T4F Series
PAPER N. 10
a.a. 2023/2024

Trento BioLaw Selected Student Papers

Deepfake e pornografia: percezione del fenomeno e aspetti normativi in chiave comparata

MICHELA D'ERCOLE, ELEONORA DI BONAVENTURA, IRIS DI GIACOMO

Jean Monnet EUROPEAN CENTRE UNIVERSITY OF TRENTO
Co-funded by the Erasmus+ Programme of the European Union
UNIVERSITÀ DI TRENTO Facoltà di Giurisprudenza

I paper sono stati selezionati a conclusione del corso *Diritto e Intelligenza Artificiale* a.a. 2023-2024, organizzato all'interno della Cattedra Jean Monnet "T4F – TrAIning 4 Future. Artificial Intelligence and EU Law", coordinato presso l'Università di Trento dal docente Carlo Casonato.

Deepfake e pornografia:
percezione del fenomeno e aspetti normativi in chiave comparata

Michela D'Ercole, Eleonora Di Bonaventura, Iris di Giacomo[*]

ABSTRACT: L'articolo si pone l'obiettivo di sviscerare il significato, l'origine e l'evoluzione del fenomeno *deepfake*. Si analizzano, in particolare, i casi che hanno portato all'emersione e diffusione dello stesso. Particolare attenzione è rivolta alla percezione comune di questo fenomeno. A ciò ha contribuito una ricerca condotta dalle stesse autrici (sotto forma di sondaggio), che ha avuto ad oggetto un campione di 200 persone. Segue l'analisi normativa interna del *deepfake*, tra approcci volti ad un'applicazione analogica delle norme già vigenti, e una controtendenza che mira alla creazione di una disciplina *ex novo*. Si conclude con l'analisi normativa del fenomeno in chiave comparata.

PAROLE CHIAVE: Deepfake; Pornografia; Intelligenza Artificiale; Normativa sull'IA; Comparazione

DEEPFAKE AND PORNOGRAPHY: PERCEPTION OF THE PHENOMENON AND REGULATORY ASPECTS IN A COMPARATIVE PERSPECTIVE

ABSTRACT: This paper aims to eviscerate the meaning, origin and evolution of the deepfake phenomenon, analyzing the cases that led to its emergence and spread. Also, particular attention is directed to the common perception of the phenomenon, shown by a survey edited by the authors themselves, involving 200 people of different age groups. Then, the script focuses on the regulatory aspects of Italian policy, analyzing two different tendencies: one directed to extend the meaning of the rules already in force, the other one directed to create a whole new discipline. Finally, it concludes with a comparative analysis of the deepfake regulation.

KEYWORDS: Deepfake; Pornography; Artificial Intelligence; AI Regulation; Comparison

1. Il fenomeno del *deepfake*, tra illusione e realtà

Il fenomeno della divulgazione di contenuti noti come *deepfake*, pur avendo conosciuto enorme diffusione a livello globale, sembrerebbe tutt'oggi circondato da un'alea di dubbi e incertezze. Alcuni dati elaborati ai fini del presente scritto[1] attestano come la conoscibilità e la conoscenza del fenomeno siano riservate in

[*] *Studentesse dell'Università di Trento, Facoltà di Giurisprudenza. Email: michela.dercole@studenti.unitn.it, e.dibonaventura@studenti.unitn.it, iris.digiacomo@studenti.unitn.it* (contributo paritario).

[1] I dati circa la consapevolezza del fenomeno *deepfake* sono stati raccolti da un questionario sottoposto a 200 campioni a cura delle autrici, disponibile al link: https://forms.office.com/Pages/AnalysisPage.aspx?AnalyzerToken=scQXLlENFFiuaWFFtaYV8UcSLn7GgzlH&id=DQSIkWdsW0yxEjajBLZtrQAAAAAAAAAAAO__VRRZKlUODI5NlpGWldCM0xSMkQ1RlJXNk03OEs3SS4u. (ultima consultazione 27/09/2024).

misura preponderante ai cultori della materia, mentre per le persone "laiche" siano oggetto di incomprensioni, alle quali la ricerca e il dibattito in tema di intelligenza artificiale tentano di sopperire.

Per meglio delineare la dimensione - amplissima e variegata, seppur piuttosto recente - dei contenuti *deepfake*, è utile darne definizione: il neologismo, composto dai termini *deep learning* e *fake*, descrive il fenomeno della generazione o alterazione di contenuti multimediali attraverso *software* di intelligenza artificiale, in grado di riprodurre fedelmente e in modo estremamente realistico le caratteristiche e i comportamenti di un volto o di un corpo umano[2]. Com'è stato rilevato, si tratta di una tendenza destinata a dilagare senza sosta: si attesta che in appena quattro anni, dal 2019 al 2023, il numero di *deepfake* creati, ed eventualmente messi in circolazione attraverso i mezzi di comunicazione di massa, sia cresciuto del 550%[3]. La realizzazione dei *deepfake* è resa possibile da una particolare forma di tecnica di apprendimento applicata all'IA, il cosiddetto *deep learning*, il quale, nell'elaborazione degli input per mezzo di complesse reti neurali, simula le caratteristiche del cervello umano[4]. È proprio in questa caratteristica che risiedono i rischi della diffusione dell'utilizzo dell'IA, trattandosi di un'intelligenza che si pone l'ambizioso obiettivo di ricreare fedelmente e simulare le connessioni neuronali.

Un dato preoccupante relativo al fenomeno è da individuarsi nell'insufficiente presenza di regolamentazioni adeguate al rischio costituito dai video generati dall'IA, specie per quanto riguarda le piattaforme quotidianamente frequentate come i social media e le piattaforme di *streaming*.

La prima piattaforma social ad aver adottato misure di tutela e informazione a riguardo è *Tiktok*: le linee guida[5] comprendono una sezione dedicata ai «Contenuti multimediali modificati e contenuti generati dall'intelligenza artificiale (AIGC)», sottoposti all'obbligo di apposizione di un'apposita etichetta AIGC, ovvero mediante una didascalia, una filigrana o un adesivo personali qualora mostrino scene o persone in maniera realistica. La politica di *Tiktok* si spinge anche oltre, prevedendo una lista ad hoc di comportamenti vietati in relazione ai media sintetici, ottenuti con IA, vietando in particolare la riproduzione di persone minori dei 18 anni o di individui privati che non abbiano prestato il loro consenso alla creazione del contenuto, nonché di figure pubbliche, qualora i contenuti riguardino il contesto politico o riproducano atteggiamenti contrari alla legge o alla società. Sulla stessa scia, nell'aprile 2024 anche *Meta* ha annunciato l'introduzione di etichette per i contenuti creati con l'intelligenza artificiale su *Facebook*, *Instagram* e *Threads*[6].

---

[2] Cfr. GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Vademecum Deepfake Il falso che ti «ruba» la faccia (e la privacy)*, dicembre 2020, disponibile al link https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9512278 (ultima consultazione 27/09/2024).

[3] REDAZIONE ANSA, *Deepfake, oltre 900 esperti chiedono una regolamentazione*, in www.ansa.it, 23 febbraio 2024, disponibile al link: https://www.ansa.it/osservatorio_intelligenza_artificiale/notizie/societa/2024/02/23/deepfake-oltre-900-esperti-chiedono-una-regolamentazione_12e01844-0c1f-4241-98fd-d64668b51b08.html (ultima consultazione 27/05/2024).

[4] Y. LECUN, Y. BENGIO, G. HINTON, *Deep learning, in Nature*, 521, 2015, pp. 436-444, disponibile al link: https://doi.org/10.1038/nature14539 (ultima consultazione 27/09/2024).

[5] In https://www.tiktok.com, disponibile al link: https://www.tiktok.com/community-guidelines/it/integrity-authenticity#3 (ultima consultazione 27/09/2024).

[6] In www.meta.it, disponibile al link: https://www.meta.com/it-it/help/artificial-intelligence/recognize-digitally-altered-content/ (ultima consultazione 27/05/2024).

2. La percezione del fenomeno

Al fine di accertare la concreta conoscenza del fenomeno e la sua percezione da parte dell'opinione pubblica è stata condotta una ricerca, sotto forma di sondaggio, che ha coinvolto 200 persone, dai 18 anni fino ad oltre i 60 anni di età. Dai risultati dell'indagine è emerso come solo il 64% di esse sia a conoscenza del fenomeno, mentre un buon 36% ammette di non averne mai sentito parlare. In particolare, si è appurato come la conoscenza del *deepfake* diminuisca con l'aumentare dell'età delle persone che hanno partecipato all'indagine. Infatti, sebbene all'interno del campione di età compresa tra i 18-29 anni e 30-40 anni le percentuali di persone consapevoli del fenomeno risultino essere rispettivamente del 71,4% e del 60%, all'interno delle fasce di età più avanzate, comprese tra i 41-60 anni e dai 61 anni in su, la percentuale di persone informate dello stesso si ferma rispettivamente a 37,5% e 33,3%, costituendo invece un'ampia maggioranza - pari al 62,5% e 66,6% - le persone ignare dello stesso.
L'ampia inconsapevolezza che circonda il *deepfake* è conseguentemente seguita da una scarsa percezione circa la diffusione dello stesso. Di fatto, si ferma a 5,78 - su una scala da 0 a 10 - la media di punteggio delle 200 risposte alla domanda: «Quanto pensi sia dilagante questo fenomeno su una scala da 0 a 10?». Rispetto a questo risultato, decisamente poco soddisfacente, è il dato percentuale già ricordato nel precedente paragrafo che attesta un aumento della diffusione dei *deepfake* del 550% tra il 2019 e il 2023: un fenomeno tutt'altro che poco dilagante, come invece emerge nella percezione comune del campione esaminato. Un ulteriore dato raccolto, corollario del precedente, è dato dal basso timore che il campione ha in relazione all'ipotesi che i loro contenuti personali possano essere manipolati dall'IA, e quindi divenire oggetto di *deepfake*: su una scala da 1 a 10, il timore di distorsione della propria immagine è pari ad un mero 6,58[7]. Una statistica che rispecchia ben poco l'effettiva ed esponenziale diffusione del fenomeno, da considerare non come un potenziale minaccia diretta alle sole celebrità, un tempo maggiormente colpite, ma come un pericolo che incombe, potenzialmente, su ognuno di noi. A tal proposito risultano fondamentali i dati statistici, di cui al paragrafo successivo, che confermano come il *deepfake* sia un fenomeno che coinvolge soprattutto persone "comuni".

3. I potenziali utilizzi, positivi e non, del deepfake

La connotazione, prevalentemente negativa, che è attribuita al *deepfake* nella prassi trova indubbiamente motivo di esistere nell'origine del fenomeno, il quale prende avvio in un contesto tutt'altro che positivo. Nel 2017, un utente del social media *Reddit* ottiene grande popolarità a seguito della pubblicazione di una serie di video pornografici falsi, che vedono come protagonisti note personalità statunitensi, dal mondo della politica alle attrici hollywoodiane[8]. L'associazione del *deepfake* a un fenomeno negativo spesso causa di gravi lesioni del diritto alla personalità legate in larga parte agli scopi cui esso si presta, permane anche in

[7] I dati fanno riferimento al sondaggio sopra menzionato: https://forms.office.com/Pages/AnalysisPage.aspx?AnalyzerToken=scQXLlENFFiuaWFFtaYV8UcSLn7GgzlH&id=DQSIkWdsW0yxEjajBLZtrQAAAAAAAAAAAO__VRRZKlUODI5NlpGWldCM0xSMkQ1RlJXNk03OEs3SS4u.

[8] R. A. DELFINO, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, in *Fordham Law Review*, 3, 2019, disponibile al link: https://ir.lawnet.fordham.edu/flr/vol88/iss3/2/ (ultima consultazione 30/09/2024).

tempi più recenti: secondo alcuni dati statistici[9], il 96% della totalità dei *deepfake* generati è a contenuto pornografico *non consensuale* (per i quali è stata coniata l'espressione *deep-porn* o *deep-nude*) e il 100% di essi vede le donne come protagoniste degli abusi. Inoltre, il 70% di queste sono donne comuni, e non celebrità, che in una certa misura sono protette dalla notorietà stessa. Di fatto, è più probabile che saremmo indotti a ritenere falso un *deepfake* pornografico che ritrae un personaggio pubblico, piuttosto che di un individuo privato di cui non conosciamo il nome, né le abitudini o lo stile di vita.
Un'ulteriore - e doverosa - precisazione riguarda la difficoltà oggettiva dei sistemi *deepfake* di riprodurre un corpo nudo maschile: a ciò consegue inevitabilmente la settorializzazione della problematica al solo genere femminile[10]. Il fatto che il fenomeno abbia ad oggetto una così ampia disparità tra i soggetti colpiti spiegherebbe anche lo squilibrio di genere in relazione al grado di inquietudine che il fenomeno del *deep-porn* ingenera. All'interno del sondaggio realizzato è stato infatti posto il seguente quesito: «Su una scala da 0 a 10, quanto timore hai che i tuoi contenuti possano essere manipolati con IA?». I risultati raccolti attestano come ci sia un maggior timore all'interno del campione femminile, rispetto a quello maschile. Difatti, all'interno del primo il 69% teme che i propri contenuti possano essere oggetto di manipolazione, contro il 55% del campione maschile. Nonostante l'alto grado di rischio costituito dall'attività manipolativa posta in atto per mezzo dell'IA, di fatto non devono, né possono, essere trascurate le potenzialità di un utilizzo positivo dei *deepfake*. Un esempio degno di nota è rappresentato dal già menzionato *Tiktok* e la sua politica sui «media manipolati», la quale consente la pubblicazione di video *deepfake* ritraenti personaggi pubblici con finalità artistiche o educative, qualora adeguatamente etichettati (così come prevede lo stesso articolo 50, paragrafo 2 della proposta dell'AI Act)[11].
Gli esempi di un utilizzo creativo del *deepfake* sono molteplici e si riscontrano nei più diversi scenari che vanno dal mondo del cinema alla produzione di *spot* pubblicitari fino a giungere ai più banali video o vignette (i c.d. "*meme*") creati a scopo satirico e umoristico. Così, ad esempio, nella serie televisiva "*A Killer Paradox*", grazie all'IA è stato possibile ottenere un fedelissimo ringiovanimento degli attori per la realizzazione di scene ambientate nel passato; oppure, nella campagna di sensibilizzazione contro la malaria il volto del protagonista, David Beckham, è modificato nelle diverse versioni tradotte, adeguando i movimenti della bocca alla lingua parlata[12].

4. Casistica del fenomeno

Sfortunatamente, non sono rari i casi in cui questa tecnologia viene sfruttata per fini deprecabili. Innumerevoli celebrità sono state vittime di *deepfake* utilizzati per la creazione di falsi filmati pornografici. Incredibile risonanza mediatica ha avuto il caso di Taylor Swift, cantautrice e attrice statunitense, che nel gennaio 2024 è diventata bersaglio di alto profilo di immagini *deepfake* sessualmente esplicite, nonché non

[9] H. AJDERL, G. PATRINI, F. CAVALLI, L. CULLEN, *The State of Deepfakes*, in *Sensity*, 2019, pp. 7-8, disponibile al link: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf (ultima consultazione 29/09/2024).
[10] S. COLE, *This Horrifying App Undresses a Photo of Any Woman With a Single Click*, 2019, disponibile al link https://www.vice.com/en/article/deepnude-app-creates-fake-nudes-of-any-woman/ (ultima consultazione: 27/09/2024).
[11] In https://www.tiktok.com, disponibile su: https://www.tiktok.com/community-guidelines/it/integrity-authenticity#3 (ultima consultazione: 27/09/2024).
[12] V. AZZALI, N. ELLECOSTA, *La questione deepfake in Italia: una panoramica*, in *Media Laws – Rivista di Diritti dei Media*, 3, 2023, p. 77.

consensuali, realizzate utilizzando l'intelligenza artificiale. Le immagini, generate dall'intelligenza artificiale, sono state viste decine di milioni di volte: secondo quanto ricostruito dal New York Times, le immagini diffuse sulla piattaforma *X* sarebbero state viste fino a 47 milioni di volte prima che diversi account coinvolti venissero sospesi dal social network, ponendo un freno alla divulgazione delle stesse. Tuttavia, come riporta ancora il quotidiano newyorkese, i contenuti sono stati condivisi anche su altre piattaforme, continuando così a circolare sulle stesse, nonostante gli sforzi delle compagnie per rimuoverli[13]. Sul caso, riporta il Guardian, è intervenuta anche la portavoce della Casa Bianca, Karine Jean-Pierre, definendo «allarmanti» le immagini *fake* e affermando che le piattaforme social hanno la responsabilità di prevenire la circolazione di questo tipo di contenuti. In aggiunta, diversi politici statunitensi hanno lanciato appelli per condannare la pratica di creare immagini sintetiche realistiche a sfondo sessuale[14].

Ciononostante, un aspetto certamente positivo che consegue alla diffusione di *deepfake* di personalità note è dato dall'aumento esponenziale dell'attenzione del pubblico sul fenomeno del *deepfake*, con un'amplificazione delle richieste di intervento da parte dei legislatori. Perfettamente in linea con la tendenza del fenomeno del *deepfake* sono le parole di Mary Anne Franks, professoressa alla George Washington University Law School e presidente della Cyber Civil Rights Initiative: «*Siamo troppo pochi, troppo tardi a questo punto, ma possiamo ancora provare a mitigare il disastro che sta emergendo*». Aggiunge: «*Le donne sono canarini nella miniera di carbone quando si tratta di abuso dell'intelligenza artificiale*» e che «*non si tratterà solo della quattordicenne o di Taylor Swift. Saranno i politici. Saranno i leader mondiali. Saranno le elezioni*[15]». Nelle medesime dinamiche si è ritrovata recentemente anche una cantante italiana, Rose Villain, che nel marzo 2024 ha denunciato la circolazione di immagini *deepfake*[16]. Un episodio, l'ennesimo, di una serie già nutrita, che alza il livello di allarme sulle potenziali conseguenze dell'intelligenza artificiale generativa e di come questa, essendo accessibile a tutti, si presti alla creazione di abusi come *deepfake* sempre più difficili da decifrare. Di fatto, finché la vittima non si rende conto della manipolazione, la rete diventa il bacino di immagini illecite che possono ledere dignità e la reputazione della persona offesa con conseguenze gravi, soprattutto per i giovanissimi. Per questo la stessa Rose Villain, nelle parole di denuncia pubblicate sulle sue piattaforme social, si è soffermata sulle persone più fragili e inconsapevoli che non sanno come accedere agli strumenti per difendersi.

[13] K. CONGER, J. YOON, *Explicit Deepfake Images of Taylor Swift Elude Safeguards and Swamp Social Media*, in *The New York Times*, 2024, disponibile al link: https://www.nytimes.com/2024/01/26/arts/music/taylor-swift-ai-fake-images.html (ultima consultazione 27/09/2024).

[14] K. VINER, *Taylor Swift searches blocked on X after fake explicit images of pop singer spread*, in *The Guardian*, 2024, disponibile al link: https://www.theguardian.com/music/2024/jan/28/taylor-swift-x-searches-blocked-fake-explicit-images (ultima consultazione 27/09/2024).

[15] Cfr. ABC News Live, *Pornographic deepfakes target women across the country*, 2024, disponibile al link: https://www.law.gwu.edu/pornographic-deepfakes-target-women-across-country (ultima consultazione 27/09/2024).

[16] C. BARISON, *Rose Villain vittima di deepfake che la ritraggono nuda: «È violenza»*, in *Corriere della Sera*, 2024, disponibile al link: https://www.corriere.it/tecnologia/24_aprile_04/rose-villain-vittima-di-deepfake-che-la-ritraggono-nuda-e-violenza-b500bcc3-3dac-4827-a8b4-05550a44bxlk.shtml (ultima consultazione 27/09/2024).

5. Deep-porn e revenge porn: analogie e divergenze

Nella maggior parte dei casi, il *deepfake* viene sfruttato in contesti patologici che finiscono per rendere la tecnologia, di per sé non di natura incriminante, uno strumento funzionale alla commissione di reati. La sfida principale risiede nel qualificare giuridicamente lo strumento tecnologico per ricomprenderlo all'interno delle norme dell'ordinamento, con finalità sanzionatorie qualora questo venisse sfruttato per finalità illecite. Lo strumento del *deepfake*, infatti, può essere combinato con condotte lesive, come il reato di *revenge porn*, e portare all'inquadramento di nuove fattispecie, colmando l'attuale assordante silenzio normativo. Il termine *revenge porn*, anch'esso un neologismo, indica la realizzazione o sottrazione, e successiva condivisione pubblica, di immagini o video sessualmente espliciti senza il consenso delle persone rappresentate. Per questa fattispecie non è necessario uno sforzo interpretativo analitico pari a quello dell'inquadramento del *deepfake*: di fatto, tenendo conto della tempestiva espansione del fenomeno, nel 2019 in Italia è stata introdotta a opera del c.d. Codice Rosso[17] un'apposita fattispecie penale, la quale, in aggiunta, prevede un aumento di pena nel caso in cui i fatti siano commessi attraverso strumenti informatici o telematici[18]. Lo stretto legame tra un fenomeno ancora privo di limiti come quello del *deepfake* e una fattispecie delittuosa ben delineata come quella del revenge porn consiste nel fondato timore che la diffusione e lo sviluppo delle tecnologie di *deepfake* spianerà la strada a una divulgazione incontrollata della pornografia non consensuale. Già nel 2019, uno studio dell'American Psychological Association ha riscontrato che una donna su dodici, ad un certo punto della propria vita, può essere vittima di revenge porn[19].

È proprio alla luce dei numerosissimi punti di contatto tra le due ipotesi, quelle del *deep-porn* e del *revenge porn*, che la dottrina si è interrogata circa la possibilità di ricomprendere nell'alveo della tutela dell'art. 612-*ter* anche il *deepfake* a sfondo pornografico[20]. Il confronto tra le assonanze e dissonanze tra le due fattispecie ha però portato ad un esito negativo, dal momento che la lettera dell'art. 612-*ter* sembrerebbe far emergere la necessità che i contenuti pornografici in questione siano anzitutto reali, e in secondo luogo

[17] Legge 19 luglio 2019, n. 69, Modifiche al Codice penale, al codice di procedura penale e altre disposizioni in materia di tutela delle vittime di violenza domestica e di genere.

[18] Art. 612-*ter*, «Diffusione illecita di immagini o video sessualmente espliciti»: «Salvo che il fatto costituisca più grave reato, 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 la multa da euro 5.000 a euro 15.000. La stessa pena si applica a chi, avendo ricevuto o comunque acquisito le immagini o i video di cui al primo comma, li invia, consegna, cede, pubblica o diffonde senza il consenso delle persone rappresentate al fine di recare loro nocumento. La pena è aumentata se i fatti sono commessi dal coniuge, anche separato o divorziato, o da persona che è o è stata legata da relazione affettiva alla persona offesa ovvero se i fatti sono commessi attraverso strumenti informatici o telematici. La pena è aumentata da un terzo alla metà se i fatti sono commessi in danno di persona in condizione di inferiorità fisica o psichica o in danno di una donna in stato di gravidanza. Il delitto è punito a querela della persona offesa. Il termine per la proposizione della querela è di sei mesi. La remissione della querela può essere soltanto processuale. Si procede tuttavia d'ufficio nei casi di cui al quarto comma, nonché quando il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio».

[19] F.M.R. LIVELLI, *Deepfake e Revenge Porn, come Combatterli con la Cultura Digitale*, in *CyberSecurity360*, 2021, disponibile al link: https://www.cybersecurity360.it/nuove-minacce/deepfake-e-revenge-porn-combatterli-con-la-cultura-digitale-ecco-come/ (ultima consultazione 30/09/2024).

[20] M. MARTORANA, Z. SICHI, *Deepfake e Revenge Porn: Punti di Contatto*, in *Altalex*, 2021, disponibile al link: https://www.altalex.com/documents/news/2021/06/03/deepfake-e-revenge-porn-punti-di-contatto (ultima consultazione: 27/09/2024).

realizzati o sottratti da parte di una persona fisica, rimanendo esclusa la possibilità di manipolazione da parte dell'intelligenza artificiale[21]. A ben vedere, dalle considerazioni finora esposte sembrerebbe delinearsi uno scenario piuttosto sinistro, in cui l'entità dei fenomeni di *revenge porn* connessi all'utilizzo di *deepfake* appare destinata ad aumentare. Ciononostante, a difesa del progresso tecnologico consapevole, si pone la c.d. teoria dell'effetto contrario: questa enuncia come, nonostante l'atteggiamento spontaneo dinanzi a immagini o video riprodotti online sia quello di ritenere vero ciò che viene rappresentato, in uno scenario dove un'immagine o un video generati artificialmente siano indistinguibili da fotografie o video reali e altrettanto (se non addirittura maggiormente) diffusi, la nostra attitudine sarebbe, invece, più scettica[22].
Alcuni filosofi hanno osservato il fenomeno in modo analitico. A titolo di esempio, il professore di filosofia e informatica alla Northeastern University degli Stati Uniti, Don Fallis, ha notato come, in un mondo dove i video cc.dd. *fake* diventano la norma, le persone tenderanno sempre di più a diffidare anche dei video reali, indebolendo il valore informativo di questi ultimi[23]. Questo indebolimento, seppur problematico per il giornalismo, potrebbe invece costituire un possibile vantaggio nel campo della diffusione non consensuale di contenuti pornografici. È vero che chiunque potrebbe facilmente contraffare e diffondere una fotografia o un video di carattere pornografico, ritraendoci nudi ovvero in atteggiamenti intimi: ed è proprio questa facilità che farebbe sì che il materiale così prodotto non risulterebbe troppo credibile. Anzi, potrebbe essere la stessa facilità di contraffare siffatte immagini a indurci a ritenere non veritieri i video o le fotografie che non sono frutto di interpolazione, bensì reali[24].

6. I tentativi di tutela dell'ordinamento italiano

Per quanto concerne la tutela nei confronti dei video generati da intelligenza artificiale da parte dell'ordinamento italiano, la questione è tanto attuale e dibattuta quanto acerba: se da un lato si avverte la necessità di introdurre *ex novo* una fattispecie di reato a sé stante, dall'altro la dottrina[25] cerca di interpretare in via analogica alcune norme già esistenti, nonostante il divieto di analogia *legis* in materia penale. Il dibattito dottrinale[26] ha formulato numerose proposte, in aggiunta al 612-*ter* già esaminato *supra*. In primo luogo, è stata richiamato l'art. 595 del codice penale[27], riguardante il reato di diffamazione

[21] *Ibidem*.
[22] M. VIOLA, *Il deepfake può anche aiutare nella lotta al revenge porn,* in *Agenda Digitale*, 2023, disponibile al link: https://www.agendadigitale.eu/sicurezza/privacy/deepfake-arginare-revenge-porn/ (ultima consultazione 28/09/2024).
[23] D. FALLIS, *The Epistemic Threat of Deepfakes*, in *Philosophy & Technology*, 4, 2021, pp. 623-6432.
[24] M. VIOLA, *op. cit.*
[25] D. POLIDORO, *Come ci si Può Difendere se si è Vittima di Deepfake Porno, come Rose Villain*, in *Wired*, 2024, disponibile al link: https://www.wired.it/article/deepfake-rose-villain-come-proteggersi-denuncia-reati/ (ultima consultazione 26/09/2024).
[26] *Ivi*.
[27] Art. 595, «Diffamazione»: «Chiunque, fuori dei casi indicati nell'articolo precedente, comunicando con più persone, offende l'altrui reputazione, è punito con la reclusione fino a un anno o con la multa fino a milletrentadue euro. Se l'offesa consiste nell'attribuzione di un fatto determinato, la pena è della reclusione fino a due anni, ovvero della multa fino a duemilasessantacinque euro. Se l'offesa è recata col mezzo della stampa o con qualsiasi altro mezzo di pubblicità, ovvero in atto pubblico, la pena è della reclusione da sei mesi a tre anni o della multa non inferiore a cinquecentosedici euro. Se l'offesa è recata a un Corpo politico, amministrativo o giudiziario, o ad una sua rappresentanza, o ad una Autorità costituita in collegio, le pene sono aumentate».

quale offesa all'altrui reputazione, aggravato se i fatti sono commessi «con qualsiasi mezzo di pubblicità»: tuttavia, pur costituendone un elemento indubbiamente centrale, la norma di cui al 595 c.p. non è di per sé sufficiente a compendiare la complessità della fattispecie del *deep-porn*, che necessiterebbe di una sanzione *ad hoc* nei confronti di chi vi contravviene. Un'ulteriore ipotesi, certamente più affine alla fattispecie in esame dal punto di vista della descrizione del fatto tipico, è l'art. 640-ter[28], ossia il reato di frode informatica, aggravato nel caso di commissione del fatto con furto d'identità (c.d. digitale). Infine, è stato sostenuto[29] che sarebbe assimilabile ai casi di *deep-porn* aventi ad oggetto pornografia minorile il reato di cui all'art. 600-quater.1[30], c.d. di pornografia virtuale, il cui riferimento alle «tecniche di elaborazione grafica» appare indubbiamente riconducibile ai sistemi di intelligenza artificiale che oggigiorno permettono la creazione dei *deepfake*. Nonostante l'evidente sforzo dottrinale di ricomprendere nelle fattispecie di cui sopra anche quella più specifica del *deep-porn*, l'esigenza di una tutela *ad hoc* non sembrerebbe venir meno. Al contrario, si sottolinea l'inadeguatezza della legislazione vigente a far fronte ad una simile problematica. Di fatto, il Governo italiano ha presentato una proposta di legge, che, qualora approvata, vedrebbe modificato il Codice penale con l'aggiunta dell'art. 612-*quater*. La bozza della proposta[31], dalla rubrica «Illecita diffusione di contenuti generati o manipolati artificialmente», punirebbe la condotta di chi «cagioni ad altri un danno ingiusto, mediante invio, consegna, cessione, pubblicazione o comunque diffusione di immagini o video di persone o di cose ovvero di voci o suoni in tutto o in parte falsi, generati o manipolati mediante l'impiego di sistemi di intelligenza artificiale, atti a indurre in inganno sulla loro genuinità o provenienza» attraverso la reclusione da uno a cinque anni. Inoltre, con l'intento di stemperare la rigidità codicistica, la proposta comprende delle modifiche ulteriori al codice penale, tra le quali: l'apposizione della dicitura «ovvero mediante l'impiego di sistemi di intelligenza artificiale» ad una serie di reati "tradizionali", come il già menzionato 640-*ter*, e una nuova aggravante comune di cui all'art. 61 («*1*1-decies: l'avere commesso il fatto mediante l'impiego di sistemi di intelligenza artificiale, quando gli stessi, per la loro natura o per le modalità di utilizzo, abbiano costituito mezzo insidioso, ovvero quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa, ovvero aggravato le conseguenze del reato»). La tutela predisposta dall'ordinamento italiano, in ambito penalistico piuttosto esigua, offre

[28] Art. 640-*ter*, «Frode informatica»: «Chiunque, alterando in qualsiasi modo il funzionamento di un sistema informatico o telematico o intervenendo senza diritto con qualsiasi modalità su dati, informazioni o programmi contenuti in un sistema informatico o telematico o ad esso pertinenti, procura a sé o ad altri un ingiusto profitto con altrui danno, è punito con la reclusione da sei mesi a tre anni e con la multa da euro 51 a euro 1.032. La pena è della reclusione da uno a cinque anni e della multa da euro 309 a euro 1.549 se ricorre una delle circostanze previste dal numero 1) del secondo comma dell'articolo 640, ovvero se il fatto produce un trasferimento di denaro, di valore monetario o di valuta virtuale o è commesso con abuso della qualità di operatore del sistema. La pena è della reclusione da due a sei anni e della multa da euro 600 a euro 3.000 se il fatto è commesso con furto o indebito utilizzo dell'identità digitale in danno di uno o più soggetti. Il delitto è punibile a querela della persona offesa, salvo che ricorra taluna delle circostanze di cui al secondo e terzo comma o la circostanza prevista dall'articolo 61, primo comma, numero 5, limitatamente all'aver approfittato di circostanze di persona, anche in riferimento all'età».

[29] *Ivi*.

[30] Art. 600-*quater*.1: «Pornografia virtuale»: «Le disposizioni di cui agli articoli 600-ter e 600-quater si applicano anche quando il materiale pornografico rappresenta immagini virtuali realizzate utilizzando immagini di minori degli anni diciotto o parti di esse, ma la pena è diminuita di un terzo. Per immagini virtuali si intendono immagini realizzate con tecniche di elaborazione grafica non associate in tutto o in parte a situazioni reali, la cui qualità di rappresentazione fa apparire come vere situazioni non reali».

[31] Art. 1, proposta di legge ordinaria 30 marzo 2021, n. 2986, *Introduzione dell'articolo 612-quater del codice penale, in materia di manipolazione artificiale di immagini di persone reali allo scopo di ottenerne rappresentazioni nude.*

maggiori garanzie sul versante civilistico: un aspetto decisivo che la questione dei *deep-nude* pone in rilievo è quello della tutela dei dati personali e, più in generale, della tutela della propria immagine e dell'identità personale. In particolare, risulterebbero senz'altro violati, o quantomeno minacciati, da un uso improprio dei *deepfake* il diritto alla propria identità personale e i suoi corollari: il diritto ad essere sé stessi e ad avere una rappresentazione veritiera del proprio io nella vita di relazione, nonché a non veder alterato, modificato, offuscato il proprio patrimonio intellettuale, ideologico, politico, etico, religioso, professionale.
In quanto compresi nel novero dei cc.dd. diritti "inviolabili" della personalità[32], il diritto all'identità personale e i diritti che ne conseguono sono garantiti e tutelati non soltanto dal codice civile (all'art. 10) e dalla legge 633/1941 (agli art. 96 e 97), ma anche a livello costituzionale, con il combinato disposto degli art. 2 e 3. Ulteriormente, non dev'essere trascurata la protezione apprestata dalle Carte internazionali: la Convenzione Europea sui Diritti dell'Uomo (riconosciuta e applicata anche dall'UE, attraverso l'art. 6 TUE, con riguardo ai diritti inviolabili dell'individuo) all'art. 8 protegge la «vita privata e familiare», includendo anche tutte «le informazioni personali che un individuo può legittimamente aspettarsi non vengano pubblicate senza il suo consenso[33]», così come l'art. 12 della Dichiarazione dei diritti dell'Uomo[34]. Muovendo da queste considerazioni, è agevole comprendere le ragioni per cui l'ordinamento disciplini, all'art. 494 c.p.[35], il reato di «*furto d'identità*», intendendo per tale la sostituzione della propria persona a quella altrui, ovvero l'attribuzione a sé o ad altri di un nome, uno stato o una situazione giuridica falsi, inducendo taluno in errore con l'obiettivo di procurarsi un indebito vantaggio o arrecare danno ad altri. Certamente i contenuti *deepfake* implicano una violazione dell'identità personale: ma non solo. È doveroso fare riferimento ad un'ulteriore fattispecie: quella dell'identità c.d. digitale, definita come «la rappresentazione informatica di ciascun cittadino mediante i suoi dati identificativi[36]», che dunque consente l'individuazione del soggetto a mezzo di strumenti informatici. Si tratta di una sfumatura del diritto all'identità personale, tuttavia decisamente più problematica, se si tiene conto di due aspetti tra essi collegati: anzitutto, che nel mondo *virtuale* le informazioni acquisiscono una capacità espansiva vertiginosa se confrontata con il mondo *reale*; e che, di conseguenza, risulta ben più arduo il tentativo di procedere alla cancellazione dei dati presenti nella rete Internet. Non a caso, i profili virtuali dell'individuo tendono ad essere molto più rigidi nel mutamento rispetto alla dinamicità dell'identità personale.

7. Il silenzio dell'AI Act

In apertura della disamina della disciplina dei *deepfake* da parte dell'Unione Europea, è opportuno evidenziare la tutela piuttosto scarna, senz'altro insufficiente, predisposta dal Regolamento europeo

[32] Si vedano, fra tutte: Corte cost. 5 aprile 1973 n. 38 e Cassazione Civile, Sez. I, sentenza 3769 del 22 giugno 1985.
[33] Corte EDU, IV sez., Sent. 6 aprile 2010, *FLinkkila and Others contro Finlandia*.
[34] «Nessun individuo potrà essere sottoposto ad interferenze arbitrarie nella sua vita privata, nella sua famiglia, nella sua casa, nella sua corrispondenza, né a lesione del suo onore e della sua reputazione. Ogni individuo ha diritto ad essere tutelato dalla legge contro tali interferenze o lesioni».
[35] Art. 494 c.p.: «Chiunque, al fine di procurare a sé o ad altri un vantaggio o di arrecare ad altri un danno, induce taluno in errore, sostituendo la propria all'altrui persona, o attribuendo a sé o ad altri un falso nome, o un falso stato, ovvero una qualità a cui la legge attribuisce effetti giuridici, è punito, se il fatto non costituisce un altro delitto contro la fede pubblica, con la reclusione fino a un anno».
[36] *Identità digitale* in Treccani.it – Vocabolario Treccani on line, disponibile al link https://www.treccani.it/enciclopedia/identita-digitale_%28altro%29/ (ultima consultazione 29/09/2024).

approvato il 13 giugno di quest'anno in materia di intelligenza artificiale: l'AI Act[37]. Anzitutto, si tratta di un atto che non crea diritti in capo agli individui, bensì è destinato agli sviluppatori dei sistemi di IA e a coloro che li utilizzano a titolo professionale. Il Regolamento prevede una classificazione dei prodotti dell'Intelligenza Artificiale basata sul rischio significativo che essi possono costituire per la salute, la sicurezza o i diritti fondamentali dei cittadini dell'Unione, spaziando da sistemi che presentano un rischio minimo fino alle c.d. pratiche vietate. In quest'ambito, i sistemi generativi di *deepfake* sarebbero da individuare come sistemi a rischio limitato, tenuto conto del fatto che la sola disposizione del Regolamento in tema di *deepfake* prevede per gli sviluppatori non più di un mero obbligo di trasparenza: i «fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici» devono garantire che «gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente» e che «le loro soluzioni tecniche siano efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche[38]». Tale obbligo invece non si applica se i sistemi di IA svolgono una funzione di assistenza per l'*editing standard* o non modificano in modo sostanziale i dati di input forniti dal *deployer* o la rispettiva semantica, o se autorizzati dalla legge ad accertare, prevenire, indagare o perseguire reati. Inoltre, al fine di contribuire alla corretta attuazione degli obblighi relativi alla rilevazione e all'etichettatura dei contenuti generati o manipolati artificialmente, l'AI Act incoraggia e agevola l'elaborazione di «codici di buone pratiche» a livello dell'Unione[39]. Alla Commissione è poi conferito il potere di adottare atti di esecuzione per approvare tali codici di buone pratiche secondo la procedura di cui all'art. 56, p. 6, 7 e 8[40].

Ad oggi, provare a formulare una previsione sull'impatto dell'entrata in vigore della norma sembrerebbe ancora prematuro. Senz'altro l'obbligo di etichettare i prodotti generati da IA - e il connesso regime di sanzioni, perlopiù pecuniarie, che costituiscono un forte deterrente - porterà ad un aumento della consapevolezza circa i contenuti reali e quelli interpolati, ridimensionando l'incontrollata diffusione di

[37] Regolamento (UE) n. 2024/1689 del Parlamento e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'unione (regolamento sull'intelligenza artificiale).

[38] Art. 50, par. 2, Regolamento (UE) n. 2024/1689.

[39] Art. 50, par. 7, Regolamento (UE) n. 2024/1689.

[40] A. FERNANDEZ, *Regulating Deep Fakes in the Proposed AI Act*, 23 marzo 2022, disponibile al link: https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/ (ultima consultazione: 27/05/2024).
Art. 56: «6. L'ufficio per l'IA e il comitato monitorano e valutano periodicamente il conseguimento degli obiettivi dei codici di buone pratiche da parte dei partecipanti e il loro contributo alla corretta applicazione del presente regolamento. L'ufficio per l'IA e il comitato valutano se i codici di buone pratiche contemplano gli obblighi di cui agli articoli 53 e 55, nonché le questioni elencate al paragrafo 2, e monitorano e valutano periodicamente il conseguimento dei loro obiettivi; essi pubblicano la loro valutazione riguardante l'adeguatezza dei codici di buone pratiche. La Commissione può approvare, mediante atto di esecuzione, un codice di buone pratiche e conferire ad esso una validità generale all'interno dell'Unione. Tale atto di esecuzione è adottato secondo la procedura d'esame di cui all'articolo 98, paragrafo 2.
7. L'ufficio per l'IA può invitare tutti i fornitori di modelli di IA per finalità generali ad aderire ai codici di buone pratiche. Per i fornitori di modelli di IA per finalità generali che non presentano rischi sistemici, tale adesione può essere limitata agli obblighi di cui all'articolo 53, a meno che essi dichiarino esplicitamente il loro interesse ad aderire al codice nella sua interezza.
8. L'ufficio per l'IA incoraggia e agevola, se del caso, anche il riesame e l'adeguamento dei codici di buone pratiche, in particolare alla luce di norme emergenti. L'ufficio per l'IA fornisce assistenza per quanto riguarda la valutazione delle norme disponibili».

contenuti prodotti dall'IA privi dell'apposito *watermark* su canali che non presentino alcuna *policy* a riguardo, come avviene frequentemente sui social media *X* o *Telegram*. Tuttavia, non potrà essere esclusa la possibilità che continueranno ad esistere sistemi non in conformità all'AI Act, utilizzati al fine di generare contenuti illeciti, come avviene nel *dark web*, una parte dell'Internet non indicizzata dai motori di ricerca. Ed è agevole ritenere che, tra questi, sopravviveranno anche sistemi generativi di *deepfake* a sfondo pornografico. Inoltre, anche con la corretta applicazione della regola sulla trasparenza, quest'ultima non sarebbe idonea a mitigare le conseguenze che ne derivano per le vittime in termini di lesione dell'immagine, dell'identità personale, nonché della reputazione e del decoro, senza contare poi i danni psicologici[41]. L'ipotesi di prevedere delle tutele anche in quest'ambito è inserita nella direttiva del Parlamento Europeo e del Consiglio sulla lotta alla violenza contro le donne e alla violenza domestica[42], la quale prevede che un primo livello di protezione delle vittime di abusi potrebbe essere costituito da ulteriori contromisure nei confronti delle piattaforme che consentono la diffusione di tali contenuti, nell'ottica di un miglioramento nelle capacità di rilevamento dei media fonte degli abusi[43]. La Commissione Europea ha inviato una richiesta di informazioni a otto grandi compagnie del mondo tech (*Bing, Google Search, Instagram, Facebook, Snapchat, Tiktok etc.*) ai sensi della legge sui servizi digitali riguardo i rischi legati all'IA generativa, tra cui la diffusione di video *deepfake* legati alla disinformazione e alla violenza di genere. In secondo luogo, individua come auspicabile la formazione di uno standard europeo a cui gli Stati membri debbano adeguarsi, con una definizione di deep-fake comprensiva sia dell'aspetto tipologico (cioè della forma) che dell'aspetto soggettivo (dei destinatari a cui i *deepfake* si riferiscono), per garantire una parità di trattamento a tutte le donne vittime del fenomeno nell'Unione Europea. In aggiunta a ciò, gli Stati potrebbero mirare ad un incremento dell'efficienza della legislazione interna (penale e civile) a protezione del *deep-porn*. Infine, qualche voce in dottrina ha accennato al ruolo di enorme rilevanza che sarà svolto dalle nuove autorità nazionali di vigilanza del mercato, poste dall'AI Act a presidio dell'applicazione e del rispetto del Regolamento. Queste ultime potrebbero incrementare la tutela delle vittime dei deep-fake, nell'ottica di un più pregnante tutela dei diritti fondamentali, dal momento che i requisiti di «*indipendenza, imparzialità e assenza di pregiudizi*[44]» che devono guidare l'esercizio dei loro poteri rappresenterebbero la diretta implicazione della dimensione «costituzionale» della disciplina europea dell'IA[45]. Alla luce di quanto affermato, «la falsa promessa di *deepfake* trasparenti» non può essere una soluzione universale: l'obbligo formale di trasparenza è idoneo a esentare dai rischi solo una minima parte dell'universo dei *deepfake* che oggi dominano il web[46].

[41] M. ŁABUZ, *Regulating Deep Fakes in the Artificial Intelligence Act*, in *ACIG*, 1, 2023, disponibile al link: https://www.acigjournal.com/Regulating-Deep-Fakes-in-the-Artificial-Intelligence-Act,184302,0,2.html (ultima consultazione 30/09/2024).
[42] *Ibidem*.
[43] A. FERNANDEZ, *op. cit.*
[44] Art. 70, Regolamento (UE) n. 2024/1689.
[45] Cfr. GARANTE DELLA PROTEZIONE DEI DATI PERSONALI, *Segnalazione al Parlamento e al Governo sull'autorità per l'IA*, disponibile al link https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9996493 (ultima consultazione 27/05/2024).
[46] *Ibidem*.

8. Deepfake e tutela dei dati personali: il GDPR

I confini della questione della tutela dell'identità personale (e *digitale*) si confondono con il tema, di eguale rilevanza, della protezione dei dati personali. La disciplina in materia è disposta dal General Data Protection Regulation (679/2016) dell'UE e dal Codice Privacy che lo ha recepito all'interno dell'ordinamento italiano, nonché dalla Carta dei diritti fondamentali dell'Unione Europea (art. 8).
La pericolosità della creazione di *deep-porn* concerne primariamente la tipologia di dati personali che siano oggetto del trattamento illecito. Il GDPR classifica come «categorie particolari di dati personali» quelli che «rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona». Sono, quindi, compresi nel novero di tali dati anche i cc.dd. «dati biometrici» che, a norma dell'art. 4, sono «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», di cui si avvalgono i sistemi di intelligenza artificiale nella produzione di contenuti pornografici "profondamente falsi". Il GDPR pone un divieto di trattamento di tali dati, con delle limitazioni nel caso in cui sia presente una base legittima del trattamento: il secondo paragrafo dell'art. 9 annovera tra i casi di liceità del trattamento quello in cui vi sia il consenso esplicito ovvero i dati siano resi manifestamente pubblici da parte dell'interessato, o ancora nel caso di un interesse pubblico rilevante, fattispecie che non si riscontrano nel caso della pubblicazione dei *deep-nude,* ma che, al contrario, mancherebbero del tutto. Ricordiamo che il GDPR non si applica nel caso della c.d. *household exemption*, ossia al trattamento dei dati posto in essere da una persona fisica per l'esercizio di attività a carattere esclusivamente personale o domestico; al contrario, qualora tale trattamento - effettuato nell'ambito di applicazione del diritto dell'Unione - dovesse acquisire carattere pubblicistico, l'applicazione della disciplina del Regolamento europeo sarebbe inderogabile. In realtà, secondo il Considerando 18, «le attività a carattere personale o domestico potrebbero comprendere la corrispondenza e gli indirizzari, o l'uso dei *social network* e attività *online* intraprese nel quadro di tali attività»: tuttavia, l'attività interpretativa della Corte di Giustizia dell'Unione Europea ha ampliato tale fattispecie, giungendo a ritenere che la *household exemption* non si applichi «nel caso del trattamento di dati personali consistente nella loro pubblicazione su internet in modo da rendere tali dati accessibili ad un numero indefinito di persone[47]». È dunque da preferire la tesi secondo cui la diffusione su larga scala di dati personali altrui renderebbe inevitabile l'applicazione del Regolamento europeo n. 679/2016. Dal punto di vista rimediale, il GDPR offre una serie di strumenti volti a tutelare i diritti di coloro che abbiano visto sottoposti i propri dati ad un

[47] Corte di Giustizia dell'Unione Europea, sentenza del 6 novembre 2003, C-101/01 Göta hovrätt - Svezia. Il caso faceva riferimento alla normativa precedente in materia di protezione dei dati personali (direttiva 46/95/CE), così come un'altra interpretazione della Corte in senso analogo (Corte di Giustizia dell'Unione Europea, sentenza del 11 dicembre 2014, C-212/13 František Ryneš contro Úřad pro ochranu osobních údajů). Successivamente all'emanazione del GDPR i casi in materia portati all'attenzione della Corte sono stati rari. Tuttavia, la CGUE si è sempre espressa nel senso di interpretare restrittivamente qualsiasi eccezione al GDPR: si veda, fra tutti, la sentenza della Corte di Giustizia dell'Unione Europea, sentenza del 10 luglio 2018, C-25/17 Jehovan todistajat, in cui la Corte di Giustizia ha affermato che l'attività di predicazione porta a porta dei testimoni di Geova non rientra tra quelle esclusivamente personali o domestiche per le quali sarebbe applicabile l'eccezione, cfr. F. ROSSI DAL POZZO, *La tutela dei dati personali nella giurisprudenza della Corte di giustizia*, in *Eurojus*, 4, 2018, p. 8.

trattamento illecito. In primo luogo, quali estrinsecazioni del più generale diritto all'oblio, i diritti alla cancellazione e alla deindicizzazione: nel primo caso, l'interessato ha diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo, nel caso in cui non vi sia una base legittima del trattamento[48]; nel secondo, invece, sorge in capo all'interessato il diritto a non essere trovati *online*, cioè al c.d. *delisting*: viene recisa la connessione tra la ricerca di un contenuto *online* e il *link* che consente il rinvio al contenuto stesso[49]. Inoltre, il GDPR prevede un'accurata procedura in caso di violazione dei dati personali (c.d. *data breach*), quella «violazione di sicurezza che comporta accidentalmente o in modo illecito la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o comunque trattati[50]». Lo svolgimento della procedura non è dei più semplici: il titolare del trattamento (e cioè la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento), ove riscontri una violazione dei propri dati personali, deve notificarla all'autorità di controllo competente senza ingiustificato ritardo e, ove possibile, entro 72 ore dal momento in cui ne è venuto a conoscenza. Inoltre, la violazione dà luogo ad un diritto al risarcimento dei danni da parte del titolare o, solo eventualmente, del responsabile del trattamento («la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che tratta dati personali per conto del titolare del trattamento[51]»), nonché ad una sanzione penale e/o pecuniaria per il trasgressore. Infine, il caso dei *deep-porn* si inserisce nell'ambito di due degli illeciti penali previsti dal c.d. Codice privacy[52]: in primo luogo, l'art. 167 punisce la condotta di chi, «al fine di trarre per sé o per altri profitto ovvero di arrecare danno all'interessato, arreca nocumento all'interessato in violazione di specifiche disposizioni di legge», tra cui quelle che regolamentano i dati relativi al traffico telematico; l'art. 167-*bis*, invece, punisce «la comunicazione e la diffusione di dati personali oggetto di trattamento su larga scala, al fine di trarre profitto per sé o altri ovvero al fine di arrecare danno». Sarebbe dunque possibile ricomprendere nell'alveo della tutela delle norme sopracitate anche la condotta di chi diffonda *deep-nude* senza il consenso dell'interessato, attraverso un'interpretazione estensiva volta ad un allargamento della protezione predisposta dal Codice della Privacy.

9. Spunti comparatistici: normative a confronto

Dando uno sguardo oltreoceano, vediamo che il fenomeno *deepfake* è ugualmente dilagante, ma la sua diffusione non va di pari passo con un'unitarietà nella normazione di questo fenomeno. Negli Stati Uniti, per esempio, al momento manca una legislazione a livello federale per la regolamentazione del *deepfake*; tuttavia una tutela più o meno specifica viene offerta da alcuni suoi stati, tra cui il Texas. Quest'ultimo, in particolare, ha adottato due leggi riguardanti il fenomeno *deepfake*, una che disciplina l'uso politico dello stesso, e l'altra l'uso pornografico del medesimo. Venendo alla prima, con la legge SB n.751, il Texas è

[48] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati, GDPR), art. 17, par. 1.
[49] GDPR, art. 17, par. 2.
[50] GDPR, art. 4, punto 12.
[51] GDPR, art. 4, punto 8.
[52] D. Lgs. 30 giugno 2003, n. 196, Codice in materia di protezione dei dati personali.

andato a modificare l'art. 255.004 del codice elettorale, aggiungendo le sottosezioni (d) ed (e). Tale legge si applica soltanto quando il contenuto *deepfake* risulta essere un video, intenzionalmente creato e condiviso (condotta cumulativa), a 30 giorni dalle elezioni politiche, e al fine di danneggiare un candidato oppure per influenzare il risultato dell'elezione stessa[53]. Al contrario, la seconda legge, la SB 1361, ha modificato, a partire dall'1 settembre 2023, il capitolo 21, codice penale, aggiungendo la sezione 21.165. La SB 1361 punisce la persona che, intenzionalmente, produca o distribuisca (dunque non una condotta cumulativa come invece richiede la SB 751, bensì alternativa), con mezzi elettronici, un video falso che raffiguri la persona oggetto del video, senza il suo consenso, con parti intime esposte oppure impegnate in condotte sessuali[54]. Nella trattazione della normativa che dà il Texas in materia di *deepfake*, si può appurare come non si riscontri una disciplina unica del fenomeno, in quanto viene regolamentato soltanto nelle sue accezioni politiche e/o pornografiche, lasciando così aperta la possibilità ad eventuali lacune.

La scelta di questo Stato di regolamentare, seppur in modo parziale, il fenomeno non rimane però un caso isolato. Altri esempi di regolamentazione sono presenti in California, in cui, analogamente alle leggi del Texas, sono stati approvati due disegni di legge. Il primo rende illegale la pubblicazione di qualsiasi video manipolato che potrebbe, ad esempio, sostituire il volto o il discorso di un candidato al fine di screditarlo, entro 60 giorni da un'elezione[55]. Il secondo consente invece ai residenti dello stato di citare in giudizio chiunque inserisca la propria immagine in un video pornografico utilizzando la tecnologia *deepfake*[56].

Ancora, sempre negli Stati Uniti, anche lo stato della Florida ha adottato una normativa mirata per cercare di arginare il fenomeno *deepfake*. In particolare, è prevista per il 1 luglio 2024 l'entrata in vigore di una proposta di legge - SB 850 - che mira a richiedere alle campagne politiche di rivelare l'uso dell'intelligenza artificiale in qualsiasi immagine, video, audio, testo, e altri contenuti digitali, utilizzati negli annunci politici. In particolare, il testo della legge dispone che: «Se una pubblicità politica, una comunicazione elettorale o un'altra pubblicità varia, di natura politica, contiene immagini, video, audio, grafica o altri contenuti digitali creati in tutto o in parte con l'uso dell'intelligenza artificiale generativa, se il contenuto generato sembra raffigurare una persona reale che compie un'azione che in realtà non si è verificata e se il contenuto

[53] *Senate Bill* n. 751, sezione 1, punto d) ed e): «d) Una persona commette un reato se, intenzionalmente per danneggiare un candidato o influenzare il risultato di un'elezione: 1) crea un video *deepfake*; e 2) provochi la pubblicazione del video deepfake o la sua distribuzione entro 30 giorni dalle elezioni.
e) Nella presente Sezione, per "video *deepfake*" si intende un video, creato con l'intento di ingannare, che sembra raffigurare una vera e propria persona che compie un'azione che non si è verificata nella realtà».

[54] *Senate Bill* n. 361, sezione 1, punto b): «Una persona commette un reato se, senza il consenso effettivo della persona che sembra essere raffigurata, produce o distribuisce consapevolmente con mezzi elettronici un video profondamente falso che sembra raffigurare la persona con le parti intime esposte o impegnate in condotta sessuale è punito come reato c.d. di classe A, con pena detentiva fino a un anno e pena pecuniaria fino a 4.000 dollari, anche cumulative».

[55] *Assembly Bill* n. 730, sezione 493: «Questa legge vieta a una persona, a un comitato o a un'altra entità, entro 60 giorni da un'elezione in cui un candidato a una carica elettiva apparirà sulla scheda elettorale, di distribuire con effettiva malizia supporti audio o visivi materialmente ingannevoli del candidato con l'intento di danneggiare la reputazione del candidato o di ingannare un elettore a votare a favore o contro il candidato, a meno che i media non includano una dichiarazione che attesti che i media sono stati manipolati».

[56] *Assembly Bill* n. 602, sezione 1, punto 14): «Questo disegno di legge prevede che un individuo raffigurato, come definito, abbia una causa di azione contro una persona che (1) crea e divulga intenzionalmente materiale sessualmente esplicito, se la persona sa o avrebbe dovuto ragionevolmente sapere che l'individuo raffigurato non ha acconsentito alla sua creazione o divulgazione o (2) che rivela intenzionalmente materiale sessualmente esplicito che la persona non ha creato se la persona sa che l'individuo raffigurato non ha acconsentito alla sua creazione».

generato è stato creato con l'intento di pregiudicare un candidato o di ingannare in merito a una questione di voto, la pubblicità politica, la comunicazione elettorale o altra pubblicità varia deve indicare in modo ben visibile la seguente dichiarazione di non responsabilità: "Creato in tutto o in parte con l'uso dell'intelligenza artificiale generativa (AI)"»[57]. Si vedano anche qui analogie con le già presentate leggi del Texas e della California in tema di *deepfake* usato in prossimità di elezioni politiche. Tuttavia, mentre le leggi del Texas e della California si applicano solo a periodi che vanno dai 30 ai 60 giorni prima di un'elezione, al contrario, la legge della Florida non pone termini. Difatti, la maggior parte dei candidati in genere annuncia le proprie campagne mesi prima delle elezioni politiche, e non a ridosso delle stesse (30-60 giorni prima).
In aggiunta, i potenziali candidati, che non hanno dichiarato ufficialmente la loro corsa per l'ufficio, sarebbero ugualmente coperti dalla normativa della Florida, facendo, questa, riferimento più generalmente a contenuti che ritraggono «una persona reale» e non limitandosi, invece, a un «candidato», contrariamente a quanto accade nella normativa del Texas e della California. Guardando invece a cosa accade a livello federale, è stata presentata, come già anticipato, una proposta di legge: *No AI Fraud Act*. Questa proposta di legge nasce, come d'altronde tutte le normative in materia, da casi che vedono coinvolti soggetti di cui viene usata l'immagine allo scopo di creare contenuti falsi sfruttando l'intelligenza artificiale. Nel caso del *No AI Fraud Act*, l'ampia casistica viene direttamente presentata in apertura del testo della proposta di legge. All'interno di questa ricca elencazione, primo fra tutti spicca un caso, che si colloca nell'aprile del 2023, in cui l'intelligenza artificiale era stata usata per creare una canzone "*heart on my sleeve*" emulando le voci dei cantanti Drake e The Weeknd, ottenendo più di 11 milioni di visualizzazioni. Viene poi ricordato, nell'ottobre del 2023, una fattispecie che vedeva l'intelligenza artificiale usata per creare false, non consensuali, immagini intime di un gruppo di ragazze liceali, a WestField in New Jersey. Ancora, viene menzionato un rapporto del dipartimento di Sicurezza Interna (intitolato: «Aumento della minaccia delle identità *deepfake»*) in cui si afferma che, a ottobre 2020, si erano registrate più di 100mila false immagini di nudo, generate al computer, di donne, create senza il loro consenso o a loro insaputa[58].
Questa proposta di legge ha lo scopo di proteggere, a livello federale, le voci e le sembianze degli individui, che possono essere impropriamente usate e manipolate dall'intelligenza artificiale. In particolare, prevede l'esistenza di un vero e proprio diritto del singolo sulla sua voce e sembianze[59]. Ciò comporta che l'individuo ha il diritto di controllare l'uso che viene fatto delle proprie caratteristiche identitarie, e, conseguentemente, agire contro coloro che agevolano, creano e diffondono prodotti ingannevoli generati con l'intelligenza artificiale, senza il loro consenso. Il *No AI Fraud Act* si pone anche a difesa del primo emendamento, richiedendo che venga bilanciato l'interesse privato da un lato, e l'interesse pubblico, nell'uso non autorizzato, dall'altro. I fattori presi in considerazione per il bilanciamento includerebbero l'uso commerciale o meno del contenuto, la sua rilevanza sullo scopo primario, e il suo impatto sugli interessi economici del titolare dei diritti. Infine, il *No AI Fraud Act* avrebbe un impatto sulle imprese, richiedendo a queste ultime, specie quelle che appartengono al settore creativo, che l'uso che le stesse fanno dell'intelligenza artificiale sia rispettoso del diritto degli individui alla loro voce e sembianze. Ciò potrebbe comportare per le imprese di implementare nuove procedure di garanzia di questi diritti, con

[57] Senate Bill n. 850, sezione 1, punto 2).
[58] No AI Fraud Act, sezione 2.
[59] No AI Fraud Act, sezione 3, punto b).

eventuali costi aggiuntivi[60] a carico delle stesse[61]. Se da un lato c'è chi critica questa proposta di legge perché ritagliata su personaggi notori, e quindi non si porrebbe come strumento di tutela comune[62], dall'altro c'è chi si oppone alla sua approvazione perché troppo onnicomprensiva, in quanto il diritto che andrebbe a creare si applicherebbe a una quantità incredibilmente ampia di contenuti digitali[63].
Ciò rende evidente come la normazione di un simile fenomeno, cioè quello dell'intelligenza artificiale, nel caso di specie del *deepfake*, non sia di facile approccio. Tornando in Europa, un esempio di normazione stringente del fenomeno *deepfake* proviene dal Regno Unito, in cui, a partire dal 26 ottobre 2023, è entrato in vigore l'*Online Safety Act*. Si tratta di una legge che intende proteggere bambini e adulti online, ponendo una serie di doveri alle società di social media e ai servizi di ricerca, responsabilizzandoli circa la sicurezza dei loro utenti sulle loro piattaforme. L'*Online Safety Act* si applica a servizi di ricerca e ai servizi che consentono agli utenti di pubblicare contenuti *online* o di interagire tra di loro, anche quando le società che forniscono tali servizi si trovano al di fuori del Regno Unito, nel caso in cui le stesse intrattengano rapporti con il Paese. Ad esempio, l'*Online Safety Act* si considera applicabile qualora il servizio abbia un significativo numero di utenti dal Regno Unito, oppure se quest'ultimo è un mercato di riferimento o, ancora, se il servizio è accessibile agli utenti del Paese e vi è un rischio materiale di danno significativo a tali utenti[64].
I nuovi reati[65] introdotti da questa legge, ed entrati in vigore dal 31 gennaio 2024, sono molteplici e non si limitano alla normazione del solo fenomeno *deepfake*, ma, più generalmente, riguardano: istigazione o aiuto all'autolesionismo[66], *cyberflashing*[67] (il fenomeno per il quale immagini oscene vengono inviate a sconosciuti online, molto spesso attraverso *Bluetooth* o *AirDrop*), invio di informazioni false intenzionate a causare un danno non futile o superficiale, comunicazioni minacciose, abuso di immagini intime[68].
Simili reati si applicano non già solo a chi produce il contenuto, ma anche alle persone che si limitano a condividere lo stesso. Sono infatti già state emesse condanne per i reati di *cyberflashing* e per la realizzazione di comunicazioni minacciose[69]. Per fronteggiare l'ampia casistica di reati poc'anzi menzionati,

[60] L. SARNOFF, *Taylor Swift and No Ai Fraud Act: How Congress Plans to Fight Back Against Ai DeepFakes*, gennaio 2024, disponibile al link: https://www.congress.gov/118/meeting/house/116778/documents/HHRG-118-JU03-20240202-SD002.pdf (ultima consultazione 17/06/2024).
[61] In resemble.ai, disponibile al link: https://www.resemble.ai/no-ai-fraud-act/#:~:text=The%20No%20Artificial%20Intelligence%20Fake%20Replicas%20and%20Unauthorized,likenesses%20against%20misuse%20by%20artificial%20intelligence%20%28AI%29%20technologies. (ultima consultazione 21/05/2024).
[62] C. CRESCENZI, *Gli Stati Uniti Vogliono Dire Basta ai Deepfake delle Celebrità*, gennaio 2024, disponibile al link: https://www.wired.it/article/stati-uniti-stop-deepfake-celebrita/ (ultima consultazione 28/09/2024).
[63] C. MCSHERRY, *The No AI Fraud Act Creates Far More Problems Than It Solves,* gennaio 2024, disponibile al link: https://www.eff.org/it/deeplinks/2024/01/no-ai-fraud-act-creates-way-more-problems-it-solves (ultima consultazione 28/09/2024).
[64] DIPARTIMENTO SCIENZA, INNOVAZIONE E TECNOLOGIA (UK), *Online Safety Act: explainer*, maggio 2024, disponibile al link: https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer (ultima consultazione 21/05/2024).
[65] Online Safety Act*,* 2023, parte 10, sezioni 179-191.
[66] Online Safety Act, 2023, parte 10, sezione 184.
[67] Online Safety Act*,* 2023, parte 10, sezione 187.
[68] Online Safety Act*,* 2023, parte 10, sezione 188.
[69] DIPARTIMENTO SCIENZA, INNOVAZIONE E TECNOLOGIA (UK), *Online Safety Act: Explainer*, maggio 2024, disponibile al link: https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer (ultima consultazione 28/09/2024). Il riferimento è a Nicholas Hawkes, uomo di 39 anni, condannato il 19 marzo 2024 dalla Corte Southend Crown per aver commesso il reato di cyberflashing, avendo inviato immagini, non richieste, del suo pene eretto a una ragazzina di 15 anni. L'uomo è stato condannato a 66 settimane di carcere e a un ordine restrittivo

la legge richiede a tutte le aziende di intraprendere azioni rigorose contro i contenuti e le attività illegali, attraverso l'attuazione di misure per ridurre i rischi che i loro servizi vengano utilizzati per commettere atti illeciti[70]. Dovranno inoltre adoperarsi per adottare sistemi che rimuovano i contenuti illegali quando essi appaiano sulla loro piattaforma[71]. Dovranno altresì fare in modo di ridurre i rischi che gli utenti possano entrare in contatto con contenuti illegali attraverso i loro servizi[72]. Infine, dovranno rimuovere qualsiasi contenuto illegale in presenza di una vittima, effettiva o presunta, quando ciò venga segnalato dagli utenti o ne vengano a conoscenza con altro mezzo[73]. I tipi di contenuti e attività illegali, da cui le piattaforme devono proteggere gli utenti, sono stabiliti dall'*Online Safety Act*, e riguardano: abusi sessuali su minori, violenza sessuale estrema, pornografia estrema, frode, comportamento controllante o coercitivo, contenuti razziali, reati contro l'ordine pubblico aggravati dalla religione, incitamento alla violenza, immigrazione clandestina e traffico di essere umani, promozione o istigazione al suicidio, abuso di immagini intime (il sopracitato *revenge porn*), vendita illegale di droghe o armi, sfruttamento sessuale, terrorismo[74].

10. Conclusioni

Com'è stato affermato a più riprese, il punto di forza dei *deep-nude* risiede nella rapidità della loro diffusione e la speculare difficoltà delle autorità nell'inibire e rimuovere i contenuti illeciti: una velocità di propagazione tale da rendere gli strumenti di *enforcement* del tutto vani rispetto alle migliaia di violazioni commesse. Così il diritto mostra tutta la sua inadeguatezza nei confronti dello sviluppo esponenziale della tecnologia, che mal si coniuga con l'artificiosità tipica dell'attività legislativa e giurisprudenziale. Come sostenuto dall'attrice Scarlett Johansson, la quale è stata ripetutamente vittima dei *deep-nudes*, il fenomeno dei *porn deepfake* costituisce una minaccia così grave e concreta che nemmeno la legge potrebbe arginarla: «l'Internet è un abisso che rimane virtualmente privo di regolamentazione», nonostante i numerosi tentativi di normare il fenomeno da parte di vari Paesi[75]. La potenziale dannosità della questione appare accentuata se si tiene in considerazione che le ormai evolutissime tecnologie di IA permettono di ottenere risultati così veritieri da non poter più riuscire a distinguere un'immagine o un video *reale* da una sua riproduzione *virtuale*. A dimostrazione di ciò i risultati del sondaggio, che rivelano come - alla sottoposizione del campione di 200 persone di quattro immagini (due delle quali raffiguranti persone reali e due raffiguranti individui creati da IA) - solo il 17,5% delle persone è in grado di riconoscere un'immagine realizzata dall'IA, contro l'82,5% delle stesse che invece si trova in difficoltà.

---

nei confronti della vittima. Per ulteriori approfondimenti: https://www.essex.police.uk/news/essex/news/news/2024/march/cyber-flashing-conviction/ (ultima consultazione 28/09/2024).

[70] Online Safety Act, 2023, parte 3, sezione 10.

[71] *Ivi.*

[72] *Ivi.*

[73] *Ivi.*

[74] DIPARTIMENTO SCIENZA, INNOVAZIONE E TECNOLOGIA (UK), *op. cit.*

[75] Cfr. D. HARWELL, *Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*, in *The Washington Post*, 2018, disponibile al link: https://www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/ (ultima consultazione 30/09/2024).