T4F Series
PAPER N. 1
a.a. 2022/2023

Trento BioLaw Selected Student Papers

Il fenomeno del deepfake nell'attuale ordinamento giuridico

SANTE PUGLISI, CRISTIANO VERGARI

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. 2022-2023, 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.

Il fenomeno del *deepfake* nell'attuale ordinamento giuridico

Sante Puglisi, Cristiano Vergari[*]

ABSTRACT: Taking as a reference the phenomena of creation, with the aid of deepfake technology, of sexually explicit materials on the one hand, and fake news on the other, this paper aims to provide an overview of the regulatory tools that the internal and European Community legal systems offer to combat these abusive uses of technology. In particular, the recent regulatory work carried out within the European Union will be illustrated, highlighting its strengths, weaknesses and limitations. Finally, the behavior of the main players in the social network market will be briefly observed in relation to the problems raised by deepfake technology.

KEYWORDS: Deepfake; fake news; sexually explicit materials illegal diffusion; AI Act; Digital Services Act

1. Introduzione

L'avanzamento della c.d. intelligenza artificiale (IA) e dei suoi prodotti sempre più innovativi costituisce il fronte di una nuova rivoluzione silente, che prelude ad una forte invasione della vita delle generazioni presenti e future.

Si tratta della combinazione di due fattori o paradigmi, costituiti dall'automazione dei processi e dallo sviluppo di *software* improntati al *problem solving,* ossia capaci di generare soluzioni ragionate rispetto ai quesiti posti.

L'intelligenza artificiale è attualmente ancora una sfida, una novità in costruzione, che ruota attorno all'obiettivo minimo di eguagliare i processi del pensiero umano, per replicarne le logiche, i criteri e i segni distintivi e procedere al suo progressivo affiancamento.

La sfida successiva, tuttavia, sarà immancabilmente quell di riuscire a superare i confini del pensiero umano, cercando di ottenere dalle macchine la produzione di un pensiero più performante ed efficace. È quella che

[*] *Studenti dell'Università degli Studi di Trento, Facoltà di Giurisprudenza. E-mail:* sante.puglisi@studenti.unitn.it; cristiano.vergari@studenti.unitn.it.

si definisce l'intelligenza artificiale "forte", ovvero quello sviluppo dell'automazione connotato da fattori di imprevedibilità, generati dall'evoluzione autonoma dei sistemi, potenzialmente anche al di fuori della capacità di controllo umano[1].

Ciò richiama futuri strumenti tecnologici capaci di garantire prestazioni di apprendimento precluse all'uomo, in quanto basate sulla lavorazione istantanea di quantità enormi di dati, e idonei a produrre combinazioni veloci di dati e soluzioni applicative impensabili per la mente umana.

La replica ed il superamento delle capacità cognitive dell'uomo assomigliano, peraltro, a quanto di più sconvolgente si possa immaginare, poiché l'uomo si prepara a generare qualcosa di altro da sé, provvisto di potenzialità più elevate e finanche incontrollabili.

È chiaro, una volta di più, come tale prospettiva evolutiva produca problemi enormi, di natura non solo tecnica, ma soprattutto etica e di rispetto della dignità umana, che sono totalmente inediti e a cui l'ordinamento giuridico non è ancora preparato. Ad oggi, tutta l'attenzione mediatica relativa ai prodotti di intelligenza artificiale sembra riguardare l'aspetto tecnologico e la meraviglia sui risultati conseguibili attraverso le nuove tecniche di lavorazione dei dati.

Si guardi, per fare un esempio, ai filmati recentemente costruiti a simulazione dell'arresto dell'ex presidente degli Stati Uniti Donald Trump[2] o della resa del presidente ucraino Volodymyr Zelensky[3].

Molto debole appare, invece, il fronte del pensiero critico sul necessario bilanciamento tra le potenzialità di utilizzo delle nuove tecnologie ed i limiti da introdurre a tutela di valori e diritti riconosciuti sino ad oggi al genere umano, a cominciare dal valore della sua dignità.

L'obiettivo di questo lavoro si colloca esattamente in questa seconda prospettiva ed ambisce a mettere in luce, oltre al fenomeno del *deepfake*, le azioni sin qui condotte a livello europeo per iniziare a costruire una cornice giuridica allo sviluppo della c.d. intelligenza artificiale.

Per altro verso, ciò che appare più sfidante è la comprensione di quanto l'attuale ordinamento giuridico italiano sia resiliente alle nuove derive tecnologiche e commerciali.

---

[1] Per una trattazione maggiormente esaustiva sulla distinzione tra IA forte e IA debole, con particolare riferimento alla tematica del potenziale esercizio di funzioni giurisdizionali da parte dell'intelligenza artificiale, cfr. C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in DPCE online, 3, 2020, 3369 ss.

[2] REDAZIONE ANSA, *'Trump arrestato', ma sono foto fake di una intelligenza artificiale,* in *ansa.it*, 24 marzo 2023, disponibile su: https://www.ansa.it/sito/notizie/mondo/2023/03/23/trump-arrestato-ma-sono-foto-fake-di-una-intelligenza-artificiale_4ec43642-cda5-4121-88dc-a83ac0a0d2c6.html (ultima consultazione 10/04/2023).

[3] L. NICOLAO, *Zelensky chiede agli ucraini di arrendersi: è il primo (e malriuscito) deepfake sulla guerra*, in corriere.it, 17 marzo 2022, disponibile su: https://www.corriere.it/tecnologia/22_marzo_17/zelensky-chiede-ucraini-arrendersi-primo-malriuscito-deepfake-guerra-ad181d44-a5db-11ec-b9d0-9b9e3bb8f215.shtml (ultima consultazione 10/04/2023).

Occorre chiedersi, in particolare, in quale misura la normativa esistente sia idonea a comprendere e regolare i nuovi fenomeni. In questa direzione, si prenderà a particolare riferimento il diritto penale per provare a verificare se i reati attualmente previsti possano trovare applicazione anche rispetto alle pratiche nuove.
Tra queste, sarà indagato l'impiego del *deepfake* nella creazione di materiali sessualmente espliciti e nella produzione di *fake news*.

2. Il fenomeno del *deepfake*

Il *deepfake* è una tecnica per la sintesi di immagini basata sull'intelligenza artificiale impiegata per generare video o immagini originali combinando e sovrapponendo video o immagini già esistenti, il tutto attraverso una tecnica di apprendimento automatico, definita rete antagonista generativa[4]. Il Garante della privacy la definisce come un insieme di «foto, video e audio creati grazie a software di intelligenza artificiale 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»[5].
Nel tempo sono state introdotte specifiche applicazioni per produrre *deepfake*, a cominciare da quella, comparsa nel 2018, denominata FakeApp. Ad essa ne sono state affiancate via via altre, reperibili anche in modalità *open source*. Attraverso di esse lo sviluppo di *deepfake* è concesso di fatto a chiunque abbia capacità e competenze per la manipolazione di immagini e suoni.
La tecnica in esame si presta a molti usi, leciti ed illeciti. Tra i primi vanno annoverati quelli commerciali, poiché i prodotti del *deepfake* si prestano naturalmente alla registrazione del *copyright* e tale presupposto, unito all'interesse per i prodotti dell'intelligenza artificiale, può condurre inevitabilmente ad un nuovo mercato specifico delle immagini create artificialmente.
Si è ricordato, in proposito, che «tale tecnologia può avere applicazioni molto nobili, come la possibilità di restituire un corpo digitale a persone con gravi disabilità o di semplificare la ricerca medica»[6].
Vi sono, per altro verso, anche i possibili usi illeciti, quando l'immagine di un soggetto sia creata artificialmente per riferire allo stesso filmati, parole o pratiche comportamentali false e, quindi, per far apparire come vere e attendibili azioni o comportamenti mai accaduti nella realtà. Si tratta di pratiche volte a generare prolifici filoni di disinformazione, che sfruttano l'immagine di personalità pubbliche per veicolare

[4] N. BARNEY, *What is deepfake AI? A definition from TechTarget*, marzo 2023, disponibile su: https://www.techtarget.com/whatis/definition/deepfake (ultima consultazione 10/04/2023).
Oltre all'assenza di una definizione normativa di deepfake difetta l'esistenza anche di una definizione generalmente condivisa di intelligenza artificiale, come sottolinea F. DONATI, *Intelligenza artificiale e giustizia*, in *Riv. AIC,* 2020, 1, 415 ss.
[5] Cfr. scheda informativa del Garante della Privacy del 28 dicembre 2020, reperibile in www.garanteprivacy.it.
[6] S. FONTANA, *I deepfake stanno diventando una cosa seria*, 10 marzo 2023, in *facta.news*, disponibile su: https://facta.news/storie/2023/03/10/deepfake-cosa-seria/ (ultima consutazione 10/04/2023).

messaggi falsi e pericolosi. Si pensi ai casi di utilizzazione dei volti di persone note nell'ambito di filmati a contenuto pornografico o ai casi, più recenti, di diffusione di immagini di esponenti politici ed istituzionali intenti a realizzare azioni inesistenti[7].

Ebbene, in entrambi i casi, tanto per le finalità positive quanto per quelle negative, non esiste ancora una definizione normativa delle fattispecie. Ciò rende chiaro lo sforzo richiesto agli operatori del diritto per realizzare l'inquadramento giuridico delle pratiche di *deepfake* in continua formazione.

Nel nostro caso, l'impegno verrà qui rivolto all'approfondimento dei casi di produzione ed uso di *deepfake* per finalità pornografiche, tenuto conto dell'attenzione recentemente dedicata dal legislatore italiano al fenomeno del c.d. *revenge porn*.

Di seguito, come preannunciato, sarà indagato il possibile inquadramento giuridico dell'utilizzo del *deepfake* per la produzione di *fake news*.

3. Deepfake e diffusione non consensuale di contenuti sessualmente espliciti

Tra i vari impieghi della tecnologia *deepfake* ha registrato un enorme successo il suo utilizzo finalizzato a sovrapporre il volto di una persona ad una fotografia o un video sessualmente espliciti, in modo tale che la vittima si trovi ad essere rappresentata durante un atto sessuale mai avvenuto. Addirittura, l'intelligenza artificiale è capace di ''spogliare'' digitalmente la vittima, partendo da una foto che ritrae la stessa vestita e restituendo un'immagine di quello che potrebbe essere il suo corpo in nudità.

Secondo il report ''*The State of Deepfakes*'' [8] del progetto Sensity, a fine 2019 le foto e i video falsi realizzati attraverso l'utilizzo di tecniche basate sull'IA sono circa 15.000 (quasi il doppio rispetto all'anno precedente), e il 96% di questi materiali sono a contenuto pornografico. Un altro report di Sensity[9] ha poi evidenziato che a luglio 2020 oltre 104 mila donne sarebbero state vittime di *bot* di Telegram capaci ''spogliare'' le persone ritratte in una foto: il numero sarebbe poi cresciuto esponenzialmente nei successivi mesi, collocando l'Italia tra i primi quattro paesi al mondo per incidenza del fenomeno, insieme a Russia, Stati Uniti e Argentina.

Inoltre, nel 70% dei casi si tratterebbe di vittime ''comuni'', non di persone famose, come avveniva invece molto frequentemente all'inizio della diffusione degli algoritmi capaci di creare contenuti *deepfake*: questo perché, se inizialmente la quantità di dati di *training* che bisognava fornire al sistema di IA per consentirgli di realizzare contenuti che fossero credibili era molto elevata, oggi, grazie alle migliori capacità computazionali

[7] Un'applicazione esemplare di questa tecnica è quella inaugurata all'indomani dell'invasione russa dell'Ucraina, in un contesto bellico già preso di mira da disinformazione e propaganda. Il 17 marzo 2022 viene pubblicato un video *deepfake* nel quale un uomo con le sembianze del presidente ucraino Volodymyr Zelensky annuncia la resa del suo Paese.

[8] H. AJDER, G. PATRINI, F. CAVALLI, L. CULLEN, *The State of Deepfakes: Landscape, Threats, and Impact*, Sensity, 2019.

[9] H. AJDER, G. PATRINI, F. CAVALLI, *Automating Image Abuse: Deepfake bots on Telegram*, Sensity, 2020.

dei sistemi informatici, i dati sufficienti ad ottenere anche ottimi risultati sono decisamente di quantità inferiore.

Da quanto finora detto emerge che, in larghissima parte, il *deepfake* viene sfruttato in contesti patologici che rendono tale tecnologia uno strumento funzionale alla commissione di reati.

L'obiettivo che ci si pone in questo paragrafo è quello di individuare i possibili punti di contatto tra il fenomeno dei *deepfake* e quello della diffusione non consensuale di immagini e/o video sessualmente espliciti, nonché tentare di comprendere se l'ordinamento giuridico fornisca o meno adeguati strumenti per contrastare tali fenomeni.

La 'Diffusione illecita di immagini o video sessualmete espliciti', oltre ad essere una pratica drammaticamente diffusa nel nostro paese, costituisce una condotta specificamente sanzionata dal delitto di cui all'art. 612-ter c.p[10].

La disposizione in esame incrimina le condotte consistenti nell'invio, consegna, cessione, pubblicazione o, in generale, nella diffusione di «immagini o video a contenuto sessualmente esplicito, destinati a rimanere privati, senza il consenso delle persone rappresentate», sanzionandole con la reclusione da uno a sei anni e con la multa da 5.000 a 15.000 euro.

La norma mira a sanzionare non soltanto coloro i quali abbiano diffuso i contenuti dopo averli personalmente 'realizzati o sottratti' (ipotesi di cui al comma 1), ma anche tutti quei soggetti che, pur non avendo materialmente contribuito alla produzione o al furto degli stessi, abbiano invece contribuito a farli circolare al fine di recare nocumento alle persone rappresentate in tali contenuti (ipotesi di *'revenge porn'*, di cui al comma 2).

Adesso, si pensi ad un malintenzionato che intenda commettere un atto di *revenge porn*, ma che non abbia a disposizione materiali sessualmente espliciti della vittima designata: gli basterà ottenere una qualunque foto di quella persona (cosa semplicissima, vista la mole di immagini e video che pubblichiamo quotidianamente sui *social network*s) e creare tali materiali espliciti attraverso l'IA.

Ecco che il punto di contatto tra la tecnologia che crea i *deepfake* e la diffusione illecita di materiale sessualmente esplicito emerge chiaramente. Il problema, tuttavia, è capire se sia possibile sussumere la condotta di creazione e diffusione di pornografia *deepfake* entro la fattispecie di cui all'art. 612-ter c.p.: la disposizione in esame, infatti, non fa alcun riferimento a contenuti multimediali creati digitalmente o comunque non reali.

[10] Articolo 612-ter c.p., introdotto con art. 10, comma 1, L. 19 luglio 2019 n. 69. Per un approfondimento sull'art. 612-ter c.p., cfr. F. PANIZZO, *Luci ed ombre sulla 'vendetta pornografica' disciplinata dall'art. 612 ter c.p.,* in *Rivista AIAF*, 1, 2020.

Il principio costituzionale di legalità in materia penale, sancito al comma 2 dell'art. 25 Cost., in particolare nei suoi corollari di tassatività e divieto di analogia, dovrebbe distogliere da qualsiasi tentativo di estensione del perimetro applicativo della norma incriminatrice. Ad alimentare il dubbio la totale mancanza di giurisprudenza in materia, dovuta anche al fatto che l'art. 612-ter c.p. è norma introdotta solo di recente.

È interessante notare come invece i reati di pornografia minorile (art. 600-ter c.p.) e di detenzione di materiale pornografico minorile (art. 600-quater c.p.) siano integrati, in forza di quanto espressamente previsto dall'art. 600-quater 1 c.p., anche qualora le immagini siano ''virtuali'', cioè «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».

Quindi, sarebbe forse auspicabile, come primissimo passo, un'integrazione normativa dell'art. 612-ter, con la stessa logica che ha ispirato l'art. 600-quater 1.

Quanto poi al profilo del consenso della vittima, tale consenso si riferisce espressamente al solo momento della diffusione del materiale esplicito: è la mancanza di consenso alla diffusione a rendere la condotta illecita e quindi perseguibile.

Nel caso di *deepfake porn*, invece, il problema della mancanza del consenso potrebbe benissimo collocarsi già nel momento in cui avviene la creazione dell'immagine/video, ovvero nel momento in cui si fornisce all'algoritmo il materiale da manipolare: rispetto a tali momenti l'art. 612-ter tace, mostrandosi radicalmente inadeguato a fronteggiare il fenomeno già a monte.

Infine, si segnala che, nell'ottobre del 2020, il Garante per la protezione dei dati personali ha avviato un'istruttoria, chiedendo a Telegram (applicazione che è stata il principale canale di diffusione del *bot* capace di ''spogliare'' persone ritratte in foto) «di fornire informazioni, al fine di verificare il rispetto delle norme sulla protezione dei dati nella messa a disposizione agli utenti del programma informatico, nonché di accertare l'eventuale conservazione delle immagini manipolate e le finalità di una tale conservazione»[11].

Quanto al ruolo, agli obblighi e alle responsabilità dei fornitori di piattaforme *online* si dirà di più nei successivi paragrafi.

[11] *Deep fake: il Garante privacy apre un'istruttoria nei confronti di Telegram per il software che "spoglia" le donne*, Comunicato stampa del 23 ottobre 2020, disponibile su: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9470722 (ultima consultazione 10/04/2023).

4. La fattispecie del deepfake nella produzione di fake news

L'uso del *deepfake* si sta diffondendo con estrema facilità anche nella produzione delle *fake news*. Non a caso, la creazione di immagini artificiali si presta a conferire alle notizie false una forza di persuasività nettamente maggiore, ingigantendo nel lettore o nell'ascoltatore la credenza di trovarsi di fronte alla verità. Tutto ciò enfatizza le problematiche relative all'inquadramento giuridico di tali pratiche.

Una prima necessaria valutazione porta a chiedersi se vi sia rilevanza penale nell'uso del *deepfake* associato alla produzione di *fake news*. La risposta non può che muovere dall'individuazione delle ipotesi di reato emergenti nei casi di elaborazione e diffusione di notizie false.

Guardando ai reati contro la persona, la produzione di *fake news*, alimentata anche dall'uso di deepfake, può integrare il delitto di diffamazione (art. 595 c.p.), quando l'azione determini la lesione dell'altrui reputazione. Si tratta di un tipico reato contro l'onore della persona.

Peraltro, quando l'offesa venga arrecata mediante notizie su Internet e, dunque, «a mezzo di pubblicità», come recita l'art. 595, comma 3, c.p., ricorre necessariamente l'ipotesi più importante della diffamazione aggravata, per la quale è prevista la pena è della reclusione da sei mesi a tre anni o della multa non inferiore a euro 516.

Oltre alla persona e al suo onore, l'azione delle notizie false può mirare facilmente ad altri bersagli. Si pensi ai casi in cui la produzione di *fake news*, comprensive di notizie parziali o esagerate, anche accompagnate dall'uso di *deepfake*, miri all'obiettivo più alto dell'alterazione dell'ordine pubblico, inteso come garanzia di pace, di tranquillità e sicurezza collettiva[12]. È possibile richiamare per tali casi reati specifici, come quello di pubblicazione di notizie false, esagerate o tendenziose atte a turbare l'ordine pubblico (art. 656 c.p.) o quello di procurato allarme presso l'Autorità (art. 658 c.p.) o, ancora, quello di abuso della credulità popolare (art. 661 c.p.).

In tutti e tre i casi è fatto richiamo a reati contravvenzionali, per i quali le condotte punite sono quelle non solo di diffusione consapevole di notizie false, ma anche di diffusione colposa, dipendente da negligenza imprudenza o imperizia. La natura colposa del reato non sembra tuttavia compatibile con l'uso di *deepfake*. Infatti, se diffondere *fake news* può costituire reato anche solo procedendo alla condivisione di post sui social senza verificare preventivamente la notizia, nel caso dell'uso di *deepfake* appare inverosimile che la diffusione della notizia possa avvenire al di fuori di un comportamento intenzionale e quindi doloso.

Con riferimento specifico al reato di procurato allarme presso l'Autorità, si è affermato, in giurisprudenza, che esso «è configurabile anche nel caso in cui l'annuncio di un disastro, di un infortunio o di un

[12] Cfr. F. ANTOLISEI, *Manuale di diritto penale*, Parte Speciale, II, a cura di L. CONTI, Milano 1991, 227 ss.; C. FIORE, *Ordine pubblico* (dir. pen.), in *Enc. Dir.*, XXX, Milano, 1980, 1084 ss. Cfr. anche Corte cost. 16 Marzo 1962, n. 19 che definisce l'ordine pubblico come «ordine legale su cui poggia la convivenza civile».

pericolo inesistente sia "mediato", cioè non effettuato direttamente alle forze dell'ordine, ma ad un privato, purché, per l'apparente serietà del suo contenuto, risulti idoneo a provocare allarme nelle Autorità, determinandone l'intervento anche d'ufficio»[13].

Il pensiero conduce alle numerose notizie false, prodotte e diffuse a mezzo dei social media durante la pandemia da Covid-19, contenenti l'annuncio di pericoli o allarmi specifici e tali da raggiungere, oltre ad un vasto pubblico, anche la pubblica autorità, generando allarme presso essa.

Si può immaginare quale potenza possa essere conferita a messaggi di tale tenore dalla associazione di false immagini o foto, riferite a personaggi pubblici, ritratti in condizioni di grave pericolo o grave disagio fisico. Alla forza della tecnologia pare contrapporsi, peraltro, l'insospettata modernità del nostro ordinamento giuridico penale, che pare dimostrare una buona resilienza, quantomeno in astratto, a questi nuovi fenomeni degenerativi collegati all'uso dell'intelligenza artificiale. In termini di politica del diritto ci si può, tuttavia, interrogare sull'efficacia dell'impostazione sanzionatoria del nostro ordinamento rispetto a possibili approcci alternativi di stampo più prevenzionistico. Si pensi a quelli previsti dal Digital Services Act[14], in particolare ai poteri di coordinamento e direzione della Commissione europea in situazione di crisi o alle iniziative di sensibilizzazione degli utenti promosse spontaneamente dagli operatori dei di servizi *online*.

5. Il contributo dell'Unione europea alla regolamentazione del fenomeno

Nei precedenti paragrafi, si è osservato come la tecnologia dei *deepfake* sia potenzialmente funzionale alla commissione di gravi reati. Inoltre, si è constatato come l'ordinamento giuridico italiano non disciplini questi fenomeni con una normativa *ad hoc*, e come gli strumenti, soprattutto di natura penale, che abbiamo a disposizione presentino lacune forse insuperabili e siano quindi inadeguati a fronteggiare gli utilizzi abusivi della tecnologia.

In questo paragrafo, invece, si tenterà di illustrare le prese di posizione dell'Unione europea, facendo riferimento in particolare all'AI Act e al Digital Services Act.

Anzitutto, l'AI act[15] prevede che, per taluni sistemi specifici di IA, vengano imposti soltanto obblighi minimi di trasparenza: ciò avviene «in particolare quando vengono utilizzati chatbot o deep fake»[16]. L'art. 52 paragrafo 3, infatti, si limita a stabilire che «gli utenti di un sistema di IA che genera o manipola immagini o

[13] Cassazione penale, Sez. I, sentenza n. 26897 del 12 giugno 2018.
[14] Regolamento (UE) 2022/2065 del Parlamento Europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).
[15] Proposta di Regolamento del Parlamento Europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'unione ("AI Act").
[16] Relazione introduttiva all'AI Act, paragrafo 1.1.

contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente», salvo poi prevedere delle eccezioni a tale regola, nel caso in cui la manipolazione di tali contenuti sia autorizzato dalla legge per «accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione, delle arti e delle scienze […]».

Si può dubitare che questa previsione riuscirà ad avere un impatto significativo nei confronti dei fenomeni patologici analizzati nei precedenti paragrafi.

Rispetto alla diffusione illecita di materiale sessualmente esplicito realizzato tramite l'IA, si ritiene: da un lato, che il semplice obbligo di rendere noto che il contenuto sia finto difficilmente spiegherà efficacia deterrente verso gli autori di tale condotta, dato che molto spesso la domanda di questi contenuti non dipende dalla loro autenticità[17]; dall'altro lato, che si sia persa un'occasione per prendere una posizione netta e specifica sui sistemi di IA che creano *deepfake* sessualmente espliciti.

Piuttosto, sarebbe stato opportuno introdurre una disciplina, se non di divieto assoluto di questi sistemi di IA, quantomeno rigorosa, che responsabilizzasse tanto i detentori dell'algoritmo che crea i *deepfake* quanto chi richiede il ''servizio'' senza avere il consenso delle persone rappresentate nei contenuti.

In sintesi, bisognerebbe prevenire il più possibile la realizzazione non consensuale, tramite sistemi di IA, di contenuti sessualmente espliciti, essendo qualsiasi intervento *ex post*, come è la semplice indicazione del fatto che quel contenuto sia falso, inidoneo a garantire una tutela effettiva della dignità delle persone.

Per quanto concerne il problema delle *fake news*, invece, la questione si fa più delicata, e prendere una decisione netta in materia è probabilmente più complicato. La tecnologia del *deepfake*, come si è già avuto modo di osservare, può dare adito a fenomeni di diversa natura: ad esempio, una cosa è la foto che ritrae Papa Francesco con un piumino bianco *oversize*, altra cosa è il video in cui Zelensky ordina all'esercito ucraino di deporre le armi di fronte all'avanzata dei russi. Questi sono esempi estremi, ma spesso il confine tra diffusione di contenuti a sfondo satirico e diffusione di notizie false si fa estremamente labile, con la conseguenza che introdurre discipline eccessivamente rigorose e limitative rischia di scontrarsi con diritti costituzionalmente rilevanti (e lo stesso avviene con discipline eccessivamente larghe e indulgenti): si tratta del classico problema del bilanciamento tra diritto alla libera manifestazione del pensiero e tutela della reputazione e della riservatezza delle persone, nonché dell'ordine pubblico e del regolare e pacifico svolgimento della vita politica, economica e sociale della collettività. Problema rispetto al quale, come si è

[17] P. GRADY, *EU Proposals Will Fail to Curb Nonconsensual Deepfake Porn*, 2023, disponibile su: https://datainnovation.org/2023/01/eu-proposals-will-fail-to-curb-nonconsensual-deepfake-porn/ (ultima consultazione 10/04/2023).

già visto all'inizio di questo paragrafo, l'AI act non prende una posizione netta, limitandosi ad affermare che l'uso del sistema di IA non incontra i limiti dell'art. 52, paragrafo 3, comma 1, se necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti, «fatte salve le tutele adeguate per i diritti e le libertà dei terzi».

Quanto al Digital Services Act, già in vigore dal 16 novembre 2022, tale regolamento si pone anche l'obiettivo di migliorare la moderazione dei contenuti sulle piattaforme *online* (tra cui i *social networks*), al fine di affrontare in maniera più efficace le preoccupazioni sulla diffusione di contenuti illegali e sulla disinformazione.

Il regolamento stabilisce, come regola generale[18], che il prestatore del servizio non può essere considerato responsabile delle informazioni memorizzate su richiesta di un destinatario del servizio, a condizione che il prestatore stesso non sia effettivamente a conoscenza delle attività o dei contenuti illegali e che, non appena ne venga a conoscenza agisca immediatamente per rimuoverli o per disabilitare l'accesso agli stessi, lasciando comunque impregiudicata la possibilità che un'autorità giudiziaria o amministrativa dei singoli Stati membri esiga che il prestatore impedisca o ponga fine ad una violazione.

Tuttavia, tale conoscenza o consapevolezza effettiva non può essere considerata acquisita per il solo motivo che tale prestatore è consapevole, in senso generale, del fatto che il suo servizio è utilizzato anche per memorizzare contenuti illegali.

Infatti, ai prestatori di servizi intermediari non è imposto alcun obbligo generale di sorveglianza sulle informazioni che gli stessi trasmettono o memorizzano, anche se si tratta di contenuti o attività illegali[19].

Il prestatore può acquisire la conoscenza della natura illegale del contenuto attraverso indagini volontarie oppure tramite le segnalazioni presentategli da persone o enti: infatti, è importante che tutti i prestatori di servizi di memorizzazione di informazioni, indipendentemente dalle loro dimensioni, predispongano meccanismi di segnalazione e azione di facile accesso e uso che agevolino la notifica al prestatore di servizi di informazioni specifiche che la parte notificante ritiene costituiscano contenuti illegali, in base alla quale il prestatore può decidere se condivide o meno tale valutazione e se intende rimuovere detti contenuti o disabilitare l'accesso agli stessi[20].

Il regolamento richiede anche che i fornitori di piattaforme *online* di dimensioni molto grandi valutino i rischi sistemici derivanti dalla progettazione, funzionamento e uso dei loro servizi, nonché dai possibili abusi da parte dei destinatari dei servizi, in modo tale da poter adottare misure adeguate ad attenuare questi rischi.

[18] Art. 6, Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, 19 ottobre 2022.
[19] Art. 8, Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, 19 ottobre 2022.
[20] Considerando n. 50 Regolamento (UE) 2022/2065.

Nel valutare questi rischi sistemici, i fornitori devono prestare particolare attenzione al modo in cui i loro servizi siano utilizzati per diffondere o amplificare la diffusione di contenuti «fuorvianti o ingannevoli»[21]: infatti, è risaputo come la maggior parte delle piattaforme utilizzi algoritmi personalizzati, progettati in modo da mostrare agli utenti i contenuti più rilevanti e coinvolgenti. Allo stesso modo, le campagne di disinformazione vengono strutturate in modo tale da sfruttare il funzionamento di questi algoritmi, rendendo le notizie false attraenti per gli utenti, favorendone la rapida diffusione[22].

Ancora, il regolamento chiede alla Commissione di incoraggiare l'elaborazione di codici di condotta volontari: l'adesione a un determinato codice ed il suo rispetto da parte di una piattaforma *online* di dimensioni molto grandi possono essere considerate misure adeguate di attenuazione dei rischi[23].

Inoltre, si prevede un apposito meccanismo di risposta alle crisi per le piattaforme *online* di dimensioni molto grandi: quando si verificano circostanze eccezionali che possano comportare una minaccia grave per la sicurezza pubblica o la salute pubblica nell'Unione, la Commissione, su raccomandazione del Comitato europeo per i servizi digitali, può chiedere ai prestatori di piattaforme *online* di avviare con urgenza apposite misure, che possono includere, ad esempio, l'adeguamento dei processi di moderazione dei contenuti e l'aumento delle risorse destinate alla moderazione, l'adeguamento dei sistemi algoritmici, l'adozione di misure di sensibilizzazione e la promozione di informazioni affidabili[24].

Infine, oltre al meccanismo di risposta alle crisi, la Commissione può avviare, per le piattaforme *online* di dimensioni molto grandi, l'elaborazione di protocolli di crisi volontari per coordinare una risposta rapida e collettiva. I protocolli dovrebbero essere attivati solo per un periodo di tempo limitato e le misure adottate dovrebbero essere limitate a quanto strettamente necessario per far fronte alla situazione di crisi e coerenti con il regolamento: in particolare non devono costituire per i fornitori di piattaforme online un obbligo generale di sorveglianza sulle informazioni che trasmettono o memorizzano[25].

Alla luce di quanto detto, sembra che il Digital Services Act introduca strumenti molto incisivi, anche se con degli evidenti limiti.

In particolare, con riferimento al fenomeno della diffusione di *fake news*, il regime di «responsabilità condizionata»[26] dei fornitori, coniugato con gli obblighi procedurali e sostanziali di valutazione e di

[21] Considerando n. 84, Regolamento (UE) 2022/2065.
[22] Si pensi alle attività di propaganda della *Internet Research Agency* russa già agli inizi dell'occupazione della Crimea e del sostegno alla campagna elettorale di Donald Trump. Il *modus operandi* dell'agenzia consiste nell'ingannare gli algoritmi attraverso la gestione coordinata di moltissimi *account* falsi: non a caso, è stata definita una «*troll factory*».
Si veda anche: M. BASTOS, J. FARKAS, *"Donald Trump Is My President!": The Internet Research Agency Propaganda Machine*, disponibile su: https://journals.sagepub.com/doi/pdf/10.1177/2056305119865466 (ultima consultazione 10/04/2023).
[23] Cfr. Considerando n. 104 Regolamento (UE) 2022/2065.
[24] Art. 36, Regolamento (UE) 2022/2065.
[25] Art. 48, Regolamento (UE) 2022/2065.
[26] Audizione del Presidente del Garante per la protezione dei dati personali, Pasquale Stanzione, 12 gennaio 2021.

mitigazione dei rischi di cui sopra, restituisce l'immagine di un sistema capace di rendere più sicuro e trasparente lo spazio digitale. Allo stesso tempo, però, il DSA ha deluso alcuni commentatori[27], soprattutto per quanto riguarda il meccanismo di risposta alle crisi e la possibilità di adottare protocolli di crisi: si avverte il rischio che la Commissione, sentendosi sotto pressione, obblighi le grandi piattaforme a adottare come misura la semplice rimozione delle *fake news* e dei loro canali di diffusione[28].

Tuttavia, la richiesta di rimozione sistematica di tali informazioni dovrebbe inevitabilmente fare affidamento su sistemi informatici basati sul *machine-learning*, che sono notoriamente imprecisi, non tengono conto del contesto, e quindi hanno spesso un impatto su contenuti genuini o comunque leciti, in totale spregio alla necessità di ricercare, nel singolo caso concreto, quel bilanciamento tra tutela del diritto alla libera manifestazione del pensiero e tutela del pacifico svolgimento della vita sociale e politica della collettività [29].

Con riferimento, invece, al fenomeno della diffusione illecita di materiale sessualmente esplicito, il regolamento richiede che i fornitori di piattaforme *online* di grandi dimensioni, e in particolare quelli utilizzati per la diffusione al pubblico di contenuti pornografici, adempiano diligentemente a tutti gli obblighi stabiliti dal regolamento stesso in relazione ai contenuti pornografici illegali, garantendo il trattamento rapido delle segnalazioni da parte delle vittime e la rimozione di tali contenuti senza indebito ritardo[30]. Qui il limite rimane quello per cui il ''semplice'' tentativo di contenere la diffusione del contenuto illecito non è di per sé sufficiente a tutelare la dignità della vittima: rispetto a questo fenomeno, si ribadisce quindi che si dovrebbe lavorare anche e soprattutto sull'adozione di meccanismi e strumenti di natura preventiva e/o deterrente.

Al riguardo, si segnala che la proposta di Direttiva sulla lotta alla violenza contro le donne e alla violenza domestica chiede agli Stati membri di provvedere affinché siano punite come reato le condotte intenzionali consistenti nella produzione o manipolazione e successiva diffusione di materiali tali da far credere che una persona partecipi ad atti sessuali, senza il consenso dell'interessato/a[31]: in tal modo si coprirebbe la lacuna normativa presente nel nostro ordinamento segnalata al paragrafo 3 di questa trattazione[32].

---

[27] Cfr. D. BUIJS, I. BURI, *The DSA's crisis approach: crisis response mechanism and crisis protocols*, 21 febbraio 2023, in *DSA Observatory*, disponibile su: https://dsa-observatory.eu/2023/02/21/the-dsas-crisis-approach-crisis-response-mechanism-and-crisis-protocols/ (ultima consultazione 10/04/2023).

[28] Un esempio: subito dopo l'inizio dell'invasione dell'Ucrania del 2022, la Commissione europea dispone il ban dell'agenzia di stampa russa *Sputnik* e del canale televisivo *Russia Today*.

[29] Z. MEYERS, ''*Will the Digital Services Act save Europe from disinformation?*'', 2022, disponibile su: https://www.cer.eu/insights/will-digital-services-act-save-europe-disinformation (ultima consultazione 10/04/2023).

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

[31] Art. 7, lett. b), Proposta di direttiva del Parlamento europeo e del Consiglio sulla lotta alla violenza contro le donne e alla violenza domestica COM/2022/105 final.

[32] Si fa tuttavia notare come la norma in parola faccia riferimento esclusivamente a contenuti che ritraggono una persona mentre partecipa «ad atti sessuali», lasciando così scoperto il caso di contenuti che rappresentino la persona in nudità senza che vi sia un comportamento sessualmente esplicito in atto.

6. Il mercato dei servizi online di fronte alle sfide poste dalla tecnologia *deepfake*

Le cronache degli ultimi mesi segnalano un grande interesse dell'imprenditoria per i prodotti dell'intelligenza artificiale, che si sviluppa su tutto il pianeta, dagli Stati Uniti, alla Cina, all'Europa. È un interesse che alimenta speranze di sviluppo, ma suscita, insieme, nuove paure.

Le cronache di questi giorni raccontano le grandi attese e gli investimenti che le grandi aziende dell'*e-commerce* e dello sviluppo tecnologico, come Alibaba, Baidu, Tesla, Google, Apple, stanno compiendo nella creazione di piattaforme di intelligenza artificiale[33].

Emergono, tuttavia, forti preoccupazioni sui livelli di controllo di tale sviluppo, a cominciare dal ruolo riservato all'etica quale regolatore e limitatore delle nuove derive tecnologiche.

Lo scontro tra queste due dinamiche è destinato a creare molta confusione nei mercati, se è vero che chi oggi si espone ad invocare una moratoria temporanea dello sviluppo degli investimenti in intelligenza artificiale, richiamando un «pericolo per l'umanità»[34], è anche il protagonista di disinvestimenti nel settore dello sviluppo etico[35].

Anche la diffusione della tecnologia *deepfake* espone il mercato a grandi opportunità, ma anche a gravi rischi, che riguardano la distorsione della concorrenza e la tutela della reputazione aziendale.

Su tale presupposto, proprio intorno alla linea sottile che divide le opportunità dai rischi prende corpo l'interrogativo su quali politiche aziendali siano in atto rispetto alla diffusione del *deepfake*.

Emerge, in argomento, una sorta di dualismo: da un lato si pone chi invoca una precisa regolamentazione del fenomeno, per reprimerne gli usi distorti e abusivi; dall'altro, si afferma l'opinione di chi ritiene che il *deepfake* costituisca prevalentemente una tecnologia da sostenere, nell'ottica di sfruttare il più possibile le opportunità commerciali che essa è capace di offrire.

Nel mezzo si assiste all'iniziativa di singole aziende, che sperimentano iniziative in proprio per dimostrare sensibilità ai pericoli dell'intelligenza artificiale.

Meta (ex Facebook), ad esempio, ha comunicato, già nel 2020, il divieto di *deepfake* sulle proprie piattaforme, preannunciando la rimozione dei contenuti modificati dall'intelligenza artificiale a protezione del rischio di inganno delle persone. Essa, tuttavia, concede l'utilizzazione di *deepfake* nell'ambito di messaggi contenenti

[33] Cfr., ad esempio, l'inchiesta "*Le frontiere della tecnologia*", in *Il Sole 24 ore* del 30 marzo 2023 e del 12 aprile 2023.

[34] Il riferimento è alla lettera aperta, firmata da oltre mille figure molto influenti della c.d. *Silicon Valley*, inoltrata alle aziende impegnate nello sviluppo dell'intelligenza artificiale per chiedere uno stop di sei mesi degli investimenti per fare il punto, darsi delle regole e poi ripartire con maggior controllo e consapevolezza.

[35] Cfr. L. De Biase, *Cruciale che anche i privati tornino a pensare all'etica*, in *Il Sole24 ore*, 30 marzo 2023, disponibile su: https://24plus.ilsole24ore.com/art/intelligenza-artificiale-perche-e-cruciale-che-anche-privati-tornino-pensare-all-etica-AEwXjQBD?s=hpl (ultima consultazione 10/04/2023).

satira o parodie, in tal modo ammettendo una deroga importante ad una politica aziendale apparentemente molto rigorosa.

Altre aziende implementano politiche analoghe. Nel sito di Reddit si legge che la piattaforma «non consente contenuti che impersonano individui o entità in modo fuorviante o ingannevole». Sono inclusi esplicitamente i *deepfake* «presentati per fuorviare o falsamente attribuiti a un individuo o entità»[36].

Microsoft ha lanciato *Microsoft Video Authenticator*, un programma che può analizzare una foto o un video per fornire una percentuale di probabilità che i *media* siano stati manipolati artificialmente[37].

Anche Google ha creato un apposito *database* in grado di smascherare l'uso di *deepfake*. Così, all'interno del catalogo Google sono disponibili diversi filmati *fake* generati per allenare i meccanismi di riconoscimento dei falsi e consentire di individuare e bloccare i materiali non genuini.

Di segno opposto è la scelta di Twitter di disattivare la moderazione dei contenuti (in particolare quelli che riguardano la disinformazione sull'epidemia da Covid-19)[38], lasciando che sia la stessa comunità *online* degli utenti della piattaforma a svolgere questa attività di moderazione. Altra mossa peculiare è stata quella del *patron* di Twitter, Elon Musk, di ripristinare moltissimi *account* in precedenza sospesi dalla piattaforma (tra cui quello di Donald Trump), a seguito di una sorta di ''referendum'' [39] che ha visto il diretto coinvolgimento del ''popolo di Internet''.

In generale, la tecnica di elaborazione di software di rilevamento dei *deepfake* si presenta potenzialmente molto incisiva rispetto alla prevenzione di possibili pratiche distorte di condizionamento ingannevole. Essa assume scopi anche repressivi, ogni qual volta sia possibile smascherare condotte aziendali illecite improntate a veicolare messaggi ingannevoli e strumentali.

---

[36] REDDIT, *Updates to Our Policy Around Impersonation*, 2020, disponibile su: https://www.reddit.com/r/redditsecurity/comments/emd7yx/updates_to_our_policy_around_impersonation (ultima consultazione 10/04/2023).

[37] MICROSOFT, *New Steps to Combat Disinformation*, 2020, disponibile su: https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes (ultima consultazione 10/04/2023).

[38] *Twitter ha interrotto moderazione su disinformazione Covid,* in *ansa.it*, 30 novembre 2022, disponibile su: https://www.ansa.it/sito/notizie/tecnologia/internet_social/2022/11/30/twitter-ha-interrotto-moderazione-su-disinformazione-covid_b25b9911-f975-4d94-b172-ded97c8aaf11.html (ultima consultazione 10/04/2023).

[39] La politica perseguita da Twitter è promossa come manifestazione di un'ideologia libertaria e democratica, ma si presta a consegnare lo spazio digitale alla totale anarchia. Lo ha sottolineato il commissario europeo per il mercato interno Thierry Breton, il quale ha affermato che Twitter "dovrà implementare politiche utente trasparenti, rafforzare in modo significativo la moderazione dei contenuti e proteggere la libertà di parola, affrontare la disinformazione con determinazione e limitare la pubblicità mirata" (cfr. https://www.reuters.com/technology/twitter-has-huge-work-ahead-eus-breton-tells-musk-2022-11-30). Anche Il Garante della Privacy, con un duro intervento dell'avv. Guido Scorza, componente del Collegio, ha definito «inaccettabile» il sondaggio proposto da Musk (cfr. https://www.agendadigitale.eu/cultura-digitale/scorza-perche-il-sondaggio-di-musk-su-trump-e-inaccettabile-ne-va-dei-diritti-di-tutti/).

Si tratta, peraltro, di iniziative, spontanee ed unilaterali, che nulla garantiscono rispetto alla qualità dei prodotti e alla loro conformità a standard predefiniti e condivisi. Il problema è tanto più importante tenuto conto che non esistono controlli sulle modalità di funzionamento di tali programmi di rilevazione.

Ne consegue che non è dato sapere come avviene la rilevazione dei *deepfake* e la loro successiva gestione. In particolare, la loro rimozione, una volta riconosciutane l'esistenza, quasi mai opera automaticamente, mentre prevale la valutazione sulle intenzioni perseguite con l'uso degli stessi, che richiama operazioni altamente soggettive.

Rimane, in ogni caso, l'apporto positivo che i programmi aziendali richiamati possono fornire alle forze dell'ordine per il contrasto dell'uso illecito dei *deepfake* e la tutela delle vittime. Questo aspetto potrà molto aiutare, in attesa di regole nazionali e internazionali sull'uso dei *deepfake*, lo sviluppo responsabile e legale delle nuove tecnologie.

7. Conclusioni

I mutamenti tecnologici stanno raggiungendo livelli di complessità con una rapidità e imprevedibilità tali da rendere difficile l'opera di adeguamento e di riflessione del giurista, nonché quella decisoria e di regolamentazione dei *policy makers*. E se vale ancora la riflessione di Francesco Carnelutti, per cui chi conosce solo il diritto non conosce nemmeno il diritto, l'avvento delle nuove tecnologie impone non solo di provare a comprendere altri linguaggi, ma anche e soprattutto di prendere in considerazione l'eventualità che alcune delle categorie con le quali si è soliti ragionare possano mostrarsi inadeguate a contenere taluni fenomeni di fronte ai quali siamo posti dal progresso tecnologico. Si tratta evidentemente di questioni che evocano la necessità di un approccio multidisciplinare: tanto nella fase dello sviluppo tecnico, quanto in quella della regolamentazione delle nuove teconlogie è fondamentale coinvolgere oltre a informatici, ingegneri e giuristi, anche sociologi ed eticisti, in modo tale da non tralasciare alcun aspetto problematico che possa emergere nel corso di questi processi.

Il diritto da solo non è sufficiente a gestire il progresso tecnologico, ma continua a rappresentare ciò che deve assolutamente guidarlo, nell'ottica di promuovere i suoi vantaggi e potenzialità da un lato, e di moderare i suoi rischi dall'altro.

A tal fine, è necessario guardare all'essenza del diritto, quale baluardo posto a garanzia e difesa dei principi fondamentali che governano la vita collettiva e individuale dell'essere umano e della sua dignità.