La questione *deepfake* in Italia: una panoramica*

Veronica Azzali – Nicol Ellecosta

Abstract

Sin dalla loro comparsa nel 2017, i *deepfake* hanno attirato notevole attenzione. Questi sofisticati strumenti software consentono la manipolazione di immagini, video e audio preesistenti. Ad esempio, è possibile incollare il viso di una persona sul corpo di un'altra, ricreando, quindi, situazioni paradossali, che appaiono reali anche agli occhi più esperti. Alcuni individui percepiscono i *deepfake* come una minaccia alla democrazia, una forza inarrestabile che la società è destinata a subire irrimediabilmente. Tuttavia, negli ultimi anni, diversi governi, piattaforme online e agenzie di informazione hanno intrapreso misure per contenere questo fenomeno, adottando strategie quali il blocco degli utenti che diffondono informazioni false, il fact checking e l'introduzione di normative legislative. L'Italia è tra le giurisdizioni che stanno cercando di contrastare l'uso improprio dei *deepfake*. Nel corso di questo saggio, esamineremo come le leggi italiane si concentrino principalmente sul furto d'identità e sulle implicazioni dei *deepfake* nel contesto della pornografia.

Deepfakes have attracted considerable attention since their emergence in 2017. These sophisticated software tools allow the manipulation of pre-existing images, videos, and audio. For example, paradoxical situations that appear real to even the most experienced eyes can be created by pasting one person's face onto another's body. Some see deepfakes as a threat to democracy, an unstoppable force from which society will suffer irreparably. In recent years, however, several governments, online platforms, and news agencies have taken steps to curb the phenomenon. They have adopted strategies such as blocking users who spread false information, fact-checking, and introducing legislation. Italy is among the countries trying to combat deepfake abuse. In this essay, we will examine how Italian legislation has focused primarily on identity theft and the impact of deepfakes in the context of pornography.

Sommario

1. La tana del Bianconiglio. – 2. Il mondo dei replicanti. – 3. Profondamente falsi. – 4. *Deepfake*-news. – 5. *Deepfake* e furto d'identità – 6. La regolamentazione in Italia – 7. Conclusioni.

* L'articolo è stato sottoposto, in conformità al regolamento della Rivista, a referaggio "a doppio cieco".

Keywords

deepfake – intelligenza artificiale – fake news – furto d'identità – pornografia

1. La tana del Bianconiglio

Gli appassionati di meme saranno probabilmente a conoscenza (se non addirittura state vittime) di un fenomeno molto particolare noto come "*rickrolling*". Acquisendo importanza a metà degli anni 2000[1], il *rickrolling* è un fenomeno della cultura meme caratterizzato da uno scherzo che consiste in un collegamento ipertestuale che reindirizza le persone al video musicale del singolo di Rick Astley "*Never Gonna Give You Up*"[2]. L'essenza di un *rickroll* risiede nella sua natura inaspettata; i destinatari del collegamento ipertestuale (condiviso in post social o attraverso semplici messaggi su WhatsApp) di solito si aspettano di accedere a contenuti rilevanti o coinvolgenti[3], per poi vedersi costretti ad ascoltare alcune note del famoso pezzo del 1987.

Gli appassionati di calcio italiano avranno verosimilmente incontrato una variante specifica di *rickroll* in cui il protagonista non è Rick Astley, bensì l'allenatore della squadra di calcio Juventus, Massimiliano Allegri. Fatta eccezione per la paradossale natura della situazione, il video potrebbe persino apparire autentico, data la fedele riproduzione della voce dell'allenatore toscano e l'eccellente mimica facciale che si sovrappone non solo a quella dell'originale, ma anche al distintivo *cast* di supporto.

Sorge quindi la domanda su come sia possibile non solo replicare in modo così accurato tratti fisionomici di Massimiliano Allegri, ma anche sostituire la voce di Rick Astley. Tale fenomeno costituisce un esempio paradigmatico dell'abilità di alterare un contenuto veritiero attraverso la fedele riproduzione di volti, espressioni facciali e voci umane. Tali adattamenti, denominati *deepfake*, vengono elaborati attraverso una forma particolare di intelligenza artificiale (IA). Nel corso degli ultimi anni, numerosi esempi di video comici *deepfake* sono stati condividi nella rete web, inclusa una delle attuali tendenze virali sui social media: la ricreazione di *roster* di personaggi fittizi tratti dal popolare gioco di combattimento "*Mortal Kombat*". Questa tecnologia consente la simulazione virtuale di scontri tra figure politiche come Joe Biden e Donald Trump al di fuori del contesto elettorale, o immaginarie contese fisiche tra personalità come Jeff Bezos ed Elon Musk con azioni di combattimento.

Tuttavia, per coloro che sono a conoscenza del fenomeno dei *deepfake*, il termine non può essere semplicemente associato alla cultura dei *meme*, ma assume una connotazione nettamente più inquietante[4,5,6]. Sfortunatamente, non sono rari i casi in cui questa

[1] K. Knowles, *What the heck is… Rickrolling?*, in *forbes.com*, 2 febbraio 2016.

[2] R. Astley, *Never Gonna Give You Up*, 1987.

[3] K. Knowles, *What the heck is… Rickrolling?*, cit.

[4] D. Fallis, *The Epistemic Threat of Deepfakes*, in *Philosophy & Technology*, 34, 2020, 623 ss.

[5] M. Westerlund, *The Emergence of Deepfake Technology: A Review*, in *Technology Innovation Management Review*, 9, 2019, 39 ss.

[6] M. Yankoski - W. Scheirer - T. Weninger, *Meme warfare: AI countermeasures to disinformation should focus on*

tecnologia viene sfruttata a fini deprecabili. Celebrità, ad esempio, sono state vittime di *deepfake* utilizzati per la creazione di falsi filmati pornografici, come nel noto caso di Scarlett Johansson. L'attrice hollywoodiana è stata oggetto di numerose manipolazioni *deepfake* in cui il suo volto è stato sovrapposto a quello di autentiche attrici pornografiche in video espliciti e osceni[7,8]. Uno di tali video ha raggiunto persino 1,5 milioni di visualizzazioni attraverso un rinomato sito per adulti[9].

In conclusione, si può affermare che i *deepfake*, analogamente ad altre forme di intelligenza artificiale, presentano aspetti positivi e negativi. Tuttavia, la questione rimane irrisolta: chi sono gli artefici dei *deepfake* e quali sono le loro motivazioni? Pertanto, risulta essenziale fornire una breve spiegazione su (a) cos'è l'intelligenza artificiale; (b) come e perché sono nati i *deepfake*; (c) le modalità di utilizzo di questa tecnologia; e (d) le possibili conseguenze legali della creazione e dell'utilizzo di *deepfake*.

2. Il mondo dei replicanti

In primis è necessario partire con una breve spiegazione, sicuramente non esaustiva, di cos'è l'intelligenza artificiale[10].

L'intelligenza artificiale studia i fondamenti teorici, le metodologie e le tecniche che consentono di progettare sistemi *hardware* e sistemi di programmi software atti a fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana[11].

In sintesi, quando ci riferiamo all'Intelligenza Artificiale, facciamo generalmente riferimento alla disciplina che si occupa dello studio e dello sviluppo di sistemi tecnologici e informatici che cercano di emulare i comportamenti, il pensiero e le capacità umane. Il nucleo di questa tematica risiede proprio in questo concetto: l'IA non mira a replicare l'intelligenza umana, bensì cerca di riprodurla[12]. Occorre però individuare chi, o meglio cosa definisce ciò che è "intelligente".

Il dizionario statunitense Merriam Webster[13] definisce l'intelligenza come (a) la capacità di apprendere, comprendere e affrontare situazioni difficili; (b) la capacità di applicare le conoscenze per manipolare l'ambiente o pensare in modo astratto; e infine, (c) la capacità di utilizzare la ragione, aspetto particolarmente interessante e difficile da trasmettere a una macchina.

popular, not perfect, fakes, in *Bulletin of the Atomic Scientists*, 77, 2021, 119 ss.

7 W. Brown - D. H. Fleming, *Celebrity headjobs: or oozing squid sex with a framed-up leaky {Schar-JØ}*, in *Porn Studies*, 7, 2020, 357 ss.

8 A. Santangelo, *Il futuro del volto nell'era dei deep fake*, in M. Leone (a cura di) *Il metavolto*, FACETS Digital Press, 2022, 19 ss.

9 G. Meikle, *Deepfakes*, Cambridge, 2022.

10 J.H. Fetzer, *What is Artificial Intelligence?*, in J.H. Fetzer, *Artificial Intelligence: Its Scope and Limits. Studies in Cognitive Systems*, Dordrecht, 1990, 3 ss.

11 F. Amigoni - V. Schiaffonati - M. Somalvico, *Intelligenza artificiale*, in *treccani.it*, 2008.

12 Ivi, para. 2.

13 Merriam-Webster, *intelligence*, in *merriam-webster.com*.

Di conseguenza, ritenuto l'essere umano come essere senziente ed intelligente, è possibile che le macchine siano intelligenti? Sarebbe, infatti, difficile considerare intelligenti sentimenti come rabbia e gelosia. Sebbene occasionalmente possano risultare utili, sicuramente non vorremmo vederli espressi da una macchina. Pertanto, la ricerca si concentra sull'identificazione dei criteri che definiscono l'intelligenza[14,15].

Le applicazioni dell'IA sono molteplici. Ad esempio, nel mondo della musica, l'uso dell'IA ha permesso di creare una nuova canzone dei Beatles estrapolando la voce di John Lennon da una vecchia demo, come dichiarato da Paul McCartney[16]. Piattaforme di streaming musicale, come Apple Music e Spotify, utilizzano l'IA per generare playlist personalizzate in base alle preferenze musicali degli utenti[17]. L'IA è progredita al punto da essere in grado di comporre pezzi musicali autonomamente[18].

Anche nel settore dei videogiochi, l'IA consente ai personaggi non giocanti (NPC) di comportarsi in modo simile agli esseri umani, migliorando così l'esperienza di gioco[19,20]. Inoltre, il *pathfinding* è una tecnologia che permette ai personaggi di muoversi nel mondo di gioco senza collidere con ostacoli e raggiungere il loro obiettivo[21,22].

L'IA contribuisce anche al miglioramento delle capacità di scrittura attraverso pro-

---

14 J.H. Fetzer, *What is Artificial Intelligence?*, cit.

15 In merito a questa tematica, sono stati condotti diversi studi, tra cui il rinomato test di Turing proposto da Alan Turing nel suo articolo "*Computing machinery and intelligence*" pubblicato sulla rivista *Mind* nel 1950. Il test di Turing misura la capacità di una macchina di manifestare un comportamento intelligente attraverso il "gioco dell'imitazione". Questo coinvolge tre partecipanti: (A) una macchina, (B) una persona e (C) un valutatore umano. Se il valutatore C, interagendo con A e B, non riesce a distinguere con certezza quale sia la macchina e quale l'essere umano, il test è considerato superato.
Un altro esperimento degno di nota è quello della "*Chinese Room*" proposto da John Searle nel suo articolo "*Minds, brains, and programs*" pubblicato su *Behavioral and Brain Sciences* nel 1980. Nell'esperimento, si immagina una persona che non comprende il cinese, ma si trova all'interno di una stanza. Questa persona dispone di un libro con istruzioni su come rispondere a messaggi in cinese. Utenti esterni fanno scorrere bigliettini in cinese sotto la porta, i quali vengono decifrati dalla persona nella stanza. Pur producendo risposte corrette, la persona non comprende il cinese; ha semplicemente risposto seguendo le istruzioni del libro. L'esperimento della *Chinese Room* ha alimentato il dibattito sull'equivalenza tra il seguire meccanicamente le regole (come fa un computer) e la reale comprensione, sollevando l'interrogativo se la cognizione umana implichi qualcosa di più della mera manipolazione di simboli.

16 M. Savage, *Sir Paul McCartney says artificial intelligence has enabled a 'final' Beatles song*, in *bbc.com*, 13 giugno 2023.

17 G. Björklund - M. Bohlin - E. Olander - J. Jansson - C.E. Walter - M. Au-Yong-Oliveira, *An Exploratory Study on the Spotify Recommender System*, in A. Rocha - H. Adeli - G. Dzemyda - F. Moreira (a cura di) *Information Systems and Technologies*, Cham, 2022, 366 ss.

18 O. Lopez-Rincon - O. Starostenko - G.A. S.Martín, *Algoritmic music composition based on artificial intelligence: A survey*, in *International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 2018, 187 ss.

19 Per una breve introduzione al tema si consiglia: J. Levine - C.B. Congdon - M. Ebner - G. Kendall - S.M. Lucas - R. Miikkulainen - T. Schaul - T. Thompson, *General Video Game Playing*, in *Dagstuhl Follow-Ups*, 6, 1998, 77 ss.

20 S. Seidel - N. Berente - A. Lindberg - K. Lyytinen - B. Martinez - J. V. Nickerson, *Artifical Intelligence and Video Game Creation: A Framework for the New Logic of Autonomous Design*, in *Journal of Digital Social Research*, 2, 2020, 126 ss.

21 X. Cui - H. Shi, *A*-based Pathfinding in Modern Computer Games*, in *International Journal of Computer Science and Network Security*, 11, 2011, 125 ss.

22 Per ciò che riguarda l'AI nei videogiochi al di fuori dei personaggi non giocanti si consiglia G.N. Yannakakis, *Game AI revisited*, in *Proceedings of the 9th conference on Computing Frontiers*, 2012, 285 ss.

grammi di editing[23] come Grammarly o ProWritingAid, che controllano grammatica e ortografia. Altre forme di IA aiutano nella traduzione di testi, come DeepL o Google Translate.
Persino le automobili moderne sono dotate di IA, ad esempio attraverso sistemi ADAS (*Advanced Driver-Assistance Systems*) che migliorano l'esperienza di guida con funzionalità come la frenata automatica d'emergenza o l'avvertimento di superamento di corsia[24].
Concludiamo quindi esplorando un ulteriore aspetto dell'IA: il *deepfake*, che rappresenta una diversa forma avanzata di intelligenza artificiale.

3. Profondamente falsi

Il termine *deepfake* è composto da "*deep learning*"[25] e "*fake*". Solitamente, quando si parla di *deepfake*, ci si riferisce a: «foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce[26]».
Innanzitutto, è importante sottolineare che le materie di base su cui si fondano i *deepfake* sono costituite dai volti, dai corpi e dalle voci di persone effettivamente esistenti[27].
In pratica, i *deepfake* operano in modo simile ai filtri di alcuni noti social network, come Instagram, o alle applicazioni di riconoscimento facciale, ma con una complessità e un sistema notevolmente più articolati. Attraverso un elaborato sistema di reti neurali, che analizzano enormi quantità di dati tramite il *deep learning*, è possibile replicare in modo estremamente fedele non solo i volti, ma anche le espressioni facciali, la voce e le sue sfumature[28,29]. Di conseguenza, l'intelligenza artificiale utilizza tutti i dati di mappatura facciale raccolti per sovrapporre il viso di una persona a quello presente in

[23] T. N. Fitria, *"Grammarly" as AI-powered English Writing Assistant: Students' Alternative for English Writing*, in *Journal of English Language Literature and Teaching*, 5, 2021, 65 ss.

[24] A. Ziebinski - R. Cupek - D. Grzecha - L. Chruszczyk, *Review of advanced driver assistance systems (ADAS)*, in *AIP Conference Proceedings*, 2017.

[25] Per un approfondimento sul tema, si suggerisce la lettura di "*An Introduction to Deep Learning*" di L. Arnold - S. Rebecchi - S. Chevallier - H. Paugam-Moisy, incluso in "*Statistical Foundations on Data Science*" del 2011. In breve, il *deep learning* rappresenta una forma di intelligenza artificiale volta a consentire ai computer di prendere decisioni in modo autonomo, seguendo un approccio simile a quello del cervello umano.
Per chiarire questo concetto, prendiamo ad esempio la parola "pesce". Tipicamente, un computer richiederebbe vari *input* (ad esempio, informazioni su animali acquatici, squame, branchie, ecc.) per comprendere il concetto di "pesce". Al contrario, il *deep learning* permette ai computer di apprendere a riconoscere un pesce elaborando diverse immagini, consentendo loro di decifrare autonomamente le caratteristiche di un animale acquatico.

[26] Garante Privacy, *Deepfake – Vademecum*, 2020, 1.

[27] *Ibid.*

[28] CBS, *I never said that! High-tech deception of 'deepfake' videos*, in *cbsnews.com*, 2 luglio 2018.

[29] D. Pan - L. Sun - R. Wang - X. Zhang - R. O. Sinnott, *Deepfake Detection through Deep Learning*, in *International Conference on Big Data Computing, Applications and Technologies*, 2020, 134 ss.

un determinato video[30]. Ciò consente la creazione di situazioni mai verificatesi, in cui una persona sembra fare e dire cose che non avrebbe mai fatto o detto[31,32,33]. Ad esempio, come accennato nel capitolo primo (in merito al caso dell'attrice hollywoodiana Scarlett Johansson), è possibile sovrapporre il volto di una famosa attrice su quello di una pornoattrice, creando così un video pornografico coinvolgente attori e attrici falsi. Dal suo emergere nel 2017, quando un utente di Reddit (una famosa piattaforma di notizie e intrattenimento) pubblicò vari video pornografici falsi con protagoniste diverse celebrità[34], il termine *deepfake* è costantemente associato, non a caso, a una connotazione negativa. Tuttavia, come evidenziato nell'introduzione di questo documento, i *deepfake* possono essere impiegati in modi creativi. Un esempio notevole è il *deepfake* realizzato dallo YouTuber Shamook, che ha riprodotto in modo impressionante il volto di un giovane Mark Hamill, sovrapponendolo a una resa CGI poco convincente, regalando ai fan di Star Wars una credibile e ringiovanita interpretazione di Luke Skywalker[35]. Inoltre, uno studio interessante condotto da Gillian Murphy e colleghi[36] ha aperto nuove possibilità stimolanti: *«many participants mentioned that they do not like specific actors, and consequently do not watch their films. Interestingly, the idea of re-casting controversial actors may provide a way for people to feel they are "separating the artist from the art", a challenge that has received renewed attention since the onset of the #MeToo movement*[37]».

Un ulteriore esempio positivo rilevante in questo contesto è rappresentato da uno spot pubblicitario volto a sensibilizzare sull'importanza della lotta contro la malaria, con protagonista l'ex calciatore inglese David Beckham. Mediante l'utilizzo della tecnologia *deepfake*, sono state superate diverse barriere linguistiche. In particolare, non solo è stato possibile ricreare in modo straordinariamente accurato le espressioni facciali del calciatore inglese, ma addirittura i movimenti della bocca sono stati modificati con precisione per adattarsi perfettamente al labiale delle nove lingue interpretate[38].

Nonostante vi siano esempi come la produzione di meme, l'uso creativo nei film e la realizzazione di pubblicità a sfondo sociale possano evidenziare il potenziale positivo della tecnologia *deepfake*, non è possibile ignorare il vero motivo per cui questa tecnologia è particolarmente rinomata, ossia la produzione di notizie false e contenuti video

---

30 A. Chadha - V. Kumar - S. Kashyap - M. Gupta, *Deepfake: An Overview*, in P. K. Singh - S. T. Wierzchoń - S. Tanwar - M. Ganzha - J. J. P. C. Rodrigues (a cura di) *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, Cham, 2021, 557 ss.

31 M. J. Blitz, *Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?*, in *Yale Journal of Law & Technology*, 23, 2020, 162 ss.

32 D. Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, in *Duke Law & Technology Review*, 17, 2019, 99 ss.

33 T. Taulli, *Deepfake: What You Need to Know*, in *forbes.com*, 15 giugno 2019.

34 S. Maddocks, *'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes*, in *Porn Studies*, 7, 2020, 415 ss.

35 La scena in questione è presente nella serie televisiva *The Mandalorian* (*Capitolo 16: Il Salvataggio*, Disney, 2020). Il *deepfake* è reperibile sul sito personale di Shamook in *youtube.com*.

36 G. Murphy - D. Ching - J. Twomey - C. Linehan, *Face/Off: Changing the face of movies with deepfakes*, in *PLoS ONE*, 18, 2023.

37 *Ibid.*

38 J. Marovt, *How we made David Beckham speak 9 languages*, in *synthesia.io*, 4 settembre 2023.

che possono compromettere non solo la reputazione individuale, ma anche la stabilità delle nostre democrazie.

4. *Deepfake*-news

«Nel tempo dell'inganno universale dire la verità è un atto rivoluzionario» - George Orwell.

La parola dell'anno 2016 per l'Oxford Dictionary fu il neologismo "*post-truth*", ovvero "post-verità"[39]. Secondo la definizione del prestigioso dizionario, l'aggettivo indica «*circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief*[40]». Oggi, le *fake news* hanno ampiamente infiltrato la quotidianità di ciascun individuo, catturando l'attenzione di molte persone e influenzando la loro percezione della realtà, come dimostra l'uscita del Regno Unito dall'Unione Europea[41].

L'influenza delle notizie false, specialmente quelle derivanti dai social network, sulla Brexit è ben nota[42]. Alcuni politici sostenitori della Brexit promuovevano un maggiore controllo dell'immigrazione nel Regno Unito. Va ricordato all'uopo che Londra, non solo controllava gran parte dei flussi migratori che riguardavano il paese, ma che le contrattazioni post-Brexit non hanno fatto altro che complicare ulteriormente le cose[43]. Un ulteriore esempio emblematico è la notizia fuorviante dei 350 milioni di sterline settimanali che avrebbero dovuto essere destinati al Servizio Sanitario Nazionale dopo la Brexit[44].

I social network sono stati decisivi nella diffusione di *fake news* sulla Brexit[45]. Uno studio della Ofcom (l'ente per la regolamentazione delle emittenti radiotelevisive e delle telecomunicazioni del Regno Unito), condotto nel 2018, ha evidenziato che internet è la seconda fonte di informazione più utilizzata dagli inglesi, con i giovani e le minoranze etniche che preferiscono informarsi online. In particolare, i social network sono la fonte online prediletta per raccogliere informazioni[46]. Di conseguenza, si può

[39] Oxford University Press, *Word of the Year 2016*, in *global.oup.com*, 2016.

[40] *Ibid.*

[41] Y. Rodny-Gumede, *Fake It till You Make It: The Role, Impact and Consequences of Fake News*, in B. Mutsvairo - B. Karam (a cura di) *Perspectives on Political Communication in Africa*, Palgrave Macmillan, 2018, 203 ss.

[42] D. Dobreva - D. Grinnell - M. Innes, *Prophets and Loss: How "Soft Facts" on Social Media Influenced the Brexit Campaign and Social Reactions to the Murder of Jo Cox*, in *Policy & Internet*, 12, 2020, 144 ss.

[43] J.N. Druckman - E. Peterson - R. Slothuus, *How elite partisan polarization affects public opinion formation*, in *American Political Science Review*, 107(1), 2013, 57 ss.

[44] J. Stone, *British public still believe Vote Leave '350million a week to EU' myth from Brexit referendum*, in *independent.co.uk*, 28 ottobre 2018.

[45] A riguardo si consiglia lo studio di M. Höller, *The human component in social media and fake news: the performance of UK opinion leaders on Twitter during the Brexit campaign*, in *European Journal of English Studies*, 25(1), 2021, 80 ss., che ha analizzato i profili X (allora, Twitter) dei principali *opinion leader* britannici nel periodo pre-Brexit. Dallo studio è emerso l'effettivo contributo nella divulgazione di notizie false da parte dell'ex primo ministro Boris Johnson e di uno dei maggiori fautori della Brexit Nigel Farage.

[46] Jigsaw Research, *News Consumption in the UK: 2018*, in *ofcom.org.uk*, 2018, 1.

affermare con verosimile certezza che le *fake news* abbiano influenzato la percezione della realtà della popolazione inglese, e di conseguenza il loro voto[47,48].

Il problema delle *fake news* online è ulteriormente aggravato dalla crescente presenza di video *deepfake*, creati principalmente da *hacktivist*, vale a dire persone o gruppi che, per motivi sociali, politici o ideologici[49], promuovono campagne di disinformazione per confondere l'opinione pubblica e minare la fiducia in un determinato paese[50,51,52]. Quando i contenuti *deepfake* coinvolgono politici o opinion leader, possono causare variazioni nelle opinioni dell'elettorato, compromettendo i diritti di autodeterminazione informativa e libertà decisionale delle persone. Ne discende la privazione di due indispensabili diritti: (a) l'autodeterminazione informativa («ciò che voglio far sapere di me lo decido io[53]») e (b) la libertà decisionale («quello che penso e faccio è una scelta su cui gli altri non possono interferire[54]»). Ne è un esempio lampante il conflitto in Ucraina. Pochi mesi dopo l'inizio dell'invasione da parte delle armate di Mosca, sono apparsi due video falsi che inscenavano una presunta dichiarazione di pace da parte dei presidenti Volodymr Zelensky e Vladimir Putin[55]. Se il primo è stato reso in modo poco convincente[56], il secondo è decisamente più persuasivo (nonostante le paradossali dichiarazioni[57]).

In questa era della post-verità, emerge, quindi, una crescente sfida nel distinguere ciò che è veritiero da ciò che è falso. Il problema è ulteriormente complicato dal fatto che spesso sono proprio i politici a diffondere notizie false allo scopo di screditare i loro avversari. A solo titolo esemplificativo si rammenta il candidato alle primarie del Partito Repubblicano, Ron DeSantis, il quale ha condiviso un'immagine *deepfake* di Donald Trump apparentemente impegnato in un caloroso abbraccio e bacio con il Dottor

---

47 A. Hern, *Facebook criticised for response to questions on Russia and Brexit*, in *theguardian.com*, 13 dicembre 2017.

48 F. Safieddine, *Political and Social Impact of Digital Fake News in an Era of Social Media*, in Y. Ibrahim - F. Safieddine (a cura di) *Fake News in an Era of Social Media: Tracking Viral Contagion*, Lanham 2020, 43 ss.

49 R. Kraus - B. Barber - M. Borkin - N.J. Alpern, *Internet Information Services*, in R. Kraus - B. Barber - M. Borkin-N.J. Alpern (a cura di), *Seven Deadliest Microsoft Attacks*, Syngress, 109 ss.

50 CBS, *I never said that! High-tech deception of 'deepfake' videos*, in *cbsnews.com,* 2 luglio 2018.

51 J. Twomey - D. Ching - M. P. Aylett - M. Quayle - C. Linehan - G. Murphy, *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine*, in *Plos ONE*, 18, 2023.

52 C. Vaccari - A. Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, in *Social Media + Society*, 6, 2020.

53 Garante Privacy, *Deepfake – Vademecum*, cit., 4.

54 *Ibid.*

55 J. Twomey – D. Ching – M. Peter Aylett – M. Quayle – C. Linehan – G. Murphyal., *Do deepfake videos undermine our epistemic trust?*, 2023.

56 J. Wakefield, *Deepfake presidents used in Russia-Ukraine war*, in *bbc.com,* 18 marzo 2022. Il video è risultato scarsamente efficace data la sua pessima realizzazione: il collo del presidente ucraino risulta essere troppo sottile rispetto alla testa e la voce è decisamente più profonda di quella nella realtà.

57 A. Smith, *Fake video of Vladimir Putin declaring peace with Ukraine seeks to cause chaos online*, in *independent.co.uk*, 18 marzo 2022. Nel filmato il presidente russo non si limiterebbe a dichiarare la pace con lo stato ucraino, ma anche di restituire ufficialmente la Crimea all'Ucraina.

Anthony Fauci[58]. Questo episodio ha suscitato ulteriore tensione, considerando l'ostilità che il Partito Repubblicano nutre nei confronti del presidente della NIAID, specialmente in relazione alle misure proposte durante la pandemia di COVID-19. Di conseguenza, molti elettori repubblicani si sono sentiti traditi dall'ex presidente.

A causa di simili eventi, nel 2019, gli stati della California e del Texas hanno reagito vietando la diffusione di video *deepfake* nei 60 giorni che precedono un'elezione. La presente iniziativa evidenzia che alcuni stati, seppur tra loro politicamente diversi, sono ugualmente preoccupati per la crescente diffusione dei *deepfake* e riflette la consapevolezza dei problemi che questa tecnologia può causare[59,60].

I video *deepfake*, come per gli esempi menzionati, sono destinati a suscitare scalpore e confusione tra la popolazione, portando a inevitabili interferenze nel sistema di voto e, più in generale, nel sistema democratico[61]. Le piattaforme social hanno cominciato ad adottare misure per contrastare la diffusione di notizie false. Durante la pandemia di COVID-19 negli Stati Uniti, la piattaforma YouTube ha annunciato il blocco di diversi canali che diffondevano disinformazione sulla campagna vaccinale[62]. Gli algoritmi del social network Instagram, d'altra parte, non mostrano nei contenuti suggeriti quelli contrassegnati come falsi dai fact checker[63,64]. Anche Facebook, altra piattaforma del gruppo Meta, considerata una degli epicentri delle fake news online[65], sta cercando di migliorare la sua reputazione bloccando diversi utenti diffusori di notizie false sui social di Meta[66,67].

---

[58] N. Nehamas, *DeSantis Campaign Uses Apparently Fake Images to Attack Trump on Twitter*, in *nytimes.com*, 8 giugno 2023.

[59] California Government Code, Section 11547.5.

[60] State of Texas, Chapter 21, Penal Code, Section 21.165.

[61] M. Westerlund, *The Emergence of Deepfake Technology: A Review*, in *Technology Innovation Management Review,* 9(1), 2019, 39 ss.

[62] T. Ginossar, *Cross-platform spread: vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations*, in *Human Vaccines & Immunotherapeutics*, 18, 2022.

[63] R. Metz - D. O'Sullivan, *A deepfake video of Mark Zuckerberg presents a new challenge for Facebook*, in *edition.cnn.com*, 11 giugno 2019.

[64] N. Veerasamy - H. Pieterse, *Rising Above Misinformation and Deepfakes*, in R. P. Griffin - U. Tatar - B. Yankson (a cura di), *ICCWS 2022 17th International Conference on Cyber Warfare and Security*, 2022, ACI, 340 ss.

[65] Stando a recenti studi Facebook è il social network che diffonde il maggior numero di notizie false. In particolare, si consigliano due studi: A.M. Guess - B. Nyhan - J. Reifler, *Exposure to untrustworthy websites in the 2016 US Election*, in *Nature Human Behaviour*, 4, 2020, 472 ss. Un'interessante ricerca sulla diffusione di notizie online durante le elezioni presidenziali americane del 2016. T. Hopp-P. Ferrucci - C.J. Vargo, *Why Do People Share Ideologically Extreme, False, and Misleading Content on Social Media? A Self-Report and Trace Data–Based Analysis of Countermedia Content Dissemination on Facebook and Twitter*, in *Human Communication Research*, 46, 2020, 357 ss. Uno studio relativamente recente sui livelli individuali di radicalismo ideologico e la fiducia nel sistema mediatico in contrapposizione a quello delle fake news.

[66] J. Pamment, *How the Kremlin circumvented EU sanctions on Russian state media in the first weeks on the illegal invasion of Ukraine*, in *Place Branding and Public Diplomacy*, 19, 2023, 200 ss.

[67] Nondimeno ci sono piattaforme social che non attuano nessuna contromisura – se non addirittura endorsano – gli *account* di controinformazione. X (precedentemente noto come Twitter) di Elon Musk, ad esempio, è da parecchio tempo sotto scrutinio da parte dell'Unione Europea a causa dell'elevato numero di account (tra i quali molti sono verificati) che continuano a riempire il sito di notizie false. Molte delle fake news diffuse sul social di Musk riguardano temi scottanti e di importanza mondiale come il

Tuttavia, l'impatto dei *deepfake* non si limita alle strutture democratiche, elettorali e sociali, estendendosi altresì al campo dei media, interessando quindi i giornalisti stessi. Questi, devono affrontare una crescente diffidenza del pubblico nei confronti delle informazioni pubblicate sulla stampa[68], mentre contemporaneamente si trovano ad avere a che fare con un proliferare di fonti false. Giova menzionare, a titolo esemplificativo ed esplicativo, che durante le tensioni tra Pakistan e India, l'agenzia di stampa Reuters ha individuato oltre 30 video manipolati pubblicati da organi di informazione[69]. Con la Con la crescente ubiquità dei video *deepfake*, è inevitabile che questa problematica si aggraverà. Per fronteggiare questa sfida, agenzie di stampa di rilevanza globale, tra cui Reuters[70] e il Wall Street Journal, hanno implementato corsi formativi dedicati ai propri giornalisti per migliorare la loro capacità di riconoscere e gestire eventuali fake news. Inoltre, le stesse agenzie stanno investendo in diverse tecnologie volte al monitoraggio e al riconoscimento dei contenuti *deepfake*[71].
Va notato che nell'ordinamento italiano la creazione di *deepfake* è, in alcuni casi, considerata un reato, e nel paragrafo successivo si affronteranno le implicazioni legate al furto d'identità e all'uso di dati personali per la creazione di video falsi, in particolare nel contesto del mondo della pornografia e della pedopornografia[72].

***Deepfake* e furto d'identità**

Il connubio tra il fenomeno dei *deepfake* e il furto d'identità si presenta come un lega-

conflitto in Ucraina. M. Haigh - T. Haigh, *Fighting and Framing Fake News*, in P. Baines - N. O'Shaughnessy - N. Snow (a cura di) *The SAGE Handbook of Propaganda*, 2020, 303 ss.; P. Suciu, *X Is The Biggest Source Of Fake News And Disinformation, EU Warns,* in *forbes.com,* 26 settembre 2023 e quello tra Israele e Palestina M. Lakhani, *Fighting Disinformation in the Palestine Conflict: The Role of Generative AI and Islamic Values*, in *Al-Misbah Research Journal*, 3(6), 2023; D. Milmo, *X criticised for enabling spread of Israel-Hamas disinformation*, in *theguardian.com*, 9 ottobre 2023.

68 M. Westerlund, *The Emergence of Deepfake Technology,* cit., 40.

69 N. Jaffer, *Fake News and Disinformation in Modern Statecraft*, in *Regional Studies*, 39(1), 2021, 3 ss.

70 R. Lauren, *Will you believe it when you see it? How and why the press should prepare for deepfakes*, in *Georgetown Law Technology Review*, 4(1), 2019, 241 ss.

71 Á. Vizioso - M. Vaz-Álvarez - X. López-García, *Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation*, in *Media and Communication*, 9(1), 2021, 291 ss.

72 Va ricordato che i reati legati al mondo del deepfake non comprendono solo il furto d'identità, la pornografia e la pedopornografia, bensì possono comprendere anche crimini riguardanti l'estorsione e il bullismo. A riguardo è necessario citare almeno tre tecnologie con le quali *deepfaker* e altri criminali di internet possono ingannare gli utenti: (a) il *phishing* è una nota tecnica estorsiva con la quale dei malintenzionati convincono una o più vittime a condividere dati personali e/o finanziari (Z. Ramzan, *Phishing Attacks and Countermeasures*, in P. Stavroulakis - M. Stamp (a cura di), *Handbook of Information and Communication Security*, Berlino, 2010); (b) lo *spoofing*, invece, è il furto di informazioni attraverso la falsificazione di dati personali e dispositivi (S.A. Schuckers, *Spoofing and Anti-Spoofing Measures*, in *Information Security Technical Report*, 7(4), 2002, 56 ss.) (c) infine, il *ransomware* è un malware che impedisce o limita l'accesso a diversi tipi di dispositivi, attraverso il completo blocco dei sistemi o alla cifratura di file. Solitamente per sbloccare un qualsiasi dispositivo i malintenzionati chiedono un riscatto in denaro o in beni, servizi e/o favori (X. Luo - Q. Liao, *Awareness Education as the Key to Ransomware Prevention*, in *Information Systems Security*, 16(4), 2007, 195 ss.).

me inequivocabile[73]. Risulta sufficiente riflettere sul fatto basilare che la creazione di qualsiasi video *deepfake* implica l'artificiale "deprivazione" di un individuo del proprio volto, seguita dalla sovrapposizione dello stesso a quello di un'altra persona. Tale procedura costituisce già di per sé una chiara minaccia alla privacy. Se a ciò si aggiunge il fatto che i *deepfake* possono generare contesti e situazioni mai verificatisi nella realtà, l'entità del problema diviene ulteriormente manifesta.

Il nucleo che ogni individuo porta in sé e che, seppur lo riconduce al genere umano, alla tipologia di soggetto o di persona, lo caratterizza più specificatamente, è la sua identità. Ancora più penetrante, quindi, è la necessaria tutela dell'identità personale, definibile come dignità assoluta. In questo contesto, la tutela dell'identità personale emerge come una necessità impellente.

Sotto questo profilo, le Carte internazionali, pur adottando formulazioni diverse, riconoscono a ogni individuo la personalità giuridica e vietano interferenze nella sua vita privata, familiare e sociale, sancendo il diritto allo sviluppo della propria personalità[74].

Parte della dottrina ha sostenuto che l'identità rappresenta un bene-giuridico intrinsecamente legato alla dimensione sociale dell'individuo a cui essa si riferisce. Questo perché si ritiene che «l'identità sia il risultato delle interazioni sociali, che ha le sue radici nello spazio privato ma si manifesta e si definisce successivamente nello spazio pubblico»[75]. Proprio in virtù di questo principio, è celebre la definizione di identità personale contenuta nella pronuncia della Corte Costituzionale del 1994, ove si fa riferimento al «diritto ad essere sé stesso, inteso come rispetto dell'immagine di partecipe alla vita associata, con le acquisizioni di idee ed esperienze, con le convinzioni ideologiche, religiose, morali e sociali che differenziano, ed al tempo stesso qualificano, l'individuo».[76] L'identità è quindi un bene tutelato che, non solo individua il soggetto a cui si riferisce, ma lo rappresenta e ne afferma la sua personalità come proiezione sociale.

Ecco, quindi, che viene disciplinato il reato di furto d'identità[77], regolato dall'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[78]».

Il furto d'identità assume una gravità particolare quando si collega al contesto sessuale delle persone coinvolte nell' illecito. Sin dall'introduzione del fenomeno nel dibattito pubblico, i *deepfake*, essendo l'origine del termine strettamente legata alla pubblicazione

[73] Garante Privacy, *Deepfake – Vademecum*, cit., 1.

[74] In tal senso artt. 6, 12 e 22 della Dichiarazione universale dei diritti dell'Uomo; art. 8 della Convenzione europea dei diritti dell'uomo; artt. 6, 17 e 22 del Patto internazionale dei diritti civili e politici; art. 6 del Trattato dell'Unione europea che riconosce i diritti della CEDU e quelli della Carta dei diritti fondamentali dell'Unione europea.

[75] V. Zeno-Zencovich, *Identità personale, Digesto, privato, sezione civile*, IX, Torino, 1993, 301.

[76] Corte cost. 3 febbraio 1994, n. 13.

[77] Il furto d'identità non è il solo reato perpetrabile creando, usando o condividendo un *deepfake*. Ad esempio, se il contenuto *deepfake* va a ledere anche la reputazione dell'individuo, al reato di furto d'identità si aggiunge quello di diffamazione (art. 595 c.p.).

[78] Art. 496 c.p.

di video falsi a sfondo pornografico, hanno suscitato preoccupazioni riguardo a come prevenire e affrontare il problema del furto d'identità. La produzione di immagini fittizie con connotazioni sessuali, spesso ritraenti celebrità, in pose o situazioni intime, è comunemente denominata "*deepnude*". I *deepnude* coinvolgono l'incorporazione di un volto, attraverso l'uso di software, su corpi nudi o in ambienti a carattere pornografico. La situazione è ulteriormente complicata dalla disponibilità sempre crescente di applicazioni[79] che consentono a chiunque di creare facilmente video o immagini.

La semplicità con cui tali applicazioni possono essere installate sui dispositivi mobili e la loro facilità d'uso sono stati posti in rilievo nel corso di recente caso di cronaca che coinvolge un gruppo di studenti di una scuola media di Latina. Utilizzando un'applicazione denominata "*BikiniOff*"[80], i minori hanno posto in essere la manipolazione di fotografie di cinque studentesse e di una docente. Quest'applicazione è particolarmente apprezzata nel mondo dei *deepfaker* poiché va oltre la mera sostituzione del viso, ricreando in modo sorprendentemente realistico la posa desiderata, mantenendo le proporzioni e il colore della pelle della vittima[81]. Il *deepnude* della docente, nel caso specifico, è risultato così convincente da comparire su due rinomati siti pornografici[82].

Inevitabilmente, le discussioni su *deepnude* e *deepfake porn* sono strettamente legate al fenomeno del *revenge porn*, definito come l'«abuso digitale che comporta la divulgazione a terzi o la pubblicazione senza consenso di immagini o video a sfondo sessuale[83]».

In relazione a ciò si rammenta il caso della poetessa inglese Helen Mort, vittima di molteplici casi di *deepfake porn*: «*an acquaintance told her. But never in her life had she taken or shared intimate photos. Surely there must be some mistake. When she finally mustered up the courage to look, she felt frightened and humiliated*[84]».

Il creatore dei *deepfake* aveva pubblicato diverse fotografie della poetessa, incoraggiando altri individui a integrare il volto di Helen in contesti di natura esplicita[85]. La lotta alla diffusione di contenuti pornografici generati mediante l'impiego dell'intelligenza artificiale sembra presentare notevoli sfide. Filoni di ricerca recenti, infatti, evidenziano che oltre il 90% dei video *deepfake* si configurano come materiale a contenuto pornografico[86,87,88]. È da sottolineare, tuttavia, che in maniera analoga a quanto si riscontra sulle varie piattaforme social per quanto concerne la diffusione di *fake news*, il sito web

[79] Basta digitare la parola "*deepnude*" in un qualsiasi motore di ricerca per venire bombardati dalle più diverse app gratuite e a pagamento.

[80] Una delle applicazioni a pagamento più famose per la creazione di *deepnude*.

[81] D. Barbera, *Tutti i rischi di usare Bikinioff, il chatbot che spoglia le donne*, in *wired.it,* 19 aprile 2023.

[82] S. Matteis, *Cinque 13enni e una prof di Latina nude sul web: indagati i compagni, le foto false create con l'app BikiniOff*, in *fanpage.it*, 14 settembre 2023.

[83] L. Sartarelli, *È reato usare un deepfake?*, in *smartius.it*, 11 giugno 2019.

[84] K. Hao, *Deepfake porn is ruining women's lives. Now the law may finally ban it*, in technologyreview.com, 12 febbraio 2021.

[85] J. Laffier - A. Rehman, *Deepfakes and Harm to Women, in Digital Life and Learning*, 3(1), 2023.

[86] N. Kshetri, *The Economics of Deepfakes*, in *Computing's Economics*, 2023, 89 ss.

[87] G. MacGregor, *Gun to your head: how deepfakes and other non-consensual synthetic media hold individual autonomy hostage*, in *UMKC Law Review*, 90(2), 2021, 431 ss.

[88] Sensity Team, *How to Detect a Deepfake Online: Image Forensics and Analysis of Deepfake Videos*, in *sensity.ai,* 2021.

a carattere pornografico più frequentato a livello globale, ha vietato la pubblicazione di video *deepfake*[89,90].

Nonostante ciò, si ritiene estremamente improbabile che non esistano tracce di contenuti compromettenti generati tramite applicazioni di intelligenza artificiale su questa celebre piattaforma riservata ai maggiorenni. Ciò è in parte attribuibile alla legislazione statunitense, nello specifico la "Section 230[91]", approvata dal Congresso nel 1996, che garantisce una esclusione di responsabilità dei siti web rispetto ai contenuti pubblicati dai singoli utenti che siano in violazione con le regole imposte dalla normativa o dallo stesso sito web, come sottolineato da Dean Russell: «*[Section 230] prevents websites from being sued for hosting non-consensual deepfake porn, which means Google, Reddit, Twitter, etc. — you can ask them to take down a deepfake, but they don't have to. Pornhub says it doesn't allow deepfakes. But they are there. And then there are websites predicated on deepfake porn. That's their whole business model. Again, they are protected by Section 230*»[92].

Nonostante l'incremento della minaccia derivante dai *deepfake* a sfondo pornografico, nella totalità degli Stati Uniti sono poche le giurisdizioni che hanno adottato specifiche contromisure per affrontare questo fenomeno. Ad esempio, nello stato della California, le vittime possono ricorrere alla corte civile per ottenere un risarcimento finanziario[93]. In Virginia, invece, una legge introdotta nel 2019 impone pene detentive a coloro che diffondono *deepfake* pornografici. Tuttavia, la pubblicazione o la condivisione di tali contenuti è considerata reato solo se finalizzata a molestare o intimidire la persona rappresentata nel video[94].

6. La regolamentazione in Italia

In seguito agli eventi verificatisi a Latina sopra citati, la Procura dei Minori di Roma ha avviato un'inchiesta sulle finte foto pornografiche realizzate dagli alunni della scuola media. Sempre sulla scorta di quanto successo al gruppo di ragazzini, il 13 ottobre 2020, il Garante per la protezione dei dati personali ha avviato un'istruttoria nei confronti di Telegram[95]. Tale iniziativa ha scaturito un ulteriore intervento specifico in materia, durante il quale il Garante ha formulato un *vademecum* nominato: "*Deepfake*. Il falso che ti "ruba" la faccia (e la privacy)"[96]. L'istituzione ha mantenuto un costante interesse sull'argomento, emettendo provvedimenti, documenti ufficiali e comunicati.

[89] M. Popova, *Reading out of context: pornographic deepfakes, celebrity and intimacy*, in *Porn Studies*, 7, 2020, 367 ss.

[90] R. Winter, *DeepFakes: uncovering hardcore open source on GitHub*, in *Porn Studies*, 7, 2020, 382 ss.

[91] 47 U.S. Code, Section 230 on Protection for private blocking and screening of offensive material.

[92] D. Russell, *Is deepfake pornography illegal? It depends*, in *wbur.org*, 23 giugno, 2023.

[93] K. Farish, *Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake*, in *Journal of Intellectual Property Law & Practice*, 15, 2020, 40 ss.

[94] E. Gerstner, *Face/off: "deepfake" face swaps and privacy laws*, in *Defense Counsel Journal*, 87, 2020, 1 ss. Per quanto riguarda la legge in questione si tratta della: VA. CODE. § 18.2-386.2.

[95] Garante Privacy, *Deep fake: il Garante privacy apre un'istruttoria nei confronti di Telegram per il software che "spoglia" le donne*, 23 ottobre 2020, doc. web n. 9470722.

[96] *Ibid.*

Infatti, l'Autorità Garante ha continuato ad affrontare il tema della lesione del diritto all'identità personale in ottica divulgativa proprio per il costante emergere di molteplici nuove tecnologie, pericolosamente rischiose per il patrimonio personale. [97]

Date le ampie implicazioni dell'intelligenza artificiale sulla società, si osserva un crescente riconoscimento della necessità di una regolamentazione adeguata in questo settore, con numerosi paesi che collaborano attivamente per sviluppare normative adeguate. La regolamentazione si sforza costantemente di tenere il passo con lo sviluppo e l'innovazione tecnologica, ma la sua implementazione varia significativamente tra le diverse giurisdizioni. Questo equilibrio complesso implica che una disciplina eccessivamente restrittiva potrebbe ostacolare l'innovazione, mentre una regolamentazione troppo lenta potrebbe lasciare un vuoto normativo pericoloso con rischi non gestiti.

L'IA è in continua evoluzione e il suo sviluppo non aspetterà l'entrata in vigore delle nuove regole. Questo porta a due importanti conseguenze. In primis, quando le nuove regole saranno stabilite, potrebbero non essere in grado di regolamentare completamente o adeguatamente le applicazioni di intelligenza artificiale che non erano ancora emerse o che non sembravano richiedere inizialmente una disciplina ad hoc.

In secondo luogo, vi è l'urgenza di essere rapidi e di trovare un modo per regolare l'intelligenza artificiale il prima possibile. Urgenza dovuta al fatto che i progressi nell'IA stanno avvenendo a un ritmo accelerato e i governi non possono rischiare di assistere nuovamente a una sorta di autoregolamentazione, così come avvenuto per Internet, che, in assenza di una normativa, ha condotto tutti a una dimensione digitale che può definirsi auto-gestita, in una sorta di tecnocrazia che rischia di travolgere le democrazie attuali.

Il Consiglio dell'Unione Europea ha affrontato la questione adottando un progetto di promozione dell'innovazione responsabile nell'intelligenza artificiale per la pace e la sicurezza. In tale contesto, il Consiglio ha riconosciuto l'IA come una tecnologia abilitante con ampio potenziale d'uso generale, ma ha altresì sottolineato il rischio di abusi derivanti dall'accesso indiscriminato alle ricerche e all'innovazione in materia di IA per applicazioni civili. Nella decisione di Promozione dell'innovazione responsabile nell'IA per la pace e la sicurezza, in vigore dal 18 novembre 2022, si evidenzia il rischio rappresentato dall'uso improprio della tecnologia di IA, specialmente attraverso reti generative avversarie (GAN), per la creazione di *deepfake* utilizzati in campagne di disinformazione, un rischio che richiede particolare attenzione e che potrebbe non essere sufficientemente affrontato dagli attuali sforzi diplomatici e di controllo degli armamenti.

Senza volersi addentrare in questa sede nell'analisi specifica del documento di progetto, basti pensare che il Consiglio si è posto come obiettivo quello di promuovere: «un'innovazione responsabile, in quanto meccanismo di autogoverno, potrebbe fornire alla comunità civile mondiale dell'AI strumenti e metodi pratici per individuare e contribuire a prevenire e attenuare i rischi che la diversione e l'uso improprio della ricerca e dell'innovazione civili in materia di IA potrebbero comportare per la pace e

[97] Garante Privacy, *Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*, 28 dicembre 2020.

la sicurezza[98]».

Invece, per quanto concerne l'ordinamento italiano, l'atto di pubblicare immagini o video sessualmente espliciti senza il consenso delle persone coinvolte è divenuta una condotta delittuosa con l'introduzione nel Codice penale dell'art. 10 della legge 69/2019 (c.d. Codice Rosso), che è stato pubblicato in Gazzetta Ufficiale il 25 luglio 2019. Questa legge è stata introdotta per punire penalmente il fenomeno noto come "*revenge porn*" attraverso una normativa *ad hoc* per affrontare il problema. Oggetto del reato sono le immagini o i video a contenuto sessualmente esplicito, destinati a rimanere privati.

Secondo quanto stabilito dalla Suprema Corte, il delitto è istantaneo e si consuma nel momento in cui avviene il primo invio dei contenuti, senza che assuma rilevanza il fatto che il destinatario sia un familiare della vittima, che non abbia interesse ad alimentare una successiva diffusione delle immagini. Viene dunque punita la «diffusione illecita di contenuti sessualmente espliciti [che possono avere come] oggetto immagini o video che ritraggano atti sessuali ovvero organi genitali ovvero anche altre parti erogene del corpo umano, come i seni o i glutei, nudi o in condizioni e contesto tali da evocare sessualità[99]».

Vi è inoltre una proposta di legge, presentata il 30 marzo 2021[100], volta a contrastare «il fenomeno della diffusione del *software* chiamato "*Deep nude*"[101]» e il *revenge porn*. L'obiettivo è quello di introdurre nel codice una «fattispecie delittuosa consistente nella diffusione di immagini di persone reali manipolate artificialmente allo scopo di ottenerne rappresentazioni nude[102]». L'art. 612-*quater* (primo comma) c.p. prevedrebbe un reato perpetrabile da chiunque abbia avuto a che fare, non solo con la creazione del *deepfake*, ma anche con la diffusione, l'invio o la pubblicazione di «immagini di persone reali, comunque identificabili, manipolate artificialmente mediante l'uso di strumenti tecnologici o di sistemi di intelligenza artificiale[103]». Inoltre, il secondo comma dell'art. 612-*quater*, prevede l'aumento della pena nel caso la persona che abbia condiviso e/o creato il video sia legata (o sia stata legata) da una relazione affettiva con l'individuo presente nei contenuti *deepfake*[104].

Occorre peraltro menzionare quanto già ricordato dal Servizio Studi del Dipartimento Giustizia nella proposta di legge su citata, vale a dire il fatto che il diritto della persona alla propria immagine è già regolato negli artt. 96 e 97, l. 22 aprile 1941, n. 633 (c. d. legge sulla protezione del diritto d'autore), vietando a chiunque di «esporre o pubbli-

98 Consiglio – Decisione 18/11/2022, n. 2022/2269.

99 Cass. pen., sez. V, 7 aprile 2023, n. 14927.

100 Art. 612-*quater* c.p.

101 Camera dei deputati, *Introduzione dell'articolo 612-quater del codice penale, in materia di manipolazione artificiale di immagini di persone reali allo scopo di ottenerne rappresentazioni nude,* 21 luglio 2021, 1.

102 *Ibid.*

103 *Ibid.*

104 Si può trattare del coniuge (indipendentemente siano essi separati o divorziati) o da persone legate da relazioni affettive passate.

care il ritratto altrui […] senza il consenso dell'interessato»[105,106]. Inoltre, ci si riferisce anche alle rappresentazioni fittizie che vanno a riprendere le sembianze di una persona[107] ed è qui che entrano in gioco i *deepfake porn*. La preoccupazione, infatti, è che i filmati e le foto vengano usate per motivi di ricatto o di revenge porn, data anche la semplicità d'uso che potenzialmente può far sì che ogni persona che ha condiviso una foto online diventi vittima[108].

Dato il realismo delle immagini e dei filmati che vengono creati con le nuove tecnologie, è indubbio che l'onore e la reputazione delle persone presenti in contenuti *deepfake* porn vengano lese. «L'interessato può rivolgersi all'autorità giudiziaria per ottenere la cessazione del comportamento abusivo, il risarcimento del danno ed eventualmente la pubblicazione della sentenza di condanna[109]»[110]. Infine, la vittima ha diritto a richiedere il risarcimento per il danno causato alla sua immagine («pregiudizio economico che la vittima abbia risentito dalla pubblicazione e di cui abbia fornito la prova[111]»). Tuttavia, nel caso non persista il reato di diffamazione, il colpevole rimane obbligato a risarcire la vittima del danno non patrimoniale.

7. Conclusioni

Sin dalla comparsa dei primi *deepfake* nell'anno 2017, questa particolare forma di tecnologia ha manifestato una costante evoluzione, procedendo a un ritmo straordinariamente accelerato, costantemente anticipando coloro che cercano di mitigarne la diffusione e il miglioramento[112]. Secondo Maras e Alexandrou[113], si prospetta un futuro in cui sempre più video artificialmente modificati con l'ausilio dell'IA saranno utilizzati per orchestrare campagne di propaganda terroristica, diffondere notizie politiche false, ricattare individui e per il cyberbullismo.

---

105 Camera dei deputati, *Introduzione dell'articolo 612-quater del codice penale*, 21 luglio, 2021, 2.

106 È necessario ricordare che persistono alcune eccezioni per quanto riguarda la condivisione di immagini senza consenso. «Non occorre il consenso della persona ritratta quando la riproduzione dell'immagine è giustificata dalla notorietà […], da necessità di giustizia o di polizia, da scopi scientifici, didattici o culturali». Tuttavia, «il ritratto non può […] essere esposto o messo in commercio, quando l'esposizione o messa in commercio rechi pregiudizio all'onore, alla reputazione od anche al decoro della persona ritratta» (Art. 97, l 22 aprile 1941, n. 633, c.d. "Legge sul diritto d'autore").

107 Camera dei deputati, *Introduzione dell'articolo 612-quater del codice penale*, cit., 2.

108 Ivi, 1.

109 Ivi., 2.

110 Accanto alla persona lesa, sono autorizzati a sporgere denuncia anche i genitori o i figli.

111 Camera dei deputati, *Introduzione dell'articolo 612-quater del codice penale*, cit., 2.

112 Va comunque detto che la tecnologia *anti-deepfake* sta facendo anch'essa passi da gigante. Negli ultimi anni, infatti, è stata sviluppata una tecnologia che individua i cambiamenti nel volto attraverso l'analisi dei flussi biometrici del sangue (U. A. Çiftçi - I. Demir - L. Yin, *Deepfake source detection in a heartbeat*, in *The Visual Computer*, 2023). Altri ricercatori hanno dimostrato che i video *deepfake* (meno recenti) non riuscivano a simulare in modo perfetto la velocità della chiusura delle palpebre (T. Jung - S. Kim - K. Kim, *DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern*, in *Institute of Electrical and Electronics Engineers*, 2020).

113 M.H. Maras - A. Alexandrou, *Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos*, in *The Internet Journal of Evidence & Proof*, 23, 2018.

Come emerge dall'analisi sin qui svolta, il fenomeno di immagini e video falsi condivisi in rete, non solo non mostra segni di attenuazione, ma sembra destinato a proliferare ulteriormente. Ciò implica che i *deepfake* costituiranno una problematica continua e difficile da affrontare per le democrazie e il sistema d'informazione. Nonostante il ritardo nella definizione di misure adeguate, diversi Stati stanno ora adottando contromisure atte a contenere almeno parzialmente la diffusione di notizie false riportati sui siti web, nei forum e sui social media.

I social media stessi, seppure non tutti agiscano in modo coerente in questa direzione, stanno compiendo sforzi significativi per contrastare la diffusione di notizie false e di *deepfake* che, oltre ad integrare fattispecie delittuose, potrebbero minare il sano processo democratico di un paese.

Richiamando la necessità di normative in questo ambito, è essenziale educare non solo i giornalisti, ma anche il pubblico generale, riguardo ai rischi connessi ai *deepfake*[114]. È però possibile mitigare questo rischio seguendo alcune precauzioni e rispettando semplici norme.

La consapevolezza dell'esistenza di questa tecnologia e delle sue potenzialità aiuta a individuare i segnali e riconoscere i contenuti virtualmente manipolati[115]. Investire nell'educazione digitale è essenziale per prevenire e proteggersi dalla minaccia dei *deepfake*[116]. Una maggiore consapevolezza e conoscenza del mondo online possono rendere il nostro discernimento più attento agli errori grossolani presenti nei video *deepfake*. Gruppi di medici forensi esperti di media consigliano di prestare attenzione a dettagli come oscillazioni del volto, strani riflessi negli occhi, movimenti non congruenti della bocca e sfocature nei contorni della figura[117].

È di tutta evidenza che il *deepfake* impatta in maniera diretta sulla sfera privata, poiché ogni individuo che abbia condiviso anche una sola foto online, potrebbe diventare oggetto della creazione di video *deepfake* senza il proprio consenso. Una misura di prevenzione fondamentale è quindi rappresentata da una maggiore cautela nei comportamenti online. Evitare di condividere indiscriminatamente aspetti della vita privata e monitorare attentamente le immagini online per verificarne un uso non autorizzato rappresentano un primo passo importante[118]. L'utilizzo di password complesse e l'aggiornamento regolare delle stesse contribuiscono a preservare dati personali e bancari[119]. L'adozione di tecnologie come le VPN[120] offre ulteriore protezione contro gli hacker, rendendo più complesso il tracciamento dell'attività online.

---

[114] Stando a dei dati raccolti da iProov (una rinomata azienda di prodotti per l'autenticazione biometrica) nel 2022 su un campione di 16.000 persone in otto paesi (Italia, Spagna, Stati Uniti, Canada, Messico, Germania, Regno Unito e Australia), solamente il 29% sa cos'è un video deepfake (iProov, *How To Protect Against Deepfakes – Statistics and Solutions*, in *iproov.com*, 2022).

[115] L. Sartarelli, *È reato usare un deepfake?*, in *smartius.it,* 31 maggio 2023.

[116] A riguardo si consiglia: Y. Mirsky - L. Wenke, *The Creation and Detection of Deepfakes: A Survey*, in *Association for Computing Machinery*, 54, 2022.

[117] M. Westerlund, *The Emergence of Deepfake Technology*, cit.

[118] L. Sartarelli, *È reato usare un deepfake?*, cit.

[119] J. Wojewidka, *The deepfake threat to face biometrics*, in *Biometric Technology Today*, 2, 2021.

[120] B. Timmerman - P. Mehta - P. Deb - K. Gallagher - B. Dolan – Gavitt - S. Garg - R. Greenstadt, *Studying the Online Deepfake Community*, in *Journal of Online Trust & Safety*, 2, 2023.

In conclusione, in un mondo in cui la tecnologia è dominante e la distinzione tra verità e falsità diventa sempre più difficile, è inevitabile riflettere sulla vulnerabilità della verità e sulla fragilità dell'identità personale. Nonostante il dinamico sviluppo dei *deepfake*, è essenziale non perdere la fiducia nel progresso tecnologico e nella realtà digitale. Solo attraverso un equilibrio tra la maggiore educazione digitale e l'utilizzo di tecnologie avanzate è possibile rendere più sicura la navigazione online, preservando così l'identità personale, l'individualità e la verità nel mondo digitale, attualmente preponderante.