Digeat N.6 - 19 Giugno 2025

Deepfake pornografici: tutti più vulnerabili. Riflessioni su tutele e conseguenze sociali

Di Enrico Pelino

Abstract

L’alterazione della realtà attraverso l’intelligenza artificiale è un fenomeno apparentemente irreversibile della nostra società. L’articolo si sofferma sul fenomeno dei deepfake pornografici. È un’ampia casistica che va dal semplice innesto della fisionomia di una persona su corpi nudi altrui, alla creazione di video sofisticati che ritraggono scene sessuali mai avvenute. La finalità non è necessariamente di revenge porn o estorsiva, ma può anche rispondere al semplice capriccio gratuito di nuocere ad altri. Le fonti di approvvigionamento su cui costruire i deepfake sono abbondanti, si pensi soltanto all’amplissima disponibilità di immagini fisse e in movimento che ritraggono eventi sociali, cene tra amici, compleanni, vacanze, ecc. pubblicate su piattaforme social e di condivisione video, che forniscono materiale di qualità aggredibile da terzi. L’articolo, interrogandosi sull’adeguatezza della normativa disponibile, affronta brevemente alcune conseguenze sociali e giuridiche di lungo termine, non sempre analizzate, del fenomeno dei deepfake, e di quelli pornografici in particolare, ossia: l’esproprio del passato altrui; l’estensione del concetto di vulnerabilità; la perdita di quel collante necessario del tessuto sociale costituito dall’affidamento; l’erronea percezione di minore disvalore della pornografia non consensuale virtuale.

Indice

La Corea del Sud, attualmente uno dei paesi più colpiti dai *deepfake* pornografici, costituisce un interessante caso di studio, in quanto anticipa scenari che non vi è ragione di ritenere esclusi in Italia e offre spunti di reazione legislativa di cui è utile seguire gli effetti.

Il fenomeno è alimentato da una serie di occasioni materiali favorevoli: la normalizzazione sociale della generazione di contenuti artificiali e, anzi, la loro sperimentazione diffusa (chi non si diletta a creare immagini e video di IA?); l’accessibilità amplissima a piattaforme software dedicate; la ricchezza di foto e video altrui reperibili in rete. Non è troppo difficile ipotizzare le conseguenze quando tutto ciò incrocia disagi personali, rancori o dinamiche tossiche.

Il peccato originale

L'abbondanza di immagini private acquisibili dovrebbe far riflettere. Nella mia percezione, **il peccato originale dell'Intelligenza Artificiale consiste nell'assoluto arbitrio con cui si ritiene possibile appropriarsi di informazioni altrui per alimentare i modelli, realizzando ampie operazioni di *scraping***. Se rendiamo possibile nutrire i modelli di IA con il patrimonio di dati delle piattaforme *social* – e mi riferisco qui all'operazione che ha condotto in questi giorni Meta – perché dovrebbe essere vietato fare altrettanto ai singoli con le loro applicazioni? O perché essi dovrebbero percepirne l'antigiuridicità?

Unicamente al fine di consentire alle grandi multinazionali del settore l'approvvigionamento di dati che è essenziale ai loro servizi, abbiamo tracciato distinzioni capziose in tema di diritto d'autore, il settore finora più tutelato, quello in cui, fino a quando gli investimenti economici prevalenti venivano dalle multinazionali del settore e dalle società di gestione collettiva, si assisteva semmai a fenomeni del tutto opposti di intransigenza.

Anche sul piano della protezione dei dati personali, **si constata un generale abbandono di capisaldi giuridici, come il principio di limitazione della finalità**, riconosciuto dall'art. 5.1.b) GDPR. Un tempo, non si pensava neppure di mettere in dubbio che, se qualcuno pubblica delle immagini su un *social*, la finalità del trattamento è esclusivamente di condividerle con il proprio cerchio relazionale, non certo con il *business* del fornitore della piattaforma, indipendentemente dai termini capziosi delle condizioni generali di servizio.

Oggi si chiede a tutti di fare *opt-out*, senza che ciò determini troppo clamore istituzionale: non blocco del trattamento, non altre misure. Un tempo ci avremmo riso sopra, ve la ricordate quella catena di Sant'Antonio che cominciava pressappoco così: "*Con la presente dichiaro che Facebook non ha il mio permesso di utilizzare le mie foto personali, le informazioni, i messaggi o le pubblicazioni, passate o future*" e continuava in modo ampiamente sgrammaticato? Ecco, sgrammaticature giuridiche a parte, siamo arrivati proprio a questo. La bufala non sa più di bufala. **La bufala oggi siamo noi**.

La Commissione europea, assediata dalle lobby (o viceversa), pensa già di accomodare l'AI Act, ancora fresco di pubblicazione sulla Gazzetta ufficiale europea[1], perché non scontenti eccessivamente gli affari dei nuovi signori del mercato generativo. Il diritto è diventato un appesantimento o un orpello decorativo.

Quattro brevi riflessioni

Ma torniamo al nostro tema, la diffusione di *deepfake* pornografici, che ha il vantaggio di portare all'attenzione un piccolo, circoscritto, ma significativo campione di temi d'attualità. Vorrei, in particolare, toccare brevemente quattro corollari di una constatazione iniziale che, ritengo, tutti ormai condividiamo: **la realtà perde realtà, l'artificiale appare verosimile quanto il vero, ciò che è autentico non è più autentico**. Se ne possono trarre almeno quattro interessanti riflessioni:

1. l'esproprio del passato altrui;
2. l'estensione del concetto di vulnerabilità;
3. la perdita dell'affidamento e la conseguente disgregazione dei rapporti sociali;
4. la percezione di minore lesività del pornografico virtuale rispetto a contenuti reali.

L'esproprio del passato altrui

I *deepfake*, com'è intuibile, si inseriscono in maniera dirompente in dinamiche di *revenge porn*, poiché forniscono l'apporto di immagini mai registrate o addirittura di vicende mai avvenute e tuttavia assolutamente credibili, permettendo inoltre forme estreme di rappresentazione.

L'effetto è quello di massimizzare nella vittima tanto l'umiliazione quanto il terrore, due dei tratti salienti del ricatto pornografico. L'umiliazione è evidente, concentriamoci sul terrore. Il terrore deriva dalla perdita completa di controllo, non solo sulla circolazione delle proprie immagini condivise nell'intimità, com'è tipico della fattispecie criminosa, ma perfino sul proprio passato reale, che potrebbe essere trasformato in qualcosa che non è mai avvenuto. Viene cioè – e questo è del tutto nuovo – creato e ricreato, con infinite varianti continuamente sperimentabili, **un pregresso mai esistito**, "*un turbio pasado irreal que de algún modo es cierto*", un torbido passato irreale che in qualche modo è certo, come nel verso di Borges[2].

Qui emerge, a mio parere, uno dei tratti su cui dovremmo maggiormente concentrare la nostra attenzione nel catturare il fenomeno e le sue conseguenze sociali e umane. Non solo si usa violenza sulle decisioni attuali della vittima del ricatto ma, addirittura, **si esercita violenza su quelle passate, le si espropria il passato**. Se esisteva una cosa certa, un dominio sicuro, era il nostro storico, ciò che abbiamo o non abbiamo fatto, i conti ormai chiusi. Oggi non più.

L'estensione del concetto di vulnerabilità

Il *deepfake* pornografico va tuttavia ben oltre l'ambito del *revenge*. "*All we have to have is just a human form to be a victim*", **basta avere forma umana per essere vittime**. È la sintesi efficacissima di una collega, l'avvocato Carrie Goldberg[3].

Beninteso, quando si parla di "soggetti vulnerabili", restano valide le categorie tradizionali di riferimento, ad esempio "minori", "anziani", persone vittime di fragilità per condizioni psicofisiche o economiche o relazionali, dunque vulnerabili anche soltanto rispetto a certi rapporti giuridici, come nel caso della relazione dipendente – datore di lavoro. Tuttavia, è altrettanto vero che si assiste ora a un'estensione senza precedenti del concetto di soggetto vulnerabile.

Siamo diventati **tutti vulnerabili** rispetto alle applicazioni dell'intelligenza artificiale. In fondo, il materiale fotografico e le coordinate intime che chiunque pubblica massivamente su piattaforme, sul proprio sito, su canali di condivisione, anche professionali, è spesso più che sufficiente a fornire alimento per un *deepfake*.

La perdita dell'affidamento

La perdita di verità delle immagini, perfino dei video, porta con sé una conseguenza: **la perdita di fiducia**. Se la realtà che ci viene presentata da altri non è più credibile, viene meno, in generale, l'affidamento negli altri. La ragione per la quale soggetti anziani sono più facilmente vittima di truffe del modello *social engineering* non dipende soltanto, come si ipotizzerebbe di primo acchito, da inesperienza tecnologica, ma soprattutto da un'impostazione culturale: **gli anziani provengono da una società ampiamente costruita sull'affidamento**. Danno fiducia a chi dichiara di essere ciò che dichiara, perché si sono formati entro un modello di rapporti umani nel quale le affermazioni sono impegnative, la parola spesa ha valore, le luci sono nette e le ombre sono nette.

Il punto è che, ragionando più in generale, la fiducia è il collante che tiene insieme qualsiasi struttura sociale, qualsiasi relazione, anche la più intima. **Ogni collaborazione, ogni rapporto implica una cessione di fiducia ad altri**. L'effetto disgregante sulla coesione sociale dei *deepfake* diventa allora enorme, perché dissolve questo collante. Che cosa potrà derivarne sul lungo termine?

La percezione della minore lesività di ciò che è fittizio

Toniamo in Corea del Sud. "*Why is it a serious crime when it's not even your real body?*", perché un reato grave se non si tratta nemmeno del tuo vero corpo?, dice Kim, una delle intervistate in una recente indagine giornalistica sui *deepfake* pornografici[4].

Applicare un volto reale a un corpo altrui è davvero meno grave? Il tema merita riflessione perché tocca un aspetto giuridico decisivo, quello del disvalore sociale della condotta, ed è più complesso e sfumato di quanto a un primissimo esame non appaia.

A ben vedere, il ragionamento secondo cui la generazione di immagini artificiali che rappresentano eventi mai avvenuti sarebbe meno grave della diffusione di immagini reali poggia su una colpevolizzazione della vittima, dunque su un capovolgimento delle responsabilità, quasi che la vittima, non avendo partecipato all'evento che le si attribuisce, sia *ipso facto* assolta da un supposto peccato originario. Però – non è neppure il caso di notarlo – bisogna porsi nella prospettiva esattamente rovesciata.

Il punto centrale, comunque, **è che è vero ciò che appare vero**. Tanto più che l'intento del *deepfake* è di rendere il più realistica possibile l'immagine o il video. Basta del resto insinuare il dubbio che l'immagine possa essere vera: il dubbio persisterà, magari assottigliato, latente, ma persisterà anche dopo evidenti smentite. E tanto basta a creare un pregiudizio sociale permanente.

Inoltre, ed è questa a mio parere la ragione più persuasiva, anche ove l'immagine sia artificiale, **veri sono certamente l'umiliazione della vittima, il trauma che riceve**. Potrebbe perfino trattarsi di un falso grossolano, non per questo sarebbe meno offensivo. Emerge l'intenzione di violare, da distanza, la sfera più segreta di una persona, di "agirla" nella sua intimità, "*e il modo ancor m'offende*". **Ciò a cui assistiamo è più che un furto d'identità, è un furto di intimità, di dignità.**

Quali gli strumenti giuridici?

Esaurita questa breve analisi, è necessario chiedersi quali siano gli strumenti giuridici azionabili e con quali limiti.

Le aspettative maggiori cadrebbero sull'AI Act (regolamento UE 2024/1689), ma vanno deluse. Il Regolamento, nonostante fornisca una definizione normativa di *deepfake* all'art. 3.60), ne assoggetta la disciplina solo a un piuttosto blando obbligo di trasparenza ai sensi dell'art. 50 e non lo include direttamente, neppure quando assume forme pregiudizievoli come quello pornografico, tra le pratiche vietate dall'art. 5, nonostante sia stato autorevolmente opinato il contrario[5].

I due corpi normativi che, pur senza fornire definizioni specifiche e senza, per vero, neppure mai citare l'intelligenza artificiale, offrono tuttavia le garanzie più concrete in ambito civile e amministrativo, restano il GDPR (Regolamento UE 2016/679) e il DSA (Digital Services Act, Regolamento UE 2022/2065).

Quanto al primo, i *deepfake* che riproducono la fisionomia della vittima[6] costituiscono indubbiamente un trattamento di dati personali, la verità dell'informazione non essendo un requisito normativo rilevante nella nozione di "dato personale" e accrescendo semmai il rischio per l'interessato. L'eccezione domestica, normata all'art. 2.1.c) GDPR, scrimina tuttavia la mera attività generativa priva di immissione in circolazione.

Venendo al DSA, esso fornisce strumenti sia preventivi sia successivi, anche a vocazione cautelare, diretti agli snodi che rendono possibile la diffusione dei *deepfake* pregiudizievoli, ossia innanzitutto le piattaforme di *hosting*. È dunque intercettato anche qui non il momento creativo, ma quello della circolazione.

Sul versante penalistico, quello certamente più idoneo a una tutela, il *deepfake* lesivo non trova una disciplina specifica, se non in ambito di contrasto alla pedopornografia, attraverso la fattispecie dell'art. 600-*quater 1* c.p. e la previsione dell'ultimo comma dell'art. 600-*ter*.

Tuttavia, al momento in cui si scrive il presente articolo, sul piano generale, il *deepfake* pregiudizievole, soprattutto quello pornografico, non riceve una tutela penale dedicata, cosa che impone il ricorso a fattispecie ampie, quali, a seconda dei casi, i reati di diffamazione, molestia, estorsione. Solo in casi limitati, appare richiamabile la fattispecie della frode informatica, art. 640-*ter* c.p.. L'art. 494 c.p., sostituzione di persona, supplisce solo qualora i *deepfake* consistano nell'impersonificazione della vittima o nella falsa attribuzione di un suo stato, si pensi alle truffe del tipo "*CEO fraud*", in cui l'agente sostituisce sé stesso a un soggetto apicale per ottenere da parte di sottoposti o di *clientes* l'esecuzione di operazioni fraudolente, ma assai più difficilmente nel caso dei *deepfake* pornografici.

Ove l'immagine artificiale sia usata a scopo di *revenge porn*, la condotta risulta difficilmente sussumibile entro la formulazione dell'art. 612-*ter* c.p., che appare costruita sull'ipotesi criminosa esattamente opposta, quella di realizzazione o sottrazione non consensuale di materiale autentico, ancorché in proposito sia opportuno attendere consolidamenti giurisprudenziali. Si noti incidentalmente che lo stesso difetto costruttivo riguarda l'omologa ipotesi civilistica regolata all'art. 144-*bis* Codice Privacy (d.lgs. 196/03), pur con le mitigazioni in tal caso permesse dal ricorso all'analogia.

In chiave comparatistica, una tutela specifica in materia di *deepfake* pornografici, che ne intercetta anche il semplice possesso o la visione, è stata invece di recente introdotta in Corea del Sud, come reazione all'affermarsi di tali tipologie di condotta. Va segnalato che il DDL 1146 in materia di Intelligenza Artificiale prevede, all'art. 25, la formulazione di un art. 612-*quater* c.p. che punisce i *fake* anche oltre l'ambito pornografico, ma limitatamente alla sola circolazione (non quindi rispetto alla generazione o al possesso) e con il limite della loro ingannevolezza, oltre a una serie di integrazioni di altre disposizioni penali, nell'evidente direzione di colmare alquanto, ma non integralmente, l'attuale vuoto normativo.

Conclusioni

La percezione della realtà attraverso immagini e suoni, sulla quale abbiamo finora costruito le nostre valutazioni e le nostre decisioni, appare pregiudicata, in maniera probabilmente non reversibile, dalla diffusione di strumenti di Intelligenza Artificiale di facile accesso, a basso costo, immediata operatività, e in grado di produrre falsi assolutamente credibili. **Le conseguenze sociali e giuridiche di questo cambiamento epocale appaiono tuttora da trarre nella loro complessità**. La breve riflessione permessa dallo spazio di questo contributo ha cercato di individuarne alcune, che – almeno nella percezione di chi scrive – appaiono ancora poco esplorate nella riflessione giuridica, quantomeno rispetto ai loro effetti tossici di lungo termine, in particolare penso a temi come l'espropriazione del passato e la distruzione dell'affidamento reciproco.

Gli strumenti offerti dal diritto appaiono ampiamente perfettibili, sono tuttavia in fase di integrazione, quantomeno in ambito penale, ove ve n'è bisogno. Nel settore civile e amministrativo, la tutela più forte viene, ancora una volta, dal glorioso GDPR e dal più recente DSA, mentre deve segnalarsi la delusione che accompagna gli episodi più attuali della normazione europea, come l'AI Act, il cui formalismo ridondante non trova giustificazioni in sostanziali avanzamenti di tutela,

e l’abbandono “volontario” della proposta di direttiva sulla responsabilità dei sistemi di AI, segno entrambi di un’arrendevolezza unionale proprio nel momento in cui sarebbe necessario l’opposto.

Certo, tuttavia, è illusorio pensare che gli strumenti del diritto possano risolvere, da soli, una delle sfide del millennio, collegata a un avanzamento tecnologico procedente a un ritmo insostenibile, se non sono accompagnati da **un mutamento profondo di consapevolezza nella società**, oggi tuttavia sempre più fidelizzata da soluzioni IA al momento semigratuite, e dallo sviluppo conseguente di forme di dipendenza.

NOTE

[1] Duncan Roberts, EU to reform AI act in bid to boost competitiveness, Luxembourg Times, 9.4.2025.

[2] Jorge Luis Borges, *El tango*.

[3] Clare Duffy, *AI means anyone can be a victim of deepfake porn*, 12.11.2024.

[4] Yoonjung Seo and Mike Valerio, Deepfake porn is destroying real lives in South Korea, 25 aprile 2025.

[5] Commissione europea, *ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, 4.2.2025, C(2025) 884 final, § 145.

[6] I *deepfake* creati invece a partire da dati personali, ma anonimizzati in modo tale da non rendere riconoscibile una persona fisica individuabile si collocano (diversamente dal loro processo di generazione) fuori dell’area applicativa del GDPR.