994, ric. 16213/90, *Burghartz c. Svizzera*

CORTE EDU, sent. 25 novembre 1994, ric. 18131/91, *Stjerna c. Finlandia*

CORTE EDU, sent. 24 giugno 2004, ric. 59320/00, *Von Hannover v. Germany*

NORMATIVA

Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 1

Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 10

Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 12

Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 13

Administrative Regulations on Online Audio and Video Information Services (网络音视频信息服务管理规定), art. 11

California Civil Code, Section 3344, lett. a)

California Civil Code, Section 3344, lett. d)

California Civil Code, Section 3344, lett. e)

California Civil Code, Section 3344, lett. f)

California Civil Code, Section 3344.1, lett. a), 1), A)

California Civil Code, Section 3344.1, lett. a), 2), A), i)

California Civil Code, Section 3344.1, lett. a), 2), B), ii)

California Civil Code, Section 3344.1, lett. g)

Codice Civile (R.D. 16 marzo 1942, n. 262), art. 10

Codice dell'Amministrazione Digitale (D.lgs. n. 82/2005), art. 1, lett. u-quater)

Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 61, n. 11-undecies)

Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 612-quater

Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 640-quater 1, co. 1

Codice Penale (R.D. 19 ottobre 1930, n. 1398), art. 640-quater 1, co. 2

Communications Act of 1934 (47 U.S.C.), sez. 223, h), 1), A)

Communications Act of 1934 (47 U.S.C.), sez. 223, h), 2), C)

Communications Act of 1934 (47 U.S.C.), sez. 223, h), 3), B)

Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 1(1)

Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 3(1)(a)

Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 8

Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti umani, la Democrazia e lo Stato di diritto, art. 16

Copyright Act of 1976, 17 U.S.C., sez. 107

Direttiva 2001/29/CE, art. 2

Direttiva 2001/29/CE, art. 3

Direttiva 2001/29/CE, art. 4

Direttiva 2001/29/CE, art. 5

Direttiva (UE) 2019/790, art. 1

Direttiva (UE) 2019/790, art. 3

Direttiva (UE) 2019/790, art. 4

Regolamento (UE) 2022/2065, art. 3, lett. h)

Regolamento (UE) 2022/2065, art. 6

Regolamento (UE) 2022/2065, art. 8

Regolamento (UE) 2022/2065, art. 14

Regolamento (UE) 2022/2065, art. 16

Regolamento (UE) 2022/2065, art. 17

Regolamento (UE) 2022/2065, art. 20

Regolamento (UE) 2022/2065, art. 21

Regolamento (UE) 2022/2065, art. 22

Regolamento (UE) 2022/2065, art. 34

Regolamento (UE) 2022/2065, art. 35

Regolamento (UE) 2022/2065, art. 35, par. 1, lett. k)

Regolamento (UE) 2022/2065, art. 40

Regolamento (UE) 2022/2065, Considerando n. 87

Regolamento (UE) 2024/1689, art. 3, n. 60

Regolamento (UE) 2024/1689, art. 5, par. 1, lett. a)

Regolamento (UE) 2024/1689, art. 5, par. 1, lett. b)

Regolamento (UE) 2024/1689, art. 6, par. 3

Regolamento (UE) 2024/1689, art. 50, par. 1

Regolamento (UE) 2024/1689, art. 50, par. 2

Regolamento (UE) 2024/1689, art. 50, par. 3

Regolamento (UE) 2024/1689, art. 50, par. 4

Regolamento (UE) 2024/1689, art. 99, par. 3

Regolamento (UE) 2024/1689, art. 99, par. 4

Regolamento (UE) 2024/1689, Allegato III, n. 8, lett. b)

Regolamento (UE) 2024/1689, art. 50, par. 1, lett. a), b), d)

Regolamento (UE) 2024/1689, art. 50, par. 2

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 6

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 9

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 12

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 14

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 15

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 17

Provisions on the Administration of Deep Synthesis Internet Information Services, art. 23

Take It Down Act (Public Law 119-12), sez. 3, a), 1), A)

Take It Down Act (Public Law 119-12), sez. 3, a), 2)

Take It Down Act (Public Law 119-12), sez. 3, a), 3)

Take It Down Act (Public Law 119-12), sez. 3, a), 4)

Take It Down Act (Public Law 119-12), sez. 3, b)