UNIVERSITÀ DEL PIEMONTE ORIENTALE

DIPARTIMENTO DI GIURISPRUDENZA E SCIENZE POLITICHE, ECONOMICHE E SOCIALI

CORSO DI LAUREA IN GIURISPRUDENZA LM/01

TESI DI LAUREA

La funzione e i risvolti delle intelligenze artificiali e dei Deepfake nel condizionamento della volontà e la falsa rappresentazione della realtà

Relatore:

Chiar.ma Prof. ssa Bianca Gardella Tedeschi

Candidato:

Pietro Rodini

ANNO ACCADEMICO 2024/2025

Al mio papà e alla mia mamma;

coloro che nella loro immensa pazienza

mi hanno visto ridere, piangere e crescere

coloro che hanno creduto in me dall'inizio

e senza i quali i miei sogni sarebbero rimasti tali

INDICE

1 Introduzione

Negli ultimi decenni, il mondo ha assistito a un'accelerazione straordinaria nello sviluppo tecnologico, un progresso che ha trasformato profondamente il modo in cui pensiamo, comunichiamo e agiamo. Tra le innovazioni più impattanti, troviamo l'intelligenza artificiale e le tecnologie derivate, come i Deepfake; queste sollevano numerose questioni sulla manipolazione della volontà umana e sulla costruzione di realtà artificiali.

La volontà, che guida il comportamento umano a livello individuale e collettivo, è una forza influenzabile, soggetta all'azione di fattori esterni che possono orientare i processi decisionali. Come si approfondirà meglio nel secondo capitolo, i mass media e i social network si sono affermati come strumenti potenti per modellare opinioni e comportamenti.[1]

La formazione della volontà, che si sviluppa attraverso l'interazione di desideri, emozioni e riflessioni, può essere influenzata da stimoli esterni in maniera significativa. I mass media, grazie alla loro capacità di selezionare le notizie e di creare realtà percepite attraverso rappresentazioni ripetitive e mirate, hanno un ruolo centrale in questo processo. I social network hanno ulteriormente amplificato questo potere, utilizzando algoritmi sofisticati per costruire bolle di filtraggio e manipolare l'attenzione degli utenti, fenomeni che verranno esaminati più dettagliatamente nel secondo capitolo. Un ruolo cruciale è svolto dai bias cognitivi, scorciatoie mentali che possono semplificare i processi decisionali ma che, allo stesso tempo, rendono tali processi vulnerabili a errori sistematici. Come si dirà meglio in seguito, questi bias vengono spesso sfruttati in ambiti come il neuromarketing e le strategie persuasive, dimostrando l'intreccio tra tecnologia, psicologia e società.[2]

L'intelligenza artificiale rappresenta un elemento chiave del progresso tecnologico odierno, con l'obiettivo di replicare e amplificare le capacità cognitive umane. Si tratta di una disciplina complessa, declinata in varie tipologie e architetture, tra cui l'intelligenza artificiale "stretta", progettata per compiti specifici, e l'intelligenza artificiale "generale", capace di adattarsi a contesti diversi. Come si illustrerà nel terzo capitolo, tecnologie come il machine learning e il deep learning consentono alle macchine di apprendere dai dati e migliorarsi

[1] A. Maslow, "*Motivation and Personality"* (1954) Harper & Row, pubblicato digitalmente da Digital Library Of India (31/12/2005) <www.archive.org>, Pp. 35 - 59

[2] E. Cheli, "*La realtà mediata*" (6° edizione 2002) FrancoAngeli srl, Pp. 164 – 180.

autonomamente, trovando applicazione in molteplici settori, dalla medicina alla logistica, dall'intrattenimento alla finanza. Le sfide legate all'uso di queste tecnologie, come la protezione della privacy e la necessità di trasparenza, sono affrontate da normative come il GDPR e l'AI Act, questioni che verranno meglio approfondite nei capitoli successivi. Inoltre, i risvolti commerciali e le implicazioni economiche legate all'impiego dell'intelligenza artificiale aprono ulteriori scenari di interesse, come si discuterà nel sesto capitolo.

I Deepfake costituiscono una delle manifestazioni più controverse dell'intelligenza artificiale generativa, essendo in grado di creare contenuti multimediali falsificati con un realismo straordinario. Come si analizzerà nel quarto capitolo, questi strumenti combinano tecniche avanzate di deep learning e reti neurali per manipolare immagini, video e audio. Sebbene tali tecnologie trovino applicazioni lecite e creative, ad esempio nel cinema e nella pubblicità, emergono anche problematiche etiche e sociali rilevanti.

I Deepfake sollevano interrogativi sulla disinformazione, la violazione della privacy e la manipolazione dell'opinione pubblica, aspetti che saranno discussi più nel dettaglio nella trattazione dedicata. Ulteriori implicazioni legali e morali, che comprendono la violazione del diritto d'immagine e l'uso improprio in contesti politici e personali, saranno affrontate anche nei capitoli dedicati alla legislazione e alla regolamentazione internazionale (§6, §7, §8, §9).

Le sfide normative legate ai Deepfake, così come alle altre applicazioni dell'intelligenza artificiale, richiedono un approccio bilanciato per prevenire abusi senza ostacolare il progresso tecnologico. Soluzioni tecniche, come i watermark digitali, e strategie educative mirate a sviluppare il pensiero critico e la consapevolezza tra gli utenti, rappresentano passi fondamentali per affrontare queste questioni. Come si vedrà meglio nei capitoli successivi, la regolamentazione e la gestione di queste tecnologie devono essere integrate in una visione più ampia, che tenga conto delle loro implicazioni sociali, economiche e politiche; In particolare, i capitoli finali offriranno una panoramica sulla pericolosità dell'immagine artificiale e sugli strumenti legislativi sviluppati e in fase di sviluppo per arginare i rischi associati.

L'analisi di questi temi mette in luce l'interconnessione tra il condizionamento della volontà, l'intelligenza artificiale e i Deepfake, richiedendo un approccio interdisciplinare per comprenderne le implicazioni. Se da un lato queste innovazioni offrono opportunità straordinarie per migliorare la vita umana, dall'altro sollevano questioni etiche e legali di grande rilievo. Nei capitoli successivi si approfondiranno ulteriormente questi aspetti, esplorando

anche le potenzialità positive e i rischi connessi al loro impiego, in un contesto che richiede un equilibrio tra innovazione e responsabilità per garantire che il progresso tecnologico sia al servizio della collettività. La riflessione finale, come si vedrà nelle conclusioni, mira a tracciare possibili scenari futuri che, pur abbracciando l'innovazione, pongano al centro i diritti fondamentali dell'individuo.

2 Il condizionamento della volontà e i suoi utilizzi

.1 La formazione della volontà del singolo e della massa

In *incipit* a questo capitolo si ritiene che sia un buon punto di partenza trattare di quelli che è il processo di formazione della volontà, sia per quanto riguarda il singolo che per le masse, in modo da meglio permettere la comprensione delle influenze che le intelligenze artificiali possono avere sulla stessa.

La volontà è uno degli elementi chiave per comprendere il comportamento umano, sia a livello individuale che collettivo. È la forza interiore che ci spinge ad agire e si sviluppa attraverso un processo di sintesi tra desideri, bisogni, emozioni e riflessioni. Quando parliamo di volontà individuale, ci riferiamo al percorso di valutazione e scelta che ogni persona compie, basato su fattori personali e contesti. Nel caso delle masse, invece, la volontà emerge da dinamiche sociali più complesse, che spesso vanno oltre le capacità decisionali del singolo. Analizziamo, quindi, come si forma la volontà in entrambe le sfere, individuale e collettiva, esplorando i diversi fattori che contribuiscono alla sua costruzione.

A livello individuale, la volontà è profondamente influenzata dai bisogni e dai desideri. Abraham Maslow ha proposto un modello in cui i bisogni umani si dispongono in una gerarchia, partendo da quelli più elementari, come le necessità fisiologiche e di sicurezza, fino a quelli più complessi, come l'autorealizzazione[3]. La volontà personale si orienta secondo questa gerarchia: quando i bisogni fondamentali sono soddisfatti, l'individuo è in grado di dedicarsi a obiettivi più elevati, come la creatività o lo sviluppo personale. In questo senso, la volontà è una manifestazione della spinta dell'individuo a soddisfare i propri bisogni e raggiungere i propri obiettivi.

Non meno importante è il ruolo dei processi cognitivi nella formazione della volontà[4]. La psicologia cognitiva ci insegna che la volontà si forma attraverso un processo di valutazione

[3] Vedi: la Piramide di Maslow: A. Maslow, "*A theory of human motivation*", 1943, successivamente ampiata dallo stesso autore nella pubblicazione: "*Motivation and Personality"* del 1954, proprio con riferimento a quest'ultima pubblicazione, Maslow scrive che la volontà nasce spesso dai desideri, bisogni e aspirazioni. La piramide di Maslow ne offre una rappresentazione utile, mostrando come i bisogni di base (fisiologici, sicurezza) influenzano le motivazioni, che evolvono poi verso bisogni più complessi come l'autorealizzazione.

[4] Questa nota non si riferisce a una specifica pubblicazione, ma a un insieme di ricerche e teorie psicologiche sul controllo cognitivo e sul processo decisionale. Queste ricerche sono condotte da vari studiosi nellambito della psicologia cognitiva e delle neuroscienze, tra cui autori come, tra i molti: Daniel Kahneman, Amos Tversky, Roy Baumeister, i quali hanno studiato temi come il controllo degli impulsi, lautoregolazione e

delle alternative, pianificazione delle azioni e previsione delle conseguenze. Questa capacità di controllo cognitivo, che permette di inibire impulsi immediati per compiere scelte ponderate, è essenziale per una volontà coerente e orientata verso obiettivi a lungo termine. Le decisioni che prendiamo, infatti, dipendono in gran parte da quanto riusciamo a gestire i nostri desideri immediati in favore di scelte più razionali.

Oltre alla componente cognitiva, le emozioni svolgono un ruolo fondamentale nel plasmare la nostra volontà. Le neuroscienze hanno mostrato come il sistema limbico, la parte del cervello che gestisce le emozioni, influenzi fortemente le nostre motivazioni e decisioni. Le emozioni ci spingono ad agire, e in molti casi la nostra volontà è una risposta alle nostre emozioni piuttosto che a un calcolo razionale. Questo ci ricorda che non siamo solo esseri razionali, ma anche profondamente influenzati dai nostri stati emotivi.

Un altro aspetto fondamentale nella formazione della volontà individuale è il contesto sociale in cui cresciamo. Jean Piaget[5], attraverso i suoi studi sullo sviluppo cognitivo e morale, ha dimostrato che le nostre scelte e il nostro senso di giustizia sono modellati dall'interazione con l'ambiente sociale. Le esperienze con genitori, insegnanti e amici formano i valori e le norme che guidano la nostra volontà. Di conseguenza, la nostra volontà non è il frutto esclusivo di una riflessione interiore, ma anche del contesto in cui viviamo e delle persone che ci circondano.

Quando passiamo dalla volontà individuale a quella collettiva, il quadro cambia significativamente. La volontà delle masse è un fenomeno che si sviluppa attraverso dinamiche di gruppo, in cui le persone tendono a conformarsi ai valori e ai comportamenti condivisi. Come ha spiegato Kurt Lewin, nelle dinamiche di gruppo la volontà del singolo spesso viene assorbita da quella collettiva, portando le persone ad agire in modi che, da sole, non avrebbero mai considerato[6]. Il desiderio di appartenenza al gruppo e di accettazione diventa un motore potente che influenza la volontà.

Questo fenomeno è strettamente legato al conformismo, un aspetto ben esplorato da Solomon Asch nei suoi studi sull'influenza sociale. Asch ha dimostrato come gli individui, di

il processo decisionale. In questo particolare contesto, il concetto di controllo cognitivo va riferito alla capacità di una persona di modulare i propri pensieri e azioni in funzione degli obiettivi e delle circostanze, ed è stato esplorato in modo approfondito nei campi della psicologia e delle neuroscienze cognitive. Infra multis: R. F. Baumeister, K. D. Vohs, & D. M. Tice, "*The Strength Model of Self-Control*" (2007).

[5] J. Piaget, "*Il giudizio morale nel fanciullo"* (1932).

[6] K. Lewin, "*Field Theory in Social Science*" (1947).

fronte alla pressione del gruppo, siano spesso disposti a cambiare le proprie opinioni pur di allinearsi a quelle della maggioranza. Anche quando la volontà individuale è in contrasto con quella collettiva, la pressione del gruppo può annullare le convinzioni personali, portando a comportamenti conformisti che non sempre sono razionali[7].

Uno dei principali studiosi della psicologia delle folle, Gustave Le Bon, ha evidenziato come, nelle grandi masse, l'individuo tenda a perdere la propria identità e a dissolvere la propria volontà in quella del gruppo. Secondo Le Bon, le folle agiscono spesso in modo impulsivo ed emotivo, guidate da leader carismatici o dall'atmosfera generale del momento. Nelle folle, la razionalità individuale viene messa da parte, e le persone diventano capaci di comportamenti che, da sole, non avrebbero mai intrapreso[8].

Oltre alla spontanea formazione della volontà collettiva, esiste anche un'altra modalità attraverso cui essa si forma: la manipolazione. Antonio Gramsci ha introdotto il concetto di egemonia culturale, spiegando come le classi dominanti possano influenzare e modellare la volontà delle masse attraverso il controllo dei mezzi di comunicazione e dell'educazione. Questo processo permette alle élite di orientare le idee e i desideri della collettività in modo spesso inconsapevole, facendo sì che la volontà collettiva rifletta gli interessi di chi detiene il potere[9].

Un altro fattore importante nel contesto della volontà collettiva è il contagio sociale, ovvero la tendenza delle persone a imitare il comportamento altrui, soprattutto in situazioni di forte emotività. Stanley Milgram, con i suoi esperimenti sull'obbedienza all'autorità, ha dimostrato come gli individui siano inclini a seguire le direttive di un'autorità senza mettere in discussione la validità delle azioni richieste. Nelle folle, questo fenomeno si amplifica: l'emozione e l'impulso del momento prevalgono sulla riflessione razionale, spingendo le persone ad agire in modo coordinato ma spesso inconsapevole[10].

In conclusione, la volontà è un processo complesso che si forma sia a livello individuale che collettivo, e che coinvolge aspetti cognitivi, emotivi e sociali. La volontà individuale è influenzata dai bisogni, dalle emozioni e dalle capacità cognitive della persona, mentre la volontà delle masse si sviluppa attraverso dinamiche sociali, conformismo, leadership e,

[7] S. E. Asch, "*Effects of Group Pressure upon the Modification and Distortion of Judgments*" (1951).
[8] G. Le Bon, "*La psicologia delle folle"* (1895).
[9] A. Gramsci, "*Quaderni del carcere*" (1971).
[10] S. Milgram, "*Obedience to Authority: An Experimental View*" (1974).

talvolta, manipolazione. Quando la volontà si esprime a livello collettivo, essa può portare a comportamenti potenti e coordinati, che però non sempre riflettono la razionalità individuale.

.2 L'effetto dei mass media e dei social network sull'attenzione e la volontà

Nel corso del tempo, la transizione verso quella che Fritz Machlup ha definito "società dell'informazione" ha segnato un cambiamento epocale nel modo in cui le risorse scarse vengono definite e percepite. Machlup ha sottolineato come, nel contesto della terza rivoluzione industriale, la competizione economica si sposti dal controllo dei beni materiali verso una nuova forma di risorsa: l'informazione[11]. Mentre in passato il carbone, l'acciaio e altri beni tangibili erano al -centro dell'industria, oggi, nella società dell'informazione, ciò che è realmente prezioso e limitato è l'attenzione delle persone.

Questo cambiamento non è soltanto una questione tecnologica o economica; è una vera rivoluzione culturale che coinvolge il modo in cui viviamo e interpretiamo il mondo. La globalizzazione e l'avvento di Internet hanno abolito gran parte delle barriere fisiche che una volta limitavano l'accesso all'informazione. Oggi, le persone di ogni angolo del pianeta hanno accesso quasi istantaneo a una quantità infinita di dati, notizie e conoscenze. Tuttavia, mentre l'informazione si espande senza limiti, la capacità umana di assimilarla rimane invariata, rendendo l'attenzione il nuovo campo di battaglia.

La competizione per l'attenzione è intensa, e i media giocano un ruolo chiave in questo scenario. Qui entra in gioco la teoria dell'agenda setting, elaborata da Maxwell McCombs e Donald Shaw negli anni '70. La loro ricerca ha dimostrato che i media non si limitano a fornire informazioni, ma influenzano attivamente le priorità del pubblico, decidendo quali temi meritano attenzione. Questo non significa che i media dicano alle persone cosa pensare, ma determinano su cosa le persone debbano riflettere e discutere. Se un evento riceve grande

[11] F. Machlup, "*The Production and Distribution of Knowledge in the United States*" (1962). Princeton University Press.

copertura mediatica, diventa immediatamente il fulcro delle conversazioni, guidando le scelte sociali e politiche degli individui.[12]

Un ulteriore passo nella comprensione di come i media plasmano la nostra percezione del mondo arriva dalla teoria della coltivazione di George Gerbner. In questo studio Gerbner affermava che l'esposizione prolungata ai contenuti mediatici contribuisce a modellare una visione del mondo che può essere profondamente distorta rispetto alla realtà. Un esempio classico è quello della rappresentazione della violenza[13]. Chi consuma costantemente media che enfatizzano pericoli e crimini tende a sviluppare una percezione esagerata dei rischi nella vita reale. Questo può influenzare decisioni importanti, come il livello di fiducia verso gli altri o il modo in cui ci si approccia alle questioni politiche e sociali.

Oltre a modellare la percezione del mondo, i media e i social network sfruttano potenti meccanismi psicologici come la ripetizione e la persuasione. Lo psicologo Robert Cialdini, nel suo celebre testo "Influence: Science and Practice", ha evidenziato come la ripetizione costante di un messaggio lo renda più familiare e quindi più credibile. Questo effetto è ampiamente utilizzato nel marketing e nelle campagne politiche per convincere le persone a cambiare idea o a fare scelte che altrimenti non avrebbero considerato[14].

Con l'ascesa dei social network, queste dinamiche sono diventate ancora più pervasive. Le piattaforme come Facebook, Twitter e Instagram utilizzano algoritmi che creano delle vere e proprie "eco chambers", bolle in cui gli utenti sono esposti solo a contenuti che confermano le loro opinioni preesistenti. Questo fenomeno riduce l'apertura a punti di vista diversi e alimenta la polarizzazione delle opinioni, con conseguenze significative sulla coesione sociale[15].

---

[12] M. McCombs &D.L. Shaw, "*The Agenda-Setting Function of Mass Media*" (1972). Public Opinion Quarterly. 36, pp. 176–187

[13] G. Gerbner, & L. Gross, "*Living with Television: The Violence Profile*" (1976). Journal of Communication, 26, pp.173–199

[14] R. B. Cialdini, "*Influence: Science and Practice*" (29 luglio 2008, quinta edizione). Allyn & Bacon

[15] Eli Pariser, nella sua pubblicazione, ha esplorato come la personalizzazione dei contenuti influenzi negativamente la capacità del soggetto di vedere e percepire il mondo in modo più ampio e sfaccettato, ovvero, in questo caso permetta solo una visione più ristretta e focalizzata di una parte dello scibile permessa e concessa dal contenuto stesso, causando quella che può essere definita come una polarizzazione indotta dall'algoritmo ovvero un filtro digitale dell'informazione. Cfr. E. Pariser, "*The Filter Bubble: What the Internet Is Hiding from You*" (2011) Penguin Press, e successivamente riedito (2012) UCLA Journal of education and information studies, 8(2).

Un altro fenomeno strettamente legato all'uso dei social network è il FOMO (Fear of Missing Out), ossia la paura di essere esclusi da eventi o esperienze significative. Questo concetto è stato analizzato da Andrew Przybylski, e suoi colleghi, in uno studio del 2013, che ha dimostrato come l'ansia di non partecipare a ciò che gli altri stanno vivendo possa spingere le persone a prendere decisioni impulsive, come acquistare prodotti, partecipare a eventi o condividere contenuti per sentirsi parte di un gruppo. Questa dinamica viene ampiamente sfruttata dalle aziende attraverso strategie di marketing che creano un senso di urgenza, come offerte limitate nel tempo o promozioni esclusive.[16]

La manipolazione delle informazioni non si ferma qui. Con la diffusione delle fake news, l'ecosistema mediatico è diventato ancora più complicato. Studi condotti da Sinan Aral e Deb Roy, condotti negli anni appena precedenti al 2020, hanno dimostrato che le fake news, specialmente quelle che provocano indignazione o sorpresa, si diffondono sui social media molto più rapidamente rispetto alle notizie verificate.[17] Questo è un fenomeno preoccupante, poiché può influenzare pesantemente le decisioni politiche e sociali, spingendo le persone a compiere scelte basate su informazioni false.

Un caso emblematico di manipolazione dell'opinione pubblica attraverso i social media è stato lo scandalo di Cambridge Analytica, dove i dati personali raccolti dai profili social sono stati usati per creare messaggi mirati, capaci di influenzare le scelte politiche degli elettori. Questo scandalo, esplorato in un'inchiesta pubblicata su The Guardian e edita da Carole Cadwalladr nel 2018, ha mostrato come la profilazione psicografica possa essere sfruttata per manipolare gli individui senza che questi ne siano consapevoli [18].

Infine, l'uso eccessivo dei social media ha conseguenze anche sulla salute mentale, soprattutto tra i giovani. Come evidenziato da Jean Twenge le piattaforme social presentano spesso una versione idealizzata della vita, spingendo gli utenti a confrontarsi con standard

[16] A. K. Przybylski, K. Murayama, C. R., DeHaan, & V. Gladwell, "*Motivational, emotional, and behavioral correlates of fear of missing out*" (2013). Computers in Human Behavior, 29, Pp. 1841-1848.

[17] S. Aral & D. Roy, "*The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health - and How We Must Adapt*" (2020). Crown Publishing Group.

[18] C. Cadwalladr, & E. Graham-Harrison, "*Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.*" (2018). The Guardian.

irrealistici di successo e felicità. Questo confronto continuo può generare sentimenti di inadeguatezza, ansia e depressione, influenzando le scelte personali e sociali[19].

In conclusione, i media e i social network esercitano un'influenza profonda sulle scelte umane. Dall'impostazione dell'agenda pubblica alla manipolazione delle opinioni, passando per la polarizzazione delle idee e la diffusione di fake news, questi strumenti modellano il nostro modo di vedere il mondo e di prendere decisioni. La capacità di navigare in questo complesso panorama informativo richiede una consapevolezza critica, capace di distinguere tra verità e manipolazione, tra scelte autentiche e condizionamenti sociali.

.3 Le derivazioni dirette della volontà: I bias e le decisioni

Secondo lo psicologo e premio Nobel per l'economia Daniel Kahneman, le scelte umane sono un compromesso tra processi automatici e razionali. Nel suo studio sul giudizio in condizioni di incertezza, Kahneman evidenzia come le persone utilizzino scorciatoie mentali, o "euristiche", per facilitare le decisioni quotidiane. Questi schemi semplificati riducono il dispendio energetico mentale ma possono causare errori di valutazione, detti "bias cognitivi". Le euristiche, come quella della disponibilità, fanno sì che si scelga sulla base della facilità con cui un caso analogo viene alla mente, mentre l'euristica della rappresentatività si basa sulle somiglianze con esperienze precedenti. L'ancoraggio, invece, porta a fondare le decisioni su un primo riferimento, influenzando tutte le valutazioni successive.[20]

Nel libro "Pensieri lenti e veloci", Kahneman introduce i due sistemi di pensiero: il sistema 1, intuitivo e rapido, e il sistema 2, razionale e lento. Il sistema 1 usa euristiche per rispondere velocemente, risparmiando energia, mentre il sistema 2 interviene solo in caso di analisi più complesse. Tuttavia, l'interazione tra questi sistemi può generare errori sistematici

[19] J. M. Twenge, "*iGen: Why Todays Super -Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy – and Completely Unprepared for Adulthood*" (2017), Simon and Schuster.

[20] A. Tversky, D. Kahneman, "*Judgment under uncertainty: heuristics and biases*", (1979), in Prospect Theory: An Analysis of Decision Under Risk, Econometrica. Vol 47(2) Pp. 263 - 291.

o Bias. La fluidità cognitiva, o la diminuzione dello sforzo con l'aumento della familiarità, porta a prendere decisioni sempre più rapide per contenere lo stress cognitivo.[21]

Diversi sono i bias cognitivi che influenzano le scelte, come il bias di conferma, che porta a cercare informazioni che confermino le credenze preesistenti, e il framing, che condiziona le decisioni in base al contesto in cui un'opzione è presentata. Esistono anche il bias di desiderio, legato alla propensione per ciò che si desidera, e il bias dei costi, che tende a sovrastimare il peso delle perdite.[22] [23]

Il concetto di scelta entra in gioco ogni volta che il consumatore si trova a decidere se effettuare un acquisto, cosa acquistare e quanto spendere. Con l'aumentare delle opzioni disponibili, tuttavia, cresce anche la difficoltà della decisione, tanto da poter portare a errori di scelta, noti come "bias di scelta". Questo fenomeno è stato analizzato in una ricerca condotta nel luglio 2002 da Alistair Rennie e Jhonny Protheroe del team di Google Scholar[24]. I due studiosi, analizzando il comportamento d'acquisto attraverso i dati online, hanno osservato che più aumentano le alternative e le informazioni, più caotico e disordinato diventa il processo decisionale del consumatore, portandolo spesso a confusione e indecisione. Con l'avvento di Internet, del Web 2.0, della comunicazione digitale e dei social media, questa fase caotica del percorso d'acquisto si è accentuata, creando ciò che viene definito "Messy Middle".

Il concetto di "Messy Middle" descrive le complesse dinamiche del processo decisionale dei consumatori nel mondo digitale, evidenziate in una ricerca condotta da Google per comprendere come l'abbondanza di opzioni e informazioni crei confusione e sovraccarico cognitivo. Questo processo, definito "disordinato", indica che il percorso verso l'acquisto non è lineare ma è caratterizzato da cicli di esplorazione e valutazione ripetuti. Durante queste fasi, i consumatori si trovano a confrontare prodotti, cercare informazioni e ponderare alternative spesso in modo caotico, prima di giungere alla decisione finale. L'ampia disponibilità di scelte e informazioni può persino portare a una situazione di "paralisi da analisi", ovvero un'incapacità di scegliere causata dall'eccesso di alternative.

---

[21] D. Kahneman, "*Pensieri lenti e veloci*" (2012), Mondadori, Milano.

[22] F. fiore "*Bias cognitivi: come la nostra mente ci inganna*" (16 maggio 2018) Sigmund Freud University (Mialno) < https://milano-sfu.it/ >

[23] Cfr: A. Ceschi, R. Sartori, "*Un approccio empirico per una tassonomia delle euristiche*" (1° gennaio 2021), Università di Verona.

[24] A. Rennie, J. Protheroe, "*Decoding Decisions: making sense of the messy middle.*" (luglio 2020) Think with Google <https://www.thinkwithgoogle.com/>

La ricerca di Google identifica due fasi principali nel "Messy Middle": l'esplorazione, durante la quale i consumatori cercano il maggior numero possibile di informazioni e aprono così la porta a una varietà di opzioni, e la valutazione, che è una fase di sintesi in cui i consumatori filtrano e valutano le informazioni raccolte per arrivare a una decisione finale. Queste due fasi si alternano in modo iterativo e rappresentano i passaggi principali prima che il consumatore giunga all'acquisto.

In questo contesto di incertezza, i bias cognitivi assumono un ruolo determinante, influenzando il modo in cui i consumatori elaborano le informazioni. I bias cognitivi possono essere sfruttati dalle aziende per rendere più fluido e diretto il percorso di acquisto, influenzando le scelte dei consumatori. Tra questi bias, il bias di scarsità rende un prodotto più desiderabile quando viene percepito come disponibile in quantità limitate, attivando nel consumatore il timore di perdere un'opportunità e quindi incentivando acquisti immediati. La prova sociale rappresentata da recensioni e testimonianze di altri acquirenti fornisce una sorta di rassicurazione, fungendo da scorciatoia decisionale e riducendo il bisogno di ulteriori confronti. Il bias di autorità porta a dare maggiore rilevanza alle opinioni di figure autorevoli o esperti, rafforzando la credibilità del prodotto o del marchio.[25]

Un'altra dinamica rilevante è quella del bias di ancoraggio, che si manifesta quando le persone basano le loro valutazioni su un primo riferimento – come un prezzo iniziale – che influenza tutte le successive considerazioni. Un prezzo scontato apparirà dunque più conveniente rispetto a un prezzo "ancorato" a un valore di partenza più alto. Anche il potere dell'immediatezza è significativo, poiché una spedizione rapida riduce l'ansia di attesa e incoraggia l'acquisto. Il bias di gratuità è altrettanto potente: offrire un regalo o un vantaggio tangibile in aggiunta al prodotto stimola il desiderio di acquisto. L'inclusione di un omaggio attiva, infatti, un senso di opportunità che spinge i consumatori a preferire quella specifica opzione.

I brand che comprendono questi bias e li integrano nelle loro strategie di marketing possono semplificare il percorso decisionale del consumatore e facilitare il processo d'acquisto. Per esempio, l'ottimizzazione della presentazione dei prodotti con descrizioni concise e chiare permette ai consumatori di prendere decisioni con maggiore rapidità e sicurezza, evitando il sovraccarico di informazioni. Anche il modo in cui le informazioni sono presentate ha un ruolo

[25] Ivi pp. 18 - 25

importante: l'uso del framing, cioè la scelta di evidenziare determinate informazioni, può orientare la percezione dell'offerta. Un prezzo scontato può apparire più conveniente se viene enfatizzato il risparmio rispetto al prezzo originale, mentre un prodotto di qualità superiore può essere presentato mettendo in risalto la durabilità o le caratteristiche distintive.[26]

Le aziende possono inoltre fare leva sui dati per personalizzare le raccomandazioni, rispondendo alle preferenze individuali e riducendo così il numero di opzioni da considerare, rendendo più semplice il processo decisionale. Gli incentivi come sconti, programmi di fedeltà e vantaggi esclusivi servono a mantenere alto l'interesse e possono sostenere la scelta finale.

Il "Messy Middle" rappresenta quindi non solo una sfida per i consumatori, ma anche un'opportunità per le imprese. Coloro che riescono a comprendere e influenzare i processi decisionali caotici dei consumatori adottando tecniche mirate e strategie di marketing ben studiate possono ottenere un vantaggio competitivo significativo, conquistando la fiducia dei consumatori e incrementando le vendite. [27]

In conclusione, comprendere e sfruttare i bias cognitivi permette ai brand di semplificare il percorso di acquisto, "aiutando" il consumatore a orientarsi nel caos di alternative disponibili. L'utilizzo strategico del framing, delle recensioni, della scarsità e dell'ancoraggio, così come l'ottimizzazione della presentazione dei prodotti e la personalizzazione delle raccomandazioni, possono ridurre l'incertezza e il sovraccarico cognitivo, facilitando il processo decisionale, generando, sostanzialmente, quella che può essere definita come una veicolazione della volontà del soggetto. In questo modo, le imprese non solo rispondono alle necessità dei consumatori, ma ottengono anche un vantaggio competitivo, conquistando la loro fiducia e incrementando la probabilità di fidelizzazione, limitando la probabilità che questi si spostino verso un competitor e, al contempo, rafforzando quella che è la presa che il brand ha sullo stesso. Saper navigare il "Messy Middle" significa, dunque, saper rispondere alla complessità del comportamento d'acquisto moderno, trasformandola in una strategia vantaggiosa sia per il consumatore che, soprattutto per quel che riguarda il lato economico, per l'azienda.

[26] ivi pp.87 - 91
[27] Ivi pp 92 - 96

3 Introduzione all'intelligenza artificiale

.1 Il dato nella sua semplicità e potenza

Per comprendere meglio i temi che verranno affrontati, è opportuno illustrare la tecnologia che sta alla base delle intelligenze artificiali (IA) e dei sistemi di apprendimento automatico, ripercorrendone le origini per far luce su strumenti che a molti appaiono ancora poco trasparenti e piuttosto recenti.

Le IA e il machine learning traggono spunto da discipline come matematica, statistica e informatica, sviluppate nel corso del XX secolo e culminate nella creazione delle soluzioni innovative di cui si parlerà.[28] L'apprendimento automatico svolge un ruolo di grande rilievo perché consente alle macchine di analizzare i dati in ingresso o le risposte alle proprie azioni e di migliorare le prestazioni nel tempo senza dover essere programmate in maniera esplicita per ciascun compito. Questa caratteristica rappresenta una svolta rispetto ai sistemi informatici tradizionali, nei quali era necessario definire rigide istruzioni per ogni tipo di scenario possibile.

La crescita della potenza di calcolo e la capacità di gestire un volume sempre maggiore di dati hanno reso possibili soluzioni e applicazioni sempre più evolute, oggi impiegate in campi eterogenei, per esempio nell'analisi economica, nella personalizzazione dei servizi, nella diagnosi medica e nella guida autonoma. A svolgere un ruolo cruciale in questo progresso è stata anche la diffusione di infrastrutture di cloud computing e reti ad alta velocità, che hanno permesso di analizzare informazioni molto più voluminose di quanto fosse immaginabile in passato. Questo aumento di complessità solleva però importanti dubbi legati all'affidabilità, alla sicurezza e alla trasparenza degli algoritmi. Spesso, infatti, i modelli di apprendimento automatico non lasciano facilmente intravedere il loro funzionamento interno: è complesso ricostruire in che modo abbiano raggiunto una particolare conclusione, perfino per gli stessi sviluppatori. Da qui, la percezione di vere e proprie "scatole nere", che alimenta la diffidenza di chi non è del settore e al tempo stesso stimola istituzioni e governi a introdurre normative più chiare e vincolanti.

[28] Nonostante la tecnologia e gli strumenti siano stati studiati a partire dallo scorso secolo, scrittori e pensatori non erano estranei al concetto di una macchina che potesse generare immagini o testi; di questo, tra i moltissimi, ne è prova quanto scritto da Jonathan Swift all'interno del quarto viaggio di Gulliver, ove questi, entrato all'accademia di Lagado si trova di fronte ad un automa di legno e di ferro la quale una volta attivata riorganizza le parole fornitole, arrivando alla produzione di un nuovo ed inedito libro; seppur l'azione fosse coadiuvata da alcuni copiatori, il concetto di intelligenza artificiale è già ravvisabile.

In questo scenario, per rendere le tecnologie di IA più chiare e affidabili, è utile esaminare le fondamenta normative che ne regolano l'utilizzo. Già prima del grande successo mediatico e dell'interesse di massa per l'intelligenza artificiale, il legislatore europeo aveva emanato il Regolamento Generale sulla Protezione dei Dati (GDPR)[29], approvato il 24 maggio 2016 ed entrato in vigore il 25 maggio 2018. Questo quadro normativo si propone di salvaguardare i diritti e le libertà fondamentali dei cittadini europei in relazione ai trattamenti di dati personali. Assume così un ruolo centrale nel dibattito contemporaneo su IA e machine learning, poiché le tecnologie di analisi dei dati si basano in gran parte sull'utilizzo di informazioni che possono ricondurre a individui specifici. Tali informazioni servono per addestrare gli algoritmi, per validarne le previsioni e per migliorare la qualità delle risposte o dei servizi offerti.

Secondo il Garante della Privacy, per dato personale si intendono le informazioni che identificano o rendono identificabile, direttamente o indirettamente, una persona fisica, fornendo indicazioni sulle sue caratteristiche, le sue abitudini, i suoi rapporti interpersonali, il suo stato di salute, la sua situazione economica o altri tratti distintivi.[30] Questa definizione appare molto ampia, tanto che qualunque dettaglio che, in modo diretto o incrociato con altri elementi, permetta di risalire all'identità di una persona può rientrare nella categoria dei dati personali.

Rientrano in questo ambito non soltanto nome e cognome, ma anche l'indirizzo IP, le coordinate GPS di un dispositivo mobile o la cronologia di navigazione, se questi dettagli si rivelano sufficienti a ricostruire il profilo di un soggetto.

Il GDPR suddivide inoltre l'identificazione in diretta e indiretta,[31] mettendo in evidenza come a volte bastino semplici correlazioni tra dati apparentemente anonimi per individuare, con un certo grado di sicurezza, un individuo.

Nel Regolamento vengono elencati anche i principi base per la corretta gestione delle informazioni personali, quali la liceità, la correttezza, la trasparenza, la limitazione delle finalità e la minimizzazione dei dati. Questi principi richiamano la necessità di acquisire e utilizzare le

[29] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), generalmente detto GDPR.

[30] "*Cosa intendiamo per dati personali? **" Garante della Privacy <garanteprivacy.it >

[31] Art 4 Regolamento (UE) 2016/679 (GDPR)

informazioni dell'interessato in maniera proporzionata e nel rispetto dei suoi diritti fondamentali. Particolare attenzione è posta sui cosiddetti dati sensibili, definiti anche dati particolari, che comprendono, tra gli altri, l'origine razziale, le convinzioni religiose, l'orientamento politico, l'appartenenza sindacale, i dati biometrici, genetici o sanitari. Tali informazioni godono di una tutela rafforzata in quanto potrebbero incidere sui diritti della persona in modo significativo. Il GDPR permette comunque alcune eccezioni in cui è consentito trattare tali dati, per esempio in ambito medico o laddove sussistano condizioni di necessità, purché siano rispettati i requisiti stringenti in merito a consenso, sicurezza informatica e finalità legittime. [32]

La protezione dei dati personali assume un'importanza ancora maggiore se si considerano gli attuali sistemi di IA e apprendimento automatico, spesso addestrati su grandi raccolte di informazioni provenienti da molteplici fonti. Immagini, dati biometrici o registrazioni vocali risultano fondamentali per il funzionamento di applicazioni come il riconoscimento facciale, la comprensione del linguaggio naturale o l'analisi predittiva di condizioni mediche. In ognuno di questi casi, il GDPR esige che i soggetti che gestiscono tali dati mettano in atto misure di sicurezza adeguate, e incoraggia l'adozione di tecniche come l'anonimizzazione o la pseudonimizzazione, particolarmente utili in fase di condivisione o trasferimento dei dati.[33]

Un ulteriore aspetto significativo è rappresentato dal modo in cui le IA sfruttano i segnali derivati dalla percezione collettiva del mondo. Come già rilevato in altri contesti (§2.1), la volontà individuale non si forma in un vuoto sociale, ma è condizionata da emozioni, contesti culturali e background personali. Le IA possono utilizzare questi fattori contestuali per creare modelli in grado di offrire risposte personalizzate e predizioni sempre più raffinate. Basti pensare alle piattaforme di social media, che raccolgono in continuo dati attraverso like, condivisioni e commenti, tracciando preferenze e inclinazioni di milioni di utenti e, di fatto, delineando un ritratto piuttosto accurato di abitudini, gusti e interazioni. Questo meccanismo

[32] Art 9 Regolamento (UE) 2016/679 (GDPR): "*trattamento di categorie particolari di dati personali*".
[33] G. Sartor & F. Lagioia "*The impact of the General Data Protection regulation (GDPR) on artificial intelligence*", (giugno 2020) EPRS European Parlamentary reserch service, Pp. 73 – 77.

può dare origine a raccomandazioni di contenuto molto mirate, ma può anche generare fenomeni di profilazione aggressiva se non regolamentato in maniera opportuna (§6).[34]

Tenere conto delle implicazioni sociali e legali dell'IA è fondamentale per affrontare le sfide che si profilano, come l'equilibrio tra la protezione della privacy e la necessità di trasparenza, la questione dei bias nascosti nei dati di addestramento, la sicurezza informatica e la possibilità di utilizzare i dati per scopi secondari non autorizzati. In un mondo sempre più interconnesso, le informazioni personali spesso attraversano diversi Paesi con normative differenti, rendendo complicata una gestione armonica della privacy a livello globale.

Di fronte a questi problemi, emergono proposte che puntano a rendere gli algoritmi più trasparenti e comprensibili, cercando di spiegare i processi decisionali a posteriori. Contemporaneamente, molte realtà si muovono nella direzione della privacy-by-design,[35] ovvero la progettazione di sistemi informatici e infrastrutture che incorporino sin dall'origine requisiti di sicurezza, protezione e minimizzazione nell'uso dei dati.

Le prospettive future mostrano che l'IA e il machine learning continueranno a svilupparsi con ritmi accelerati, introducendo nuovi servizi basati su reti neurali profonde, tecniche di elaborazione del linguaggio naturale avanzate e sistemi in grado di analizzare segnali complessi come il riconoscimento visivo o l'elaborazione di emozioni.[36] Tale evoluzione richiederà che enti pubblici, aziende, centri di ricerca e società civile collaborino in maniera coesa per definire standard e procedure globali. Sarà indispensabile formare professionisti in grado di comprendere tanto le sfumature tecnologiche quanto quelle giuridiche, etiche e sociali legate ai dati personali. Solo in questo modo sarà possibile cogliere appieno le potenzialità dell'innovazione e, al contempo, tutelare i diritti fondamentali delle persone in una società sempre più automatizzata.

L'attuale cornice legislativa rappresentata dal GDPR, assieme all'AI act, come vedremo in seguito (§7) offrono già un insieme di linee guida indispensabili, tra cui il principio di limitazione della finalità, la minimizzazione dei dati e il diritto all'oblio, che mirano a garantire

---

[34] E. Sigel "*Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* di Eric Siegel" (febbraio 2016), John Wiley & Sons. Pp. XXI-XXIIX.

[35] Questo anche in osservanza dell'art 25 del Regolamento (UE) 2016/679 (GDPR), il quale prevede regole stringenti per la gestione, reperimento e conservazione dei dati personali, anche in base al livello di sensibilità delle informazioni associate ai dati stessi.

[36] Vedi: M. Martorana & R. Savella "*IA e riconoscimento delle emozioni: rischi e possibili vantaggi*" (21 settembre 2023) Agenda Digitale. <www.agendadigitale.eu >

una reale protezione della dignità umana. Questa normativa, però, non risponde a tutti gli interrogativi sollevati dalle nuove tecnologie, e si prospettano probabili aggiornamenti, in particolare nell'ambito dell'intelligenza artificiale più avanzata e delle applicazioni emergenti che potrebbero influenzare settori cruciali della vita pubblica e privata.

.2 Comprendere il fondamento delle intelligenze artificiali

Per quanto sia un campo in velocissima evoluzione in questo periodo storico, i natali dell'idea sono da ricercare, come meglio verrà spiegato nel sottocapitolo dedicato alle origini e le evoluzioni temporali (§ 3.4), all'anno 1950, anno di pubblicazione dell'articolo "Computing Machinery and Intelligence" di Alan Turing.

Per pura completezza e logicità del discorso sembra utile, se non necessario, originare quanto segue partendo da cosa sia l'intelligenza artificiale - termine che come standard viene abbreviato in IA (o AI seguendo quella che è la locuzione inglese "*artificial intelligence*"), abbreviazione che verrà largamente utilizzata anche nella stesura di questo testo -; purtroppo, ad oggi, non esiste una definizione univoca di IA, ma viene generalmente intesa come la scienza e l'ingegneria della creazione di macchine intelligenti in grado di eseguire compiti complessi, adattarsi all'ambiente e comprendere il linguaggio naturale. In sostanza, i sistemi di IA acquisiscono, elaborano e interpretano i dati per definire modelli e raggiungere un risultato, questo sotto l'assistenza umana o in semi autonomia.

Il concetto di IA si fonda, quindi sulla volontà di riprodurre attraverso un sistema artificiale le capacità cognitive umane, come l'apprendimento, il ragionamento, la pianificazione e la comunicazione. Tuttavia, il suo obiettivo non è la semplice imitazione dell'intelligenza umana, ma contempla la creazione di sistemi in grado di ottenere risultati pari o migliori, anche senza replicare fedelmente i processi mentali umani.

Come esempi possiamo pensare a sistemi più o meno moderni e applicati in campi molto differenti: si veda, infatti, l'intelligenza artificiale di IBM denominata BeepBlue, progettata e

realizzata agli inizi degli anni ‘90 per giocare a scacchi e che vinse contro il campione mondiale prima nel 1996 e poi, nelle successive iterazioni e aggiornamenti del software con sempre maggior frequenza, dimostrando la supremazia della macchina sull’uomo, ponendo, anche, questioni morali che in parte possiamo veder riprese in produzioni cinematografiche come, per citarne uno: Matrix.

Considerando campi più tecnici possiamo nominare il progetto AlphaFold di Alphabet (società controllata da Google) il quale ha “risolto”, in soli quattro anni, il problema della “protein folding”, ovvero la possibilità di predire il comportamento e la “piega” che assumerebbero proteine semplici e aggregati di queste; questa stessa questione ha attanagliato le menti dei biologi e dei bio-ingegneri per cinquant’anni, arrivando a sbrogliare una quota stimata del 17%, ovvero ottenendo la forma tridimensionale di circa 20 000 proteine, prima dell’avvento di questa intelligenza artificiale, per giungere ad una quota stimata del 98,5% nell’arco di qualche anno.[37]

Per fare una rapidissima sintesi, in quanto l’argomento verrà meglio trattato nel prossimo sottocapitolo, l’intelligenza artificiale si riferisce a sistemi automatizzati progettati per funzionare con diversi livelli di autonomia. Questi sistemi, elaborando grandi quantità di dati (input), sono in grado di generare output come previsioni, contenuti, raccomandazioni o decisioni che possono influenzare sia l’ambiente fisico che quello virtuale.

È importante considerare, molto brevemente, il significato della parola “dato”, questo, di fatti possiede diverse definizioni in base all’utilizzo che se ne vuole fare, ma ricercando la generalità più assoluta, possiamo indicare come dato una qualsiasi informazione che rappresenti un fatto o una qualsiasi idea, statistica che possa essere analizzata o archiviata al fine di essere utilizzata, fruita ovvero mantenuta da una macchina o un utente.

Tornando all’argomento principe di questo sottoparagrafo, le fonti prese in considerazione per la stesura di questo elaborato presentano diverse definizioni di IA, evidenziandone la complessità e la mancanza di un’unica interpretazione. Ad esempio, il DDL Butti definisce l’IA come un sistema automatizzato con diversi livelli di autonomia, progettato per generare output che influenzano ambienti fisici o virtuali. Un’altra definizione, proposta dall’High Level Expert Group on AI (AI HLEG), la quale richiamando quanto detto

[37] Vedi: R. Towels, “*Alphafold is the most important achivement in AI ever”* (3 ottobre 2021), Frobes (digitale).

dall'informatico Jhon McCarthy durante Il Dartmouth Summer Research Project on Artificial Intelligence del 1955[38] descrive l'IA:

> *"I sistemi di intelligenza artificiale (IA) sono sistemi software (ed eventualmente anche hardware) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il loro ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulla conoscenza o elaborando le informazioni derivate da questi dati e decidendo la migliore linea di condotta da adottare per raggiungere l'obiettivo prefissato. I sistemi di IA possono utilizzare regole simboliche o apprendere un modello numerico e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti."*[39]

È importante notare che, come indicato nella fonte, la maggior parte dei sistemi di IA attualmente esistenti svolge solo una parte delle attività descritte in questa definizione. Tuttavia, alcuni sistemi avanzati, come i veicoli a guida autonoma, possono combinare diverse di queste capacità.

L'IA non si limita a processi computazionali, ma rappresenta un "nuovo mondo artificiale della conoscenza umana"[40]. Questo nuovo mondo include l'informatica cognitiva, l'apprendimento automatico e la collaborazione uomo-macchina, portando ad evidenti benefici e semplificazioni per quel che riguarda tanto il quotidiano quanto l'esperienza lavorativa, di fatto permettendo all'utilizzatore di vivere una, a detta di chi scrive, pericolosissima simbiosi uomo-macchina, che permea moltissimi aspetti della vita di ogni giorno, inficiando, quindi, anche sulle decisioni prese in autonomia o in coadiuvati dagli strumenti.

---

[38] McCarthy, M.L. Minsky, N. Rochester & C.E. Shannon, "*A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*", originalmente pubblicato il 31 agosto 1955, e riedito in AI Magazine (2006) vol 26 - 4, p. 12, "an attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. (...). For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving", volendo dare un'estrema sintesi, concetto del discorso intavolato dall'autore risulta essere che l'intelligenza artificiale è la scienza atta a realizzare macchine che sappiano eseguire compiti che, se fatti da un essere umano richiederebbero intelligenza.

[39] G. Sartor & F. Lagioia "*The impact of the General Data Protection regulation (GDPR) on artificial intelligence*", (giugno 2020) EPRS European Parlamentary reserch service, p. 2.

[40] In questo senso V. Frosini "*Il giurista e le tecnologie dell'informazione*", (1998) Bulzoni editore, Roma, p. 38, che così prosegue: "…per cui l'uomo d'oggi può essere definito come "uomo artificiale", nel senso che egli è l'artefice della sua nuova struttura psicologica per mezzo delle macchine, che sono un prodotto dell'intelligenza e dell'operosità dell'uomo e sono pertanto anche "naturali", cioè consustanziali alla creatura umana intesa in senso morale e non fisico".

Sebbene l'IA abbia già portato a notevoli progressi, come l'automazione del lavoro manuale e intellettuale, è necessario restare in guardia dai potenziali rischi, come le violazioni della privacy, della sicurezza e dell'eguaglianza. Pertanto, è fondamentale uno sviluppo dell'IA etico e responsabile, che metta l'uomo al centro e garantisca la fiducia degli utenti.

.3 Diverse tipologie e architetture

Come già evidenziato nel sottocapitolo precedente, il campo dell'intelligenza artificiale (IA) comprende una vasta gamma di sistemi e modelli progettati per imitare le capacità cognitive umane in vari campi applicativi. Nonostante non esista un'unica definizione esaustiva di IA, è possibile delineare una classificazione basata sulle funzioni o azioni che questi sistemi cercano di replicare, fornendo così un quadro più chiaro delle loro potenzialità e limiti.

Prima di procedere a una distinzione più dettagliata, si trova utile parlare dei modelli di intelligenza artificiale in termini generali, suddividendoli in alcune categorie generali:

IA "stretta" (Narrow AI): Questo tipo di intelligenza artificiale è progettato per eseguire compiti specifici e limitati; di fatti, queste AI sono progettare con delle limitazioni strutturali estremamente rigide. Tra gli esempi più comuni troviamo il riconoscimento facciale, la traduzione automatica, e i sistemi di raccomandazione. Queste IA sono estremamente efficaci nel compito per cui sono state programmate, ma non possiedono la capacità di apprendere o risolvere problemi al di fuori del loro ambito predefinito. Ad esempio, un sistema di riconoscimento facciale non sarà in grado di tradurre un testo o risolvere problemi di altro genere. Si tratta di un'IA specializzata, che eccelle in aree ben definite ma non può operare al di fuori di esse.[41]

[41] Secondo il parere del Comitato economico e sociale europeo, 31 maggio 2017, L'intelligenza artificiale - Le ricadute dell'intelligenza artificiale sul mercato unico (digitale), sulla produzione, sul consumo, sull'occupazione e sulla società, 2017/C 288/01, punto 2.2., "l'IA può essere suddivisa schematicamente in IA "stretta" (narrow AI) e IA "generale" (general AI). L'IA stretta è in grado di portare a termine mansioni specifiche, mentre quella generale può eseguire qualsiasi compito intellettuale che un essere umano è in grado di svolgere".

IA "generale" (general AI): Al contrario, l'IA generale mira a replicare l'intelligenza umana in modo più completo e flessibile. Questa forma di IA, ancora in gran parte teorica, sarebbe in grado di apprendere, adattarsi e risolvere problemi in vari contesti, proprio come farebbe un essere umano. Si tratta di sistemi capaci di rispondere a domande o svolgere compiti senza limitarsi a uno specifico ambito. Sebbene lo sviluppo di una vera IA generale rimanga un traguardo futuro, molte ricerche si concentrano su come avvicinarsi a questo obiettivo, lavorando su algoritmi sempre più complessi e flessibili. Un sistema che può avvicinarsi a sufficienza per dare un'idea di fondo delle potenzialità di questa tipologie di intelligenze per quanto sia solo una finzione, la possiamo ritrovare nell'infame intelligenza artificiale chiamata Skynet, dal film Terminator, ovvero un sistema capace di conoscere, comprendere e reagire a qualsiasi stimolo al fine di perpetrare un proprio fine, ovviamente siamo molto lontani dalla realizzazione di un sistema simile, ma visti i ritmi di avanzamento, osservando anche le analisi strutturate dal filosofo svedese Bostrom, può già essere fonte di preoccupazione come vedremo meglio nel prossimo paragrafo.

IA generativa: queste sono un sottoinsieme delle IA generali, le IA generative sono progettate per creare nuovi contenuti, che possono includere testi, immagini, video, musica o altri media.[42] Grazie all'avanzata tecnologia sviluppata negli ultimi anni, i prodotti generati da queste intelligenze risultano spesso indistinguibili da quelli creati dall'uomo, almeno a un primo sguardo superficiale. Tuttavia, algoritmi più sofisticati o altre IA possono talvolta rilevare la differenza tra un contenuto generato artificialmente e uno autentico. In questa trattazione, ci concentreremo in particolare sulle IA generative, analizzandone le applicazioni e le implicazioni etiche, soprattutto in relazione ai cosiddetti "Deepfake".

La letteratura del settore evidenzia come l'apprendimento automatico (machine learning) rappresenti un elemento chiave dell'IA moderna. Il machine learning fornisce i mezzi per analizzare grandi quantità di dati, adattarsi ai cambiamenti, fare previsioni accurate e supportare decisioni informate, tutto ciò in modo efficiente e scalabile. Questo metodo è

[42] Questo argomento viene meglio definito al considerando 134 della rettifica del 17 aprile 2024 operata dal parlamento europeo rispetto alla posizione del Parlamento europeo definita in prima lettura il 13 marzo 2024 in vista delladozione del regolamento (UE) 2024/… del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sullintelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sullintelligenza artificiale) P9_TA(2024)0138 (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))

particolarmente prezioso in quanto permette alle macchine di migliorarsi autonomamente, senza intervento umano, perfezionando le loro risposte e reazioni agli stimoli degli utenti.

Esistono due principali tipi di apprendimento automatico:[43]

Machine learning: sostanzialmente si tratta di algoritmi che apprendono autonomamente da grandi quantità di dati per svolgere specifiche mansioni. Questo avviene attraverso l'uso di reti neurali artificiali, una serie di unità interconnesse che operano in modo simile ai neuroni del cervello umano. Questi algoritmi sono in grado di riconoscere schemi, elaborare informazioni e prendere decisioni in base ai dati precedenti, simulando in qualche misura il processo biologico del pensiero.

Deep learning: Questo è un sottoinsieme del machine learning, noto anche come apprendimento profondo. Anche in questo caso, il principio fondamentale è l'uso di reti neurali, ma la struttura delle reti è molto più ampia e complessa rispetto al machine learning tradizionale. Grazie a questa complessità, il deep learning permette l'elaborazione di dati non strutturati come immagini, video o audio e abilita funzioni cognitive più avanzate, come il riconoscimento vocale o l'analisi delle immagini. Il deep learning si è rivelato fondamentale per l'avanzamento delle IA generative, in quanto consente loro di creare contenuti di altissima qualità.

Questa breve panoramica sull'intelligenza artificiale ci introduce a uno dei temi centrali di questa trattazione: i Deepfake. I Deepfake, come approfondiremo nel prossimo capitolo (§4), rappresentano un'applicazione avanzata delle IA generative, che utilizza modelli di deep learning per creare video, immagini o audio falsificati che sembrano autentici. Questi contenuti generati artificialmente sollevano questioni etiche e legali di grande rilievo, soprattutto in ambito politico, sociale e mediatico, dove la diffusione di informazioni false può avere conseguenze di vasta portata.

[43] M. Di Paolo Emilio, "*Intelligenza artificiale, machine learning e deep learning: quali sono le differenze*", (19 dicembre 2022), Innovation Post.

.4 L'evoluzione storica delle intelligenze artificiali da Turing ad oggi

Come ultimo paragrafo dedicato alla definizione di cosa sia un intelligenza artificiale, si ritiene interessante e non privo di utilità circostanziare questo fenomeno all'interno del tempo, procedendo con la trattazione di quella che possiamo considerare l'evoluzione dell'intelligenza artificiale (IA), prendendo le mosse da colui che diede i natali al concetto: Alan Turing, per giungere fino ai giorni nostri, attraversando le diverse fasi; ciascuna di queste è stata caratterizzata da progressi significativi in termini di teoria, tecnologia e applicazioni pratiche.

Ripercorriamo le tappe principali di questo percorso:[44]

Le origini, come appena sopra già riportato, sono da far risalire al matematico e logico britannico Alan Turing è considerato il pioniere dell'intelligenza artificiale. Negli anni '30 e '40, Turing gettò le basi dell'informatica moderna con il concetto di "macchina di Turing", una macchina ipotetica in grado di eseguire qualsiasi calcolo matematico seguendo una serie di istruzioni. Questo concetto influenzò profondamente lo sviluppo dei computer e, in seguito, dell'IA.

Nel 1950, Turing pubblicò il celebre articolo "Computing Machinery and Intelligence", in cui introduceva il famoso "Test di Turing", anche detto "Imitation game", così da determinare se una macchina potesse essere considerata intelligente. Secondo Turing, se un computer fosse stato in grado di ingannare un essere umano, facendogli credere che stesse interagendo con un altro essere umano, allora si sarebbe potuto dire che la macchina possedesse un'intelligenza simile a quella umana. [45]

Negli anni '50, la nascita ufficiale del campo dell'intelligenza artificiale avvenne grazie a scienziati come John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon. McCarthy coniò il termine "intelligenza artificiale". E successivamente Nel 1956, il Dartmouth Summer

[44] Cfr L. Portinale "*Intelligenza Artificiale: storia, progressi e sviluppi tra speranze e timori*" (28 gennaio 2022) Media Laws, Focus: Blockchain e Intelligenza Artificiale (3/2022), Pp 13-28

[45] A. Turing, "*Computing Machinery and Intelligence*", (ottobre 1950), Mind, 49(236): Pp. 433 – 460, in questo articolo il professor Alan Turing si pone la famosa domanda: "can machines think?", interrogativo che ha dato le fondamenta per strutturare l'esperimento chiamato "the imitation game". L'esperimento verte sulla contrapposizione di tre soggetti, un uomo, una donna e un interrogatore (il quale può essere di ambo i sessi). Il soggetto denominato interrogatore si trova in una stanza separata rispetto a quella degli altri soggetti e suo compito sarà comprendere e discriminare il fatto che uno dei soggetti con i quali sta dialogando è o meno una macchina. Lo scopo del gioco è, appunto, rispondere alla domanda e osservare se sia possibile discriminare la risposta generata da una macchina da una umana.

Research Project on Artificial Intelligence segna ufficialmente la nascita dell'IA come campo di ricerca. Durante questo evento, John McCarthy definisce l'IA come "the science and engineering of creating intelligent machines"[46], mentre Marvin Minsky la descrive come "la scienza di fare in modo che le macchine facciano cose che richiederebbero intelligenza se fatte dagli uomini"[47].

Durante questo periodo, si svilupparono i primi tentativi di creare programmi basati sulla logica e sulla risoluzione di problemi. Newell e Simon crearono uno dei primi programmi di IA chiamato Logic Theorist, capace di dimostrare teoremi matematici. Tuttavia, nonostante l'entusiasmo iniziale, i limiti computazionali dell'epoca impedirono il raggiungimento di risultati concreti.

Negli anni '70, l'IA attraversò una fase di crisi nota come il "primo inverno dell'IA" – AI Winter[48] –. Le aspettative iniziali si rivelarono troppo ottimistiche e i finanziamenti diminuirono. Tuttavia, durante questi anni si continuarono a sviluppare teorie, tra cui la nascita dei sistemi esperti, software progettati per replicare il processo decisionale umano in contesti specifici, come diagnosi mediche e risoluzione di problemi aziendali. Il sistema esperto più noto fu MYCIN, sviluppato negli anni '70 per diagnosticare malattie infettive del sangue, questo era stato sviluppato sulla base di quella che era stata l'esperienza ottenuta con lo sviluppo di DENDRAL per quello che riguardava il riconoscimento di molecole sconosciute attraverso processi di spettrografia[49].

Convenzionalmente questo periodo inizia con la pubblicazione del report del professor Sir James Lighthill[50], nel quale questi critica il sostanziale e plateale fallimento dei sistemi di IA di rispettare quelle che furono le grandiose promesse avanzate, per poi concludersi nel 1983

---

[46] McCarty et al., "*A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*" (31 agosto 1955), e riedito in AI Magazine, 2006, vol. 26(4), p. 12.

[47] Minsky e Papert, "*Perceptrons: An Introduction to Computational Geometry*", (1969), Cambridge MA,

[48] Questo termine apparse per la prima volta nel 1984, sollevato come argomento di discussione al raduno annuale dell'AAAI (american association of artificial intelligence)

[49] Vedi: R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, & J.Lederberg. "*Dendral: A Case Study of the First Expert System for Scientific Hypothesis Formation. Artificial Intelligence*" (1993) vol. 61 2, Pp. 209-261. Dendral era un progetto di intelligenza artificiale della fine degli anni 60, sviluppato allUniversità di Stanford per aiutare i chimici organici a identificare molecole sconosciute tramite lanalisi dei loro spettri di massa. Considerato il primo sistema esperto, automatizzava il processo decisionale dei chimici. Il progetto, iniziato nel 1965, coinvolse Edward Feigenbaum, Bruce G. Buchanan, Joshua Lederberg e Carl Djerassi. Scritto in Lisp, Dendral ispirò molti altri sistemi, come MYCIN e MOLGEN. Il nome "Dendral" è un acronimo di "Dendritic Algorithm".

[50] J. Lighthill, "*Artificial Intelligence: A General Survey*" (1973), Artificial Intelligence: a paper symposium. Science Research Council.

con le sovvenzioni statali per la ricerca sull'intelligenza artificiale di Alvey (un progetto di ricerca britannico) fondato per far fronte al giapponese "Fifth Generation Project"[51].

Durante gli anni '80 dello scorso secolo che nel campo delle applicazioni dell'intelligenza artificiale viene trovato nuovo interesse e un rinnovato vigore, questo avvenne tanto nelle applicazioni finanziate dagli stati, di cui possiamo ricordare i progetti menzionati nel paragrafo appena sopra, ma anche grazie all'apporto di sistemi esperti progettati, strutturati e finanziati da aziende, oltre che allo sviluppo di nuove tecniche, come il "backpropagation" per l'addestramento delle reti neurali.

In questo periodo vi è un fondamentale cambiamento del paradigma fondante per i sistemi esperti, di fatti, se prima, seguendo quanto proponeva il test di Turing, si cercava di comprendere l'intelligenza intrinseca della macchia, si è arrivati alla comprensione per la quale la macchina sarà tanto più intelligente quanti saranno i dati che questa potrà analizzare e gestire; riprendendo le parole di Pamela McCorduck: "The great lesson from the 1970s was that intelligent behavior depended very much on dealing with knowledge, sometimes quite detailed knowledge, of a domain where a given task lay".[52]

Nasce, quindi, una nuova disciplina basata sul gestire e sfruttare il sapere, in quanto questo era "l'alimento" principale e fondante delle nuove strutture delle IA. È proprio alla fine di questo decenni che vedrà per la prima volta la luce, nonostante fosse in una sua forma particolarmente primitiva il sistema "DeepBlue", che, come avremo modo di parlare successivamente, fu il primo sistema che fece comprendere quale fosse la portata delle effettive capacità di un sistema esperto, anche se nel campo degli scacchi.

Tuttavia, il campo incontrò di nuovo difficoltà dovute ai limiti della tecnologia e alle aspettative non soddisfatte, causando il secondo inverno dell'IA verso la fine degli anni '80 [53].

---

[51] B. Oakley & O. Kenneth, "*Alvery: Britains Strategic Computing Initiative*" (26 aprile 1990), MIT Press.

[52] Cfr. P. McCorduk, "*Machines Who Think: A personal inquiry into the history and prospects of aritficial intellignece*" (2004), CRC Press, p. 421.

[53] Vedi: S. J. Russell, & P. Norvig, "*Artificial intelligence: a modern approach*." (2016), Pearson. In questa pubblicazione si approfondisce quello che è l'argomento dei sistemi esperti come MYCIN e l'idea degli inverni dell'IA, con particolare riferimento ai capitoli che trattano della storia dell'AI, dei limiti tecnologici che ancora avevano un forte peso durante quegli anni e, quindi, sulle sfide che gli sviluppatori dovevano affrontare in quegli anni.

Negli anni ‘90, l’IA tornò a farsi strada grazie al crescente interesse per il machine learning e il potenziamento delle capacità computazionali dei computer. Il machine learning, un ramo dell’IA, si concentra sullo sviluppo di algoritmi in grado di apprendere automaticamente dai dati senza essere esplicitamente programmati. L’uso di grandi dataset e nuovi approcci algoritmici contribuì a risolvere problemi reali.

Uno dei momenti storici di questo periodo fu la vittoria di Deep Blue, un computer sviluppato da IBM, che nel 1997 sconfisse Garry Kasparov, campione del mondo di scacchi, in una partita ufficiale. Questo evento dimostrò il potenziale dell’IA nel superare le capacità umane in compiti specifici e molto complessi[54].

A partire dagli anni 2000, l’IA ha registrato una crescita esponenziale grazie a tre fattori principali: la disponibilità di enormi quantità di dati (big data), l’aumento della potenza computazionale e lo sviluppo di nuovi algoritmi, in particolare il deep learning.

Il deep learning è una sottocategoria del machine learning basata su reti neurali profonde, ispirate al funzionamento del cervello umano. Queste reti possono elaborare enormi quantità di dati e riconoscere pattern complessi. Grazie al deep learning, l’IA è stata in grado di raggiungere risultati straordinari in compiti come il riconoscimento delle immagini, la traduzione automatica e il riconoscimento vocale.

Nel 2012, un importante passo avanti fu compiuto quando AlexNet, un modello di deep learning, vinse il concorso di riconoscimento delle immagini ImageNet, con un margine significativo rispetto ai concorrenti. Questo segnò l’inizio dell’era moderna dell’IA, in cui le reti neurali profonde sono diventate una tecnologia dominante.

Negli ultimi anni, l’IA ha continuato a evolversi, spostando il suo focus verso modelli sempre più avanzati e capaci, come le IA generative. Questi modelli, tra cui le GAN (Generative Adversarial Networks) e gli autoencoder, sono in grado di creare nuovi contenuti, come immagini, video, audio e testi, spesso indistinguibili da quelli creati dagli esseri umani[55].

[54] F. H. Hsu, “*Behind Deep Blue: Building the computer that defeated the world chess champion*.” (2002), Princeton University Press.

[55] Y. Bengio, I. Goodfellow & A. Courville, “*Deep learning*” (18 novembre 2016). Vol. 1, MIT press, Cambridge.

Un esempio notevole di IA generativa è GPT (Generative Pre-trained Transformer), sviluppato da OpenAI, che ha portato a rivoluzioni nel campo della generazione automatica di testi. I sistemi basati su GPT possono rispondere a domande, scrivere articoli o persino generare codice, dimostrando capacità di linguaggio simili a quelle umane.

I Deepfake, che sono un'applicazione del deep learning, sono diventati una delle manifestazioni più controverse dell'IA generativa. Queste tecnologie possono creare video e audio falsificati con un livello di realismo tale da sollevare preoccupazioni etiche e legali, specialmente in ambito politico e sociale[56] [57].

Guardando al futuro, si discute del possibile sviluppo della IA generale (AGI, Artificial General Intelligence), ovvero una forma di intelligenza artificiale capace di apprendere e risolvere problemi in qualsiasi dominio, con capacità cognitive simili o superiori a quelle umane. Sebbene oggi l'AGI sia ancora lontana, molti ricercatori stanno lavorando per avvicinarsi a questo obiettivo[58].

In conclusione, possiamo dire che l'intelligenza artificiale ha compiuto un lungo viaggio, dai concetti teorici di Turing fino alle applicazioni pratiche dei giorni nostri. Ogni fase dell'evoluzione dell'IA ha portato nuove sfide e opportunità, contribuendo a ridefinire il nostro rapporto con la tecnologia. Sebbene l'IA abbia ancora molta strada da fare, le sue capacità attuali stanno già trasformando il mondo in modi profondi e inaspettati.

---

[56] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, "*Generative Adversarial Nets. Advances in Neural Information Processing Systems*" (2014). Pp. 2672 - 2680.

[57] Y. Mirsky & W. Lee, *"The creation and detection of deepfakes: A survey*" (2021). ACM computing surveys (CSUR), vol 54(1). Pp. 1 - 41.

[58] Uno di questi ricercatori è sicuramente il filosofo Nick Bostrom, il quale nel libro *"Superintelligence: Paths, Dangers, Strategies"* edito nel 2014, tratta dei vari percorsi attraverso i quali si potrebbe giungere alla realizzazione in un intelligenza artificiale generale e dei vari rischi potenziali che potrebbero sorgere dalla realizzazione di un sistema tanto potente, anche al fine di mitigare questi pericoli ed evitare che un AGI possa assumere un controllo incontrollato di quella che è la realtà umana nella quale noi tutti viviamo; nonostante queste preoccupazioni, Bostorm afferma anche che: nonostante il velocissimo ritmo di sviluppo dell'AI siamo ancora lontani da quello che può essere considerato un sistema di questo genere.

4 Introduzione ai Deepfake

.1 Definizione e precisazioni sulla tecnologia alla base dei Deepfake

In questo sottocapitolo si intende offrire una panoramica sul fenomeno dei Deepfake; il termine composto dall'unione dei vocaboli inglesi "deep learning" (apprendimento profondo) e "fake" (falso), che esprime fin dall'origine l'idea di sfruttare sofisticati algoritmi di intelligenza artificiale per generare contenuti multimediali alterati o completamente costruiti.[59]

La definizione fornita dal Garante della Privacy chiarisce con precisione come i Deepfake siano:

> *"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."*[60]

Tale descrizione pone l'accento sulle potenzialità di un fenomeno in grado di coinvolgere non soltanto l'immagine statica, ma anche il campo del video e dell'audio, con la possibilità di riprodurre o trasformare sembianze e voci di individui reali o immaginari, creando talvolta la percezione di una realtà che in effetti non esiste.

La genesi storica dei Deepfake, nell'accezione che ha poi trovato eco sui media, è comunemente fatta risalire al 2017, quando su Reddit comparve un utente capace di realizzare video pornografici contraffatti[61]: costui sostituiva i volti degli attori originali con quelli di figure più o meno note, ricorrendo a tecniche oggi classificate come Face Swapping.

Pur essendo ancora lontani dalla sofisticazione dei moderni Deepfake, questi esperimenti rappresentarono i primi esempi popolari di manipolazione digitale basata su

---

[59] Questo fenomeno trova i natali nel 2017, anno in cui Ict Security Magazine, una delle riviste di cyber security italiane di maggior riguardo, attesta come anno di prima applicazione di intelligenze artificiale a questi fini. Tra le altre fonti possiamo sicuramente puntualizzare diverse pubblicazioni del MIT Sloan, ovvero, ancora: S. Cole, "*We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*", (*24 gennaio 2018*) *Vice*.

[60] "*Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*", (28 dicembre 20) Garante della privacy <www.garanteprivacy.it >

[61] M. Cazzaniga, "*Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfake*, Media Laws", (14 giugno 2023) Medialaws <Medialaws.eu>. Cfr: V. Giacometti, "*Il deepfake: il sottile confine fra fantasia, gioco e reato*" (13 dicembre 2023) Forensicnews <forensicnews.it>.

tecniche di deep learning.[62] Per comprendere però le radici di questa tecnologia, occorre tornare a qualche anno prima, quando Ian Goodfellow, ricercatore di machine learning, introdusse per la prima volta l'architettura detta GAN (Generative Adversarial Network).[63] Tale sistema prevede la competizione tra due reti neurali: la prima genera contenuti che si mescolano a contenuti reali, mentre la seconda cerca di distinguerli in base a caratteristiche e pattern. Il processo prosegue finché la seconda rete non è più in grado di riconoscere la differenza tra ciò che è reale e ciò che è fittizio. Questo "gioco" di sfida continua risulta essere la base su cui si fondano i moderni sistemi di Deepfake, poiché permette di addestrare il software al punto di generare manipolazioni sempre più verosimili e difficili da smascherare.

Inizialmente, si applicavano architetture GAN soprattutto per la creazione o l'alterazione di immagini statiche, come fotografie e ritratti generati dal nulla. In un breve giro di tempo, l'ambito si è esteso al settore video e all'audio, aprendo la strada alla realizzazione di prodotti multimediali di altissima qualità. Per esempio, è possibile generare voci sintetiche che imitano con straordinaria precisione il timbro e la cadenza di personaggi pubblici, cantanti o amici e familiari (§5). Con analoghi metodi, si riescono a replicare i movimenti facciali e corporei, ottenendo filmati in cui la persona ripresa compie gesti o pronuncia frasi mai detti o compiuti nella realtà. È utile ricordare che il grado di realismo ottenibile è strettamente legato alla mole e alla qualità dei dati con cui si addestrano gli algoritmi: più estesa e curata è la base di immagini, video o audio, maggiore sarà la verosimiglianza del risultato finale.[64]

Un esempio di utilizzo in apparenza più innocuo di manipolazioni digitali è rappresentato dai filtri utilizzati in molte app o piattaforme social, che permettono di trasformare l'immagine catturata dalla fotocamera, correggendo imperfezioni, cambiando i tratti somatici o rendendo la persona somigliante a creature immaginarie. Questi filtri, tuttavia, non sono paragonabili ai Deepfake più complessi, dal momento che la finalità è generalmente di intrattenimento e di editing leggero, e l'alterazione resta spesso immediatamente riconoscibile. Ciononostante, l'uso di tecniche sempre più sofisticate mostra come si possa partire da un livello ludico fino a giungere a manipolazioni gravemente lesive, soprattutto

[62] Y. Mirsky, & W. Lee, "*The Creation and Detection of Deepfakes: A Survey.*" (2021). ACM Computing Surveys (CSUR), vol. 54(1), Pp. 1 - 4.

[63] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, "*Generative Adversarial Nets. Advances in Neural Information Processing Systems*" (2014). Pp. 2672-2680.

[64] A. Shaji George e A. S.Hovan George "*Deepfakes: The Evolution of Hyper realistic Media Manipulation*", (dicembre 2023), in Partners Universal Innovative Research Publication (PUIRP), Vol. 1(2) Pp. 58 - 74.

quando coinvolgono persone reali all'interno di contesti delicati, come quello sessuale, politico o economico.

Quando parliamo di Deepfake, inoltre, non si può trascurare il ruolo dell'evoluzione tecnologica che ha portato a un'elevata disponibilità di risorse di calcolo a costi relativamente bassi.[65] Processi di elaborazione che, fino a pochi anni fa, avrebbero richiesto supercomputer esclusivi di grandi aziende, sono ora alla portata di singoli utenti, startup o piccole realtà imprenditoriali. Questa democratizzazione degli strumenti di intelligenza artificiale ha permesso alla tecnologia dei Deepfake di diffondersi in maniera capillare, con risvolti che investono contesti leciti e illeciti.

Da una parte, alcune produzioni cinematografiche sfruttano i Deepfake o le tecniche GAN per generare effetti speciali di qualità e a costi inferiori rispetto ai metodi tradizionali. Dall'altra, esiste il rischio che tali strumenti vengano sfruttati a scopo malevolo, per diffamare personaggi pubblici, influenzare elezioni, truffare aziende o ricattare singoli individui (§5.4).

Le applicazioni potenzialmente negative sono, infatti, numerose e preoccupanti. Particolarmente diffusa è la creazione di contenuti pornografici non consenzienti, che coinvolgono il volto di celebrità o persone comuni, montato sui corpi di attrici e attori porno. Un caso emblematico fu proprio quello con cui nacque il termine Deepfake: l'utente Reddit citato in precedenza sostituiva i volti di attrici celebri a quelli di interpreti di film hard, generando video apparentemente autentici. Simili pratiche provocano gravi danni all'immagine e alla dignità delle vittime, oltre a sollevare questioni di natura giuridica in merito al consenso, alla tutela dei diritti della personalità e alla protezione dei dati personali.

Un altro esempio di uso malevolo, divenuto tristemente celebre, è la manipolazione di discorsi pubblici o messaggi di leader politici per diffondere disinformazione, alterare l'opinione pubblica o semplificare la propaganda ostile. Nel corso del conflitto in Ucraina, si sono diffusi brevi clip in cui il presidente Volodymyr Zelens'kyj sembrava pronunciare discorsi che non aveva mai fatto, con l'intento di generare confusione tra i cittadini, seminare dubbi o dirottare il consenso (§5.3). Questi tentativi, sebbene spesso riconoscibili, danno un'idea del potere potenziale che i Deepfake possono esercitare qualora la qualità della manipolazione diventi tale da essere indistinguibile a occhio nudo. A ben vedere, anche contenuti contraffatti

[65] M. Brundage, S. Avin, J. clark et alt. "*The Malicious Use of Artificial Intelligence : Forecasting, Prevention, and Mitigation*" (febbraio 2018)

di minor accuratezza possono sortire effetti notevoli su un pubblico non adeguatamente informato o poco abituato a verificare la fonte.

La portata del rischio ha spinto diverse piattaforme social, a partire dai primi eclatanti casi, a introdurre divieti o restrizioni parziali alla diffusione di contenuti manipolati. In alcuni casi, i termini di servizio di siti come Facebook, Twitter o Reddit sono stati aggiornati per includere esplicitamente il bando di Deepfake che implicassero la violazione dei diritti altrui (§ .2). Tuttavia, l'implementazione di simili regole e la capacità di riconoscere con certezza contenuti creati con l'ausilio di GAN rimangono sfide notevoli. Non sempre, infatti, i sistemi di rilevamento e segnalazione automatica riescono a stare al passo con la rapidità di evoluzione degli algoritmi di generazione.

Parallelamente, i legislatori di diversi Paesi si interrogano sulla necessità di regolamentare l'uso dei Deepfake. Alcune iniziative legislative sono state messe in campo, per esempio negli Stati Uniti, dove alcuni Stati hanno proposto leggi o regolamenti specifici per criminalizzare l'impiego di Deepfake a fini di disinformazione o diffusione di contenuti pornografici non autorizzati. Allo stesso modo, l'Unione Europea ha inserito, nel cosiddetto Ai-act (reg. 2024/1689), alcune disposizioni volte a disciplinare le applicazioni di intelligenza artificiale, includendo brevi riferimenti ai rischi correlati ai Deepfake. Si tratta spesso di normative in divenire, che cercano di contemperare la necessità di non bloccare il progresso tecnologico con quella di salvaguardare la tutela dei diritti fondamentali, quali la dignità, la privacy e la reputazione delle persone.

La natura stessa del Deepfake – che per definizione esalta la capacità di creare illusioni iperrealistiche – costituisce una sfida per il sistema giuridico e per la società nel suo complesso. In passato, si tendeva a dare per scontato che un video o un file audio fossero prove più o meno incontrovertibili di un certo evento, in quanto più difficili da manipolare rispetto a fotografie o documenti cartacei. Oggi questa presunzione di autenticità non è più sostenibile. Diventa perciò essenziale sviluppare meccanismi in grado di verificare l'affidabilità delle fonti e l'identità di chi produce o condivide un contenuto multimediale. Alcune soluzioni tecnologiche stanno emergendo, come i watermark digitali o le firme crittografiche, che possono attestare l'originalità di un video, ma permane la complessità di un quadro in cui non tutti i creatori di contenuti desiderano o possono adottare procedure simili.

La questione assume notevole importanza anche sul piano della responsabilità. Se un Deepfake diffamatorio, sessualmente esplicito o gravemente offensivo viene diffuso senza autorizzazione, chi ne risponde? L'autore materiale del contenuto? L'utente che lo ha caricato su una piattaforma? Gli amministratori del servizio online che ne permettono l'hosting? Le risposte, al momento, non sono univoche e variano a seconda delle normative di ciascun Paese, nonché delle condizioni d'uso delle piattaforme. Alcune giurisdizioni si muovono verso un modello che impone una certa responsabilità in capo agli operatori del servizio, se non eliminano con prontezza i contenuti nocivi o se non predispongono strumenti adeguati di controllo. Altri contesti giuridici risultano invece più permissivi, privilegiando la libertà d'espressione e la neutralità della rete.

Un ulteriore problema consiste nelle modalità di riconoscimento e smascheramento dei Deepfake. La qualità di alcuni contenuti generati dalle intelligenze artificiali più avanzate raggiunge livelli così elevati che perfino un esame attento potrebbe non bastare a svelarne la natura artefatta.[66] Alcuni ricercatori suggeriscono di concentrarsi su particolari difetti quali il movimento degli occhi o la gestione delle ombre, ma le reti neurali sono in grado di "imparare" anche da questi errori, riducendoli in ulteriori generazioni. Pertanto, i metodi di detection devono rinnovarsi continuamente e risulta cruciale l'investimento in ricerca e sviluppo di software di contrasto, pena il rischio di un'escalation di contenuti falsi pressoché perfetti.

In considerazione di quanto esposto, si comprende quanto sia indispensabile un approccio che, oltre alla sfera giuridica e tecnologica, coinvolga anche la sfera educativa. È essenziale, infatti, formare gli utenti e i cittadini affinché imparino a porre un certo grado di scetticismo nei confronti di ciò che vedono e sentono in rete. Ciò non significa ridurre la fiducia collettiva nella circolazione delle informazioni, ma piuttosto sviluppare un pensiero critico, l'abitudine a verificare le fonti e a ricorrere a canali autorevoli prima di diffondere o credere a un contenuto particolarmente sensazionale. In parallelo, le stesse aziende che sviluppano strumenti di deep learning dovrebbero adottare prassi di responsabilità sociale, definendo limiti all'uso della tecnologia e promuovendo linee guida utili a minimizzare il rischio di abusi.

Per concludere, i Deepfake presentano opportunità e rischi ineguali: da un lato, l'industria del cinema, della pubblicità e dell'intrattenimento può beneficiare di nuove modalità espressive, più economiche e creative di quelle offerte dai software tradizionali; dall'altro, le

[66] Ivi nota 64

applicazioni malevole rischiano di crescere in maniera incontrollata, influenzando l'opinione pubblica e violando i diritti dei singoli. Le politiche di contrasto e i tentativi di regolamentazione rappresentano soltanto un inizio, mentre le problematiche più complesse inerenti alla prova dell'autenticità, alla privacy, all'onore e alla reputazione rimangono aperte. La società dell'informazione richiede dunque un approccio consapevole, uno sguardo vigile rispetto alle notizie e ai contenuti che circolano sui social network e sui media tradizionali, senza dimenticare che l'evoluzione tecnologica rende probabile l'ulteriore perfezionamento degli strumenti di generazione dei Deepfake.

.2 Tecniche e procedimenti utilizzati nella creazione dei Deepfake

Esistono diversi approcci per la creazione di Deepfake e, come già anticipato, le reti neurali GAN (Generative Adversarial Network) rappresentano l'innovazione che ha contribuito in maniera decisiva all'affermazione di questo fenomeno, affiancate dai sistemi denominati Autoencoder.

Poiché si tratta di tecnologie avanzate che richiedono una notevole quantità di risorse computazionali e di dati per funzionare al meglio, la loro rapida diffusione è stata favorita sia dall'evoluzione dell'hardware sia dalla disponibilità di database pubblicamente accessibili; in particolare, 100K-faces e iFakeFaceDB hanno rappresentato due risorse fondamentali per addestrare i modelli di rete neurale, mettendo a disposizione una grande quantità di immagini di volti realistici. Questi dataset hanno consentito alle grandi aziende, ai ricercatori e anche ai semplici appassionati di sperimentare e creare immagini e video estremamente convincenti, a volte talmente realistici da rendere difficile, anche per un occhio allenato, distinguere le creazioni artificiali da quelle autentiche.

L'accesso a questi database e il conseguente miglioramento costante degli algoritmi hanno portato effetti positivi, per esempio nello sviluppo di modelli 3D sempre più accurati, utilizzati in settori come l'animazione o la progettazione industriale. Tuttavia, non sono mancate le conseguenze negative: i Deepfake, sin dai loro esordi, come si è già discusso, si sono rivelati strumenti potenzialmente pericolosi, in grado di generare contenuti ingannevoli e di diffondere disinformazione. Questi rischi verranno approfonditi più avanti, poiché in questa sede è opportuno concentrarsi prima di tutto su come funzionano le varie tipologie di reti neurali e sulle differenze architetturali che consentono di ottenere risultati così sorprendenti.

Le reti neurali GAN, una volta completata la fase di apprendimento, creano al loro interno una competizione tra due componenti distinte: il generatore e il discriminatore. Il primo ha il compito di produrre campioni artificiali sufficientemente persuasivi da ingannare il secondo, che invece cerca di distinguere ciò che è reale da ciò che è creato dal generatore. Nel corso degli iter di addestramento, i due "giocatori" continuano a sfidarsi, migliorando reciprocamente le proprie capacità: mentre il generatore affina la propria abilità di simulare con precisione i tratti che compongono l'immagine o il video, il discriminatore diventa sempre più abile nell'individuare le incongruenze o le imperfezioni che tradiscono la natura sintetica del contenuto. L'equilibrio finale si raggiunge quando il discriminatore non riesce più a distinguere il falso dal vero, segno che il generatore è diventato tanto sofisticato da produrre un risultato quasi identico a un contenuto autentico. Questo meccanismo a "gioco competitivo" è stato paragonato all'eterna sfida tra falsari e forze dell'ordine: più i falsari migliorano le loro tecniche, più la polizia sviluppa strumenti per identificarne i prodotti, e così via in un circolo di reciproco incremento della qualità delle strategie.

Le GAN possono adottare diverse strutture: alcune fanno uso del modello minimax, in cui il generatore si sforza di massimizzare la possibilità di ingannare il discriminatore, mentre quest'ultimo cerca di minimizzare tale possibilità; altre utilizzano tecniche non saturanti, che influiscono sul metodo di aggiornamento dei pesi interni al generatore. Le differenze architetturali contribuiscono a generare risultati di vario tipo, a seconda della complessità del contenuto che si desidera creare e del livello di controllo che si vuole ottenere sui singoli aspetti dell'immagine.[67]

Anche gli Autoencoder hanno un ruolo cruciale nella produzione dei Deepfake, benché la loro logica di funzionamento si differenzi da quella delle GAN. Un Autoencoder impara a comprimere un'immagine (o qualsiasi altra tipologia di dati) in un insieme di informazioni latenti più ridotto, conservandone le caratteristiche fondamentali in uno spazio dimensionale inferiore rispetto a quello originario. Successivamente, tramite una fase di decodifica, ricostruisce l'immagine di partenza. Questo sistema è prezioso per la creazione dei Deepfake perché permette di estrarre la "firma" di un dato volto – i tratti distintivi che lo caratterizzano – in modo da poterli ricombinare o sostituire. Per produrre effettivamente un Deepfake, generalmente si utilizzano due Autoencoder: il primo viene addestrato a riconoscere la caratteristica della sorgente, cioè il volto da cui si vuole estrarre i tratti, mentre il secondo si specializza sul volto bersaglio, cioè quello in cui tali tratti verranno innestati. Tramite il passaggio nello spazio latente, si possono ottenere manipolazioni che non si limitano a una semplice sovrapposizione, ma modificano attivamente le fattezze del bersaglio in maniera coerente e realistica.

Queste tecnologie hanno visto il loro primo exploit nel 2017, con la presentazione ufficiale della prima GAN realizzata da Ian Goodfellow. Da allora, si è assistito a una sequenza di perfezionamenti tecnici che hanno reso i Deepfake sempre più complessi da smascherare. Le implementazioni successive hanno generato nuove varianti e architetture che, a loro volta,

[67] I. Goodfellow et al., "*Generative Adversarial Networks*", (novembre 2020) ACM Digital Library, communication of the ACM. Vol. 63(11)

hanno spinto l'evoluzione del campo e innalzato il livello di sfida per i ricercatori che lavorano ai sistemi di rilevamento del falso.

Uno di questi sviluppi riguarda le AttGAN[68], una tipologia di rete in grado di manipolare con estrema precisione gli attributi di un soggetto bersaglio. Questo consente di intervenire con maggiore accuratezza e di restituire risultati di alta qualità, sia nell'ambito del faceswapping sia in quello di altre operazioni, come la regressione dell'età (o antiaging), che permette di ringiovanire artificialmente un volto.

Le AttGAN si distinguono dalle versioni precedenti per la capacità di "capire" in modo più completo gli attributi da modificare, perfezionando il risultato finale e garantendo un realismo incredibilmente elevato.

Un altro passo in avanti si deve ai ricercatori di Nvidia, che hanno introdotto le StyleGAN, un'architettura che ha rivoluzionato il modo di interpretare l'immagine. Il passaggio da una concezione basata sui singoli pixel a una visione vettoriale dell'immagine ha consentito di ottenere risultati particolarmente nitidi, talmente convincenti da richiedere l'aggiunta di rumore fotografico artificiale per rendere l'immagine meno perfetta e, paradossalmente, più realistica. Grazie a StyleGAN, la creazione di volti sintetici ha superato un livello di dettaglio mai raggiunto prima, al punto che il confine tra realtà e falsificazione appare sempre più sottile.[69]

I ricercatori autori del modello STGAN hanno adottato un approccio che fonde elementi di Nvidia con le loro idee, introducendo collegamenti di salto tra i neuroni nelle reti autoencoder e combinando questi concetti con una struttura GAN. Questo permette di aumentare ulteriormente la qualità dell'immagine a scapito, in parte, della flessibilità con cui si possono modificare certi attributi. Più nello specifico, STGAN consente di lavorare su vettori etichettati, così da controllare con notevole precisione le modifiche di singoli tratti, sebbene con qualche limitazione rispetto ad altre architetture più "libere".[70]

Una successiva innovazione è giunta con StarGANv2[71], modello che permette di operare non su un singolo dominio d'immagini ma su più domini contemporaneamente. In altre parole, anziché "specializzarsi" soltanto su un tipo di volto o su una singola categoria di immagini, questa rete riesce ad abbracciare gruppi di immagini contraddistinte da caratteristiche diverse (stili, pose, e così via). Di conseguenza, si ampliando le potenzialità di generazione, offrendo la possibilità di creare un maggior numero di risultati partendo da un solo modello e dando origine a trasformazioni anche molto diverse tra loro. Questo aspetto si rivela utile in diversi contesti applicativi, dal design di personaggi virtuali per i videogiochi, alla creazione di scenari più vari nei film o nelle campagne pubblicitarie.

---

[68] H. Zhenliang et al., "*AttGAN: Facial Attribute Editing by Only Changing What You Want*", (11 novembre 2019) IEEE transaction on image progression. Vol. 28.

[69] Tero Karras et al., "*A Style-Based Generator Architecture for Generative Adversarial Networks"*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, Pp. 4401 - 4410

[70] M. Liu et al., "*STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing*" (2019), IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Pp. 3673-3682

[71] Y. Choi et al., "*StarGAN v2: Diverse Image Synthesis for Multiple Domains*" (26 aprile 2020) Arxiv.

Tra le varianti progettate per lavorare su più domini si annovera anche la CycleGAN[72], che concentra la propria azione sul trasferimento di caratteristiche facciali da un individuo a un altro, con particolare riferimento al faceswapping. I contenuti generati mediante CycleGAN risultano spesso di alta qualità, soprattutto per la capacità di mantenere una coerenza interna all'immagine nel passare da un volto all'altro. Un ulteriore tassello all'interno di questo panorama in evoluzione costante è costituito dalle RSGAN[73] (region-separative generative adversarial network), che si focalizzano in modo selettivo sulla manipolazione di aree specifiche, come il volto o i capelli, lasciando inalterato il resto dell'immagine. È un'implementazione particolarmente utile se si vuole intervenire soltanto su determinati elementi del frame, evitando le possibili distorsioni causate da un cambiamento involontario di regioni contigue.

Diversa dalle precedenti è la LipGAN,[74] che si concentra sull'allineamento tra un contenuto audio preesistente e la sincronizzazione del labiale di un volto nel video. La finalità è quella di produrre i cosiddetti video Lipsic, in cui l'audio in sottofondo è perfettamente sincronizzato con il movimento delle labbra di un soggetto che, nella versione originale, stava dicendo tutt'altro o magari non stava nemmeno parlando. Questo tipo di Deepfake si presta particolarmente a fini di parodia o propaganda, poiché trasmette l'illusione che una persona stia pronunciando determinate parole, mentre in realtà si tratta di un abbinamento arbitrario tra audio e video.

Tutte queste architetture vengono implementate dalle aziende con protocolli proprietari, ma in rete si trovano anche numerose risorse open-source. Chiunque, con le giuste conoscenze e con una macchina sufficientemente potente, può cimentarsi nella creazione di un Deepfake a partire da software come FaceApp, Reface, Deepbrain, DeepFaceLab o Deepfakes Web. Sebbene gli strumenti open-source offrano spesso meno possibilità rispetto a quelli proprietari – e, ovviamente, le prestazioni possono dipendere molto dalla potenza di calcolo disponibile – la facilità con cui è possibile fruire di queste risorse ha contribuito alla rapida diffusione del fenomeno e alla moltiplicazione di contenuti potenzialmente manipolati. Il capitolo delle conseguenze sociali, etiche e legali legate a questa accessibilità sarà affrontato successivamente, poiché merita un'analisi approfondita su come l'opinione pubblica e le istituzioni dovranno regolamentare tali tecnologie, specialmente quando i loro utilizzi sfuggono agli scopi meramente ludici o creativi.

Prima di procedere con la realizzazione effettiva di un Deepfake, occorre alimentare e addestrare il sistema. La prima fase, quella della raccolta dati, è imprescindibile: per creare un contenuto credibile, occorre disporre di un numero rilevante di immagini o registrazioni che rappresentino la persona da riprodurre. Soltanto così la rete può "imparare" ogni dettaglio delle espressioni facciali, della prossemica e delle sfumature della voce di quel singolo

---

[72] Y. J. Zhu et al., "*Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks*" (2017), Proceedings of the IEEE International Conference on Computer Vision (ICCV), Pp. 2223 - 2232.

[73] R. Natsume, T. Yatagawa & S. Morishima, "*RSGAN: Face Swapping and Editing using Face and Hair Representation in Latent Spaces*" (18 aprile 2018) Arxiv.

[74] K. R. Prajwal et al, "*Towards Automatic Face-to-Face Translation*" (October 2019), Proceedings of the 27th ACM International Conference on Multimedia. Pp. 1428 - 1436.

individuo. Più ampio è il dataset a disposizione, più accurato e difficilmente smascherabile risulterà il prodotto finale, sebbene entrare in possesso di un dataset così voluminoso sollevi questioni di privacy e diritti d'immagine. In taluni casi, il materiale è pubblicamente accessibile (ad esempio dai social network), oppure, nello scenario peggiore, potrebbe essere stato sottratto in maniera illecita.

Segue quindi la fase di apprendimento automatico: l'algoritmo viene "nutrito" con questi dati e addestrato a riprodurre la figura, il volto o la voce del soggetto. Man mano che il modello si allena e affina la propria comprensione di ogni dettaglio rilevante, fino a diventare capace di generare nuovi contenuti che si dimostrano coerenti con le caratteristiche originarie dell'individuo. Quando la rete ha completato l'apprendimento, è pronta per la fase di produzione del Deepfake vero e proprio. In questa fase, l'utente può specificare, per esempio, un testo da far pronunciare virtualmente alla persona simulata, oppure un'espressione facciale o un movimento particolare, che nella realtà non sono mai stati espressi. Se l'addestramento è stato adeguato e si dispone di sufficienti risorse di calcolo, il risultato può rivelarsi sorprendentemente verosimile. Non di rado, infatti, persino un osservatore esperto potrebbe avere difficoltà a individuare a colpo d'occhio gli indizi tipici di un contenuto manipolato.

In conclusione, i Deepfake sono il prodotto di un processo articolato che utilizza reti neurali profonde, dataset di addestramento estesi e un'accurata sintonizzazione degli algoritmi interni. Il loro realismo cresce in proporzione ai dati e alle risorse computazionali disponibili, aprendo orizzonti estremamente ampi nelle applicazioni artistiche, nell'intrattenimento, nella ricerca e, purtroppo, anche in attività poco etiche o criminali. Le reti neurali GAN costituiscono il motore principale di questa rivoluzione, mentre gli Autoencoder, con la loro capacità di estrarre e ricomporre le caratteristiche essenziali di un volto o di un contenuto audio, offrono un ulteriore strumento di flessibilità. Con l'evoluzione di architetture come AttGAN, StyleGAN, STGAN, StarGANv2, CycleGAN, RSGAN e LipGAN, le possibilità di manipolazione si sono moltiplicate, e la qualità dei risultati è aumentata esponenzialmente. Nonostante molti utenti si avvicinino a queste tecnologie per divertirsi, esistono scenari più complessi, in cui i Deepfake possono divenire veicolo di truffe, propaganda e disinformazione. Il dibattito sulla loro regolamentazione, come su come bilanciare libertà di ricerca, diritto alla privacy e tutela della verità, è appena iniziato. E sarà compito dei legislatori, degli studiosi e di tutti gli operatori del settore individuare soluzioni che salvaguardino l'enorme potenziale positivo di queste tecnologie, limitando i rischi connessi alle loro versioni più insidiose e difficili da controllare.

5 Casi pratici

.1 Introduzione ai casi nei quali l'IA è stata utilizzata per condizionare l'opinione pubblica

Visto e analizzato quanto scritto nei passati capitoli, abbiamo compreso come si formi, come si differenzi e come viene legislata questa vasta frontiera alla quale ci stiamo affacciando.

Naturalmente nasce quella che è una polemica riguardo l'utilità, ovvero sulla necessità di tutto quanto appena scritto, a ciò si vuole rispondere in questo capitolo, prendendo in esame casistiche che si sono viste succedere in questi ultimi anni, dando, quindi anche un quadro fenomenico a quella che altrimenti potrebbe sembrare una trattazione sterile e fine a sé stessa.

Negli ultimi anni, l'intelligenza artificiale ha avuto un impatto significativo su molti settori, compresa la comunicazione e la formazione dell'opinione pubblica. L'IA è stata utilizzata in vari contesti per influenzare le percezioni delle persone, spesso in modi sottili ma potenti. Uno degli utilizzi più controversi riguarda le campagne di disinformazione e manipolazione dei social media, dove algoritmi avanzati analizzano il comportamento degli utenti e generano contenuti personalizzati per rafforzare o cambiare opinioni su argomenti politici, sociali o commerciali.

Tra le principali casistiche di utilizzo dell'IA per condizionare l'opinione pubblica, troviamo l'uso di Deepfake per creare video falsi di figure pubbliche, la diffusione automatizzata di notizie false attraverso bot sui social media, e l'analisi dei big data per creare profili psicografici dettagliati degli utenti. Queste tecniche vengono impiegate per manipolare elezioni, influenzare il comportamento dei consumatori, e persino polarizzare la società. Questo fenomeno solleva interrogativi etici e richiede l'adozione di strategie per contrastare l'uso improprio dell'IA nella manipolazione dell'informazione.

.2 L'intelligenza artificiale e le ingerenze nelle elezioni

Argomento di questo paragrafo riguarderà l'ingerenza di forze esterne nelle elezioni di uno Stato. Questo, ad oggi, può avvenire sia attraverso mezzi tradizionali che tramite tecnologie avanzate ed è una questione che solleva profonde preoccupazioni sul piano giuridico e politico. Il diritto internazionale, fondato sul principio della sovranità statale, garantisce a ciascun Paese il diritto di gestire autonomamente i propri processi elettorali, senza interferenze di attori esterni. Tuttavia, con l'evoluzione delle tecnologie digitali, questa protezione si trova ad affrontare nuove sfide, poiché le forze esterne possono ricorrere a strumenti sofisticati come l'intelligenza artificiale (IA) per influenzare l'opinione pubblica, distorcere il dibattito democratico e, in ultima analisi, alterare i risultati elettorali, anche a causa di questi strumenti e tecniche, oggi, la difesa dei principi di cui a breve parleremo sembra ogni giorno più debole e manchevole.

Il principio che primo tra tutti dobbiamo considerare riguarda la non ingerenza negli affari interni, ciò è sancito dall'articolo 2(7) della Carta delle Nazioni Unite, questo vieta a qualsiasi Stato o attore esterno di interferire nelle questioni interne di un altro Stato, incluse le elezioni. Tuttavia, la diffusione di campagne di disinformazione o la manipolazione dei media, potenziate dall'IA, rappresentano una forma subdola di ingerenza, difficile da rilevare e sanzionare, ma comunque in violazione di questo principio. Queste azione sono pienamente contro il diritto all'autodeterminazione, sancito da trattati come il Patto Internazionale sui Diritti Civili e Politici (ICCPR), garantisce ai popoli il diritto di scegliere liberamente i propri leader e rappresentanti.[75] Quando tecnologie di IA vengono utilizzate per alterare la percezione degli elettori attraverso contenuti manipolati, come Deepfake o campagne di microtargeting politico basate su big data, si compromette la genuina espressione della volontà popolare, violando il principio di autodeterminazione.

Come verrà trattato in un futuro capitolo (§ 7), il diritto internazionale ha iniziato a tenere conto della minaccia rappresentata dagli attacchi informatici, che possono interferire con

[75] Il principio di autodeterminazione dei popoli è sancito dall'articolo 1, par. 2, 55 e 76 della Carta delle Nazioni Unite. Questo principio è divenuto diritto umano, formalmente riconosciuto a tutti i popoli, in virtù dell'identico articolo 1 dei due Patti internazionali sui diritti umani del l966: "1. Tutti i popoli hanno il diritto di autodeterminazione. In virtù di questo diritto, essi decidono liberamente del loro statuto politico e perseguono liberamente il loro sviluppo economico, sociale e culturale. (...) 3. Gli Stati parti del presente Patto, (...), debbono promuovere l'attuazione del diritto di autodeterminazione dei popoli e rispettare tale diritto, in conformità alle disposizioni dello statuto delle Nazioni Unite".

il processo elettorale. La Tallinn Manual[76], pur non essendo vincolante, offre un quadro interpretativo secondo cui gli attacchi cibernetici che minano l'integrità delle elezioni costituiscono una violazione della sovranità statale. Questo diventa particolarmente rilevante quando l'IA viene utilizzata per orchestrare attacchi su larga scala, come la diffusione automatizzata di contenuti falsi o la manipolazione dei dati elettorali.

L'uso di algoritmi di IA per creare contenuti personalizzati e persuasivi, progettati per influenzare le preferenze elettorali, rappresenta una sfida emergente al principio di integrità elettorale. Ad esempio, i bot automatici sui social media possono diffondere contenuti polarizzanti e falsi, amplificando artificialmente determinate narrazioni politiche. Allo stesso modo, i Deepfake possono presentare figure politiche in situazioni compromettenti, alterando la percezione pubblica in maniera fraudolenta. Questo tipo di ingerenza, alimentato dall'IA, non solo distorce il dibattito elettorale, ma mina la fiducia stessa nel processo democratico.

Iniziamo, quindi ad osservare quali siano stati alcuni dei casi che si sono venuti a presentare e quali siano state le tecniche utilizzate, e nelle casistiche in cui sarà possibile, analizzeremo anche quelli che sono stati gli effetti.

Sono molteplici i casi di ingerenza elettorale attraverso l'uso di tecnologie legate all'intelligenza artificiale (IA), spesso finalizzate alla diffusione di disinformazione, propaganda e manipolazione dell'opinione pubblica in paesi stranieri.

Uno dei primi casi e sicuramente uno tra i più noti di ingerenza elettorale con l'uso dell'intelligenza artificiale riguarda le elezioni presidenziali statunitensi del 2016, in cui sono state coinvolte campagne di disinformazione attribuite a entità russe.[77]

---

[76] Tallinn Manual è un documento non vincolante che fornisce linee guida interpretative sul diritto internazionale applicabile al cyberspazio e, in particolare, agli attacchi informatici. Originariamente sviluppata nel 2013 sotto l'egida della NATO Cooperative sotto legida del NATO Cooperative Cyber Defence Centre of Excellence (NATO CCDCOE) con la collaborazione di un gruppo di esperti internazionali, la Tallinn Manual cerca di stabilire come le norme del diritto internazionale, comprese quelle relative ai conflitti armati, si applichino al contesto del cyberspazio. Il cyber spazio è stato l'argomento fondante della seconda edizione del manuale (*Tallinn Manual 2.0*) assieme ad altri argomenti quali questioni di sovranità e responsabilità degli Stati nelle operazioni informatiche. Nonostante la Tallinn Manual non sia un trattato internazionale ufficiale, è ampiamente riconosciuta come una risorsa autorevole per comprendere come il diritto internazionale si applichi agli attacchi cibernetici e alle operazioni nello spazio digitale, soprattutto in relazione a sovranità statale, autodifesa e responsabilità internazionale.

[77] E. Kamarck, "*Malevolent soft power, AI, and the threat to democracy*", (29 novembre 2018), brookings.edu

Nonostante molti attacchi siano stati rivolti direttamente contro le infrastrutture deputate al voto attraverso sistemi di hacking molto sofisticati e permeanti [78], per altra parte di questi "attacchi" sono stati impiegati bot automatizzati alimentati da IA per diffondere disinformazione su piattaforme come Facebook, Twitter e Instagram. Questi bot pubblicavano messaggi su questioni divisive come razzismo, immigrazione e diritti civili, spesso mirati a polarizzare l'elettorato. Utilizzando tecniche di IA per creare contenuti mirati, i bot amplificavano false narrative e incrementavano la visibilità di post disinformativi

Per giungere ai fini di meglio influenzare quella che è l'opinione pubblica è necessario fare un passaggio indietro, infatti, prima di poter portare avanti una campagna così subdola, si vede necessario realizzare una profilazione psicologica, non ci sono fonti che ammettano che gli stessi agenti russi di cui abbiamo parlato al paragrafo precedente abbiano agito in questa maniera, ma è possibile trarlo come logica base per le operazioni che hanno messo in atto.

Visto quanto detto sopra, dovrebbe giungere alla memoria la controversa azienda Cambridge Analytica, la quale ha utilizzato strumenti basati su IA per analizzare i dati raccolti da milioni di profili di utenti su Facebook e creare profili psicologici dettagliati generando diversi scandali. Si sottolinea che questi dati potrebbero essere stati utilizzati per microtargeting degli elettori, inviando loro messaggi altamente personalizzati per influenzare la loro opinione politica e avendo una profilazione di questo carattere è facile comprendere che gli effetti potrebbero essere molto più incisivi rispetto ad un normale messaggio pubblicitario. Sebbene non sia stato provato che Cambridge Analytica abbia direttamente alterato i risultati elettorali, l'uso di IA per manipolare l'opinione pubblica, come è già stato detto, ha sollevato serie preoccupazioni sulla trasparenza e la democrazia.

Per pura continuità cronologica, e non perché la questione statunitense sia conclusa, di fatti verrà successivamente ripresa ed ulteriormente trattata, un altro caso legato all'uso di IA per modificare o comunque veicolare il convincimento della popolazione, è stato il referendum sulla Brexit nel 2016. Anche in questo contesto, l'uso di bot e algoritmi per amplificare disinformazione ha giocato un ruolo importante. In questo caso entità esterne, compresi gruppi legati alla Russia, hanno utilizzato bot sui social media per diffondere notizie false o fuorvianti

[78] Aa.vv. "*Report Of The Select Committee On Intelligence United States Senate On Russian Active Measures Campaigns And Interference In The 2016 U.S. Election*" (10 novembre 2020) Vol. 1: Russian Efforts Against Election Infrastructure With Additional Views Commitee sensitive – russia investigation only, 116th congress, 1st session.

riguardanti i benefici o i rischi della permanenza nell'Unione Europea. Questi bot, alimentati da IA, erano progettati per influenzare il dibattito politico e polarizzare l'opinione pubblica.[79] [80]

È stata utilizzata una strategia simile a quella di Cambridge Analytica, che ha cercato di sfruttare dati psicografici e analisi basate su IA per indirizzare messaggi su misura agli elettori indecisi.

Come parleremo più avanti, negli ultimi anni, ci sono state ripetute accuse di ingerenza russa nelle elezioni europee, come le elezioni parlamentari europee del 2019. Gruppi legati alla Russia hanno impiegato IA per generare fake news, Deepfake, e manipolare l'opinione pubblica nei paesi europei.

Le campagne di disinformazione miravano a destabilizzare le istituzioni democratiche europee, alimentando sentimenti anti-UE e sostenendo partiti populisti o estremisti. Le tecnologie basate su IA sono state utilizzate per generare automaticamente contenuti e post su misura per specifici segmenti di elettori.

Procedendo, però, in ordine cronologico, meritano di essere citati anche casi che sono passati più in sordina rispetto ai casi europei o statunitensi; in particolare nelle elezioni presidenziali brasiliane di fine del 2018, è emerso l'uso di bot e disinformazione automatizzata attraverso le piattaforme di messaggistica e social media, principalmente WhatsApp.

Il candidato vincente, Jair Bolsonaro, è stato accusato di beneficiare di campagne coordinate di disinformazione basate su IA per diffondere notizie false sul suo avversario. I bot

[79] Aa. vv., "*Intelligence and Security Committee of Parliament - Russia*", (21 luglio 2020), House of Commons, Pp. 12 - 14

[80] A questo proposito, si consiglia la lettura dell'articolo de "il sole 24 ore" in quanto, nonostante sono sia argomento da trattare in questa sede, la risposta a questo avvenimento, che allora erano solo accuse, non è stata considerata accettabile da diversi parlamentari, i quali frustrati dall'inerzia, ovvero dal rifiuto dell'azione governativa, dalla mancanza di indagini che potessero dare certezza ad accuse decisamente rilevanti, si sono mossi verso la corte di Strasburgo. Questa è la prima volta che dei deputati si sono rivolti alla Corte per "costringere" o per meglio dire correggere, l'azione del proprio governo. Per quanto, questo sviluppo sia di incredibile importanza, le accuse mosse al governo britannico riguardano il diritto dei cittadini di autodeterminare la propria idea politica senza ingerenze extra statali. N. Degli Innocenti "*Gran Bretagna: influenze russe nel voto per la Brexit. Governo Uk sotto accusa*", (19 gennaio 2023), Il Sole 24 Ore, edizione online.

sono stati utilizzati per inviare in massa messaggi ingannevoli e manipolare l'opinione pubblica, sfruttando la viralità e l'impatto delle fake news[81].

Altri casi riguardano quanto avvenuto negli stati africani, ci sono state eventualità nelle quali entità straniere hanno utilizzato tecnologie avanzate, per influenzare elezioni in questi paesi. Un esempio riguarda l'azienda "Archimedes Group", una società israeliana che, nel 2019, ha utilizzato una rete di bot e account falsi su Facebook per interferire nelle elezioni in diversi paesi africani, come Nigeria[82], Senegal, e Togo.[83]

Durante le elezioni presidenziali in Gabon, un video del presidente Ali Bongo[84] ha sollevato sospetti di essere un Deepfake. Dopo un lungo periodo di assenza dai media, è stato diffuso un video in cui Bongo teneva un discorso, ma molti analisti hanno evidenziato stranezze nelle immagini e nel movimento, suggerendo che potesse trattarsi di un Deepfake per dimostrare la sua buona salute e stabilità. Anche se non è stato confermato, il sospetto e le leggere discrepanze tra il video mostrato e i passati video del politico hanno destabilizzato la politica locale e generato confusione.

Un ulteriore esempio internazionale, che nonostante abbia influito molto sulla politica di un continente, riguarda le elezioni indiane del febbraio 2020, dove il partito politico Bharatiya Janata Party (BJP) ha usato i Deepfake in modo relativamente controllato. Il BJP ha utilizzato video modificati di uno dei suoi leader, Manoj Tiwari, che parlava in diversi dialetti per raggiungere diverse fasce di elettori. Sebbene questa fosse una forma di utilizzo "consentita", la tecnologia ha sollevato preoccupazioni su come i Deepfake possano essere impiegati per manipolare il consenso pubblico, portando alla plausibile conclusione che questa

[81] D. F. Aranha & J. van de Graaf, "*The Good, the Bad, and the Ugly: Two Decades of E-Voting in Brazil*" (Nov.-Dic. 2018) in *IEEE Security & Privacy*, vol. 16(6), Pp. 22 - 30.

[82] Questo riferimento viene inserito in nota per una questione di continuità temporale, in quanto, se fosse altrimenti fatto non si potrebbe avere una lineare, viste le necessarie eccezioni, progressione temporale del fenomeno: Durante le elezioni nigeriane di febbraio 2023, un audio manipolato con luso dellintelligenza artificiale è circolato sui social media, suggerendo falsamente il coinvolgimento di un candidato presidenziale dellopposizione in tentativi di manomissione del voto. Questo contenuto risch iava di aggravare sia le tensioni politiche tra i partiti, sia i persistenti sospetti sullintegrità del processo elettorale. A. Funk, A. Shahbaz & K. Vesteinsson, "*The Repressive Power of Artificial Intelligence*" (2023) freedomhouse.org

[83] E. Albrecht, E. Fournier-Tombs, & R. Brubaker, "*Disinformation and Peacebuilding in Sub-Saharan Africa: Security Implications of AI-Altered Information Environments*", (2024) New York: United Nations University Centre For Policy Research

[84] Y. Xu, P. Terhöst, M. Pedersen & K. Raja "*Analyzing Fairness in Deepfake Detection with Massively Annotated Databases.*" (marzo 2024). IEEE Transactions on Technology and Society, vol. 5 (1). Pp. 93-106.

mossa politica potesse ingiustamente far percepire il politico come vicino alle posizioni rurali di una zona grazie all'utilizzo del dialetto piuttosto che attraverso azioni concrete[85].

In Taiwan, durante la campagna elettorale del 2020, ci sono stati rapporti secondo cui la Cina avrebbe utilizzato Deepfake per diffondere disinformazione contro i candidati pro-indipendenza. Le tecniche di disinformazione digitale sono state combinate con Deepfake per creare confusione tra gli elettori e influenzare il risultato elettorale[86].

Dopo aver discusso di questi eventi, essendo che la questione statunitense è tornata di estrema attualità, si intende fare un piccolo balzo temporale al 2022 trattando delle elezioni francesi per poi considerare le elezioni USA

Di fatti durante le elezioni presidenziali in Francia del 2022, un Deepfake del presidente Emmanuel Macron è circolato sui social media. Questo video sembrava mostrare Macron fare dichiarazioni non veritiere, creando imbarazzo e confusione tra i suoi sostenitori. Anche se il video è stato rapidamente smentito, ha evidenziato come i Deepfake possano essere usati per attacchi di disinformazione mirati contro candidati[87].

Da ultimo torniamo a considerare la situazione d'oltre oceano: di fatti, durante le elezioni presidenziali negli Stati Uniti del 2020, si sono verificati casi di video manipolati che, sebbene non tutti definiti Deepfake, hanno utilizzato tecnologie avanzate per distorcere le dichiarazioni dei candidati. Un esempio è un video manipolato di Nancy Pelosi del 2019, che è stato rallentato per farla sembrare ubriaca durante un discorso. Questo caso non era un vero Deepfake basato su IA, ma ha mostrato come la manipolazione dei video possa avere un impatto politico significativo.

---

[85] C. Jeearchive , "*An Indian politician is using deepfake technology to win new voters",* (19 febbraio 2020), MIT technology review,.

[86] L. Sze-Fung, "*Deepfakes with Chinese Characteristics: PRC Influence Operations in 2024*", (29 marzo 2024) jamestown.org, China Brief, Vol. 24(7).

[87] Il video rappresenta Emmanuel Macron, davanti alla bandiera tricolore, rivolgersi ai francesi con i suoi "auguri", giustificando con voce esitante la "decisione di non aprire 10.000 posti aggiuntivi nei centri di emergenza". Dice: "Stiamo già spendendo una quantità folle di soldi nei sussidi sociali". Alla fine del video, pubblicato dalla deputata EELV Marie-Charlotte Garin su X (ex-Twitter), appare una smentita: "Il presidente potrebbe non aver detto queste parole, ma le sue azioni inviano questo messaggio". Dopo una controversia, Garin precisa che il video è un "deepfake", una tecnica di sintesi multimediale basata sullintelligenza artificiale. Vedi C. Gentilhomme "*«Une arme pour décrédibiliser un adversaire politique» : alerte contre les «deepfakes», ces vidéos truquées par intelligence artificielle*" (24 dicembre 2023), Le Figaro (edizione digitale)

Durante le elezioni presidenziali degli Stati Uniti nel 2020, l'intelligenza artificiale (IA) e altre tecnologie digitali hanno avuto un impatto significativo, alimentando disinformazione e manipolazione mediatica.

Le ingerenze russe nelle elezioni statunitensi erano già emerse durante le elezioni del 2016, ma nel 2020 sono state più sofisticate. Gli hacker russi, a differenza delle elezioni del 2016, ove gli attacchi erano più direzionati verso quelle che erano le sedi istituzionali del paese, hanno utilizzato social media, bot e reti di troll per diffondere disinformazione, spesso attraverso profili falsi, gruppi politicizzati e contenuti creati o amplificati tramite algoritmi di IA. Il rapporto dell'intelligence statunitense[88] ha confermato che attori legati alla Russia, tra cui l'Internet Research Agency (IRA), hanno cercato di polarizzare ulteriormente la società statunitense, usando la disinformazione per minare la fiducia nelle istituzioni democratiche e nei risultati elettorali[89].

Questo fenomeno si è manifestato attraverso diverse interferenze, alcune delle quali legate a ingerenze straniere, in particolare russe - ma anche iraniane e cinesi[90]-, e all'uso di Deepfake, che sono stati utilizzati per influenzare l'opinione pubblica e distorcere la percezione di figure politiche come Nancy Pelosi.

Prendiamo in esame questo caso, il video del 2019 raffigurante Nancy Pelosi, una Speaker della Camera dei Rappresentanti, è stato ampiamente condiviso sui social media e mostrava la Pelosi mentre parlava in modo confuso, come se fosse sotto l'influenza di sostanze alcoliche. Anche se questo video non era propriamente un Deepfake, ma più un "cheapfake" (manipolazione semplice come la modifica della velocità del video), la sua circolazione dimostrò come la distorsione dell'immagine pubblica di un politico potesse avere effetti devastanti[91].

Con l'avanzare delle tecnologie Deepfake, la preoccupazione durante le elezioni del 2020 era che video manipolati di questo tipo potessero essere usati in modo più sofisticato per

---

[88] Aa.vv., "*Report on The Investigation into Russian Interference In the 2016 Presidential Election"*, (marzo 2019), vol 2, U.S. Department of Justice, Washington D.C.

[89] A. Devy "*How Russia's Troll Farm Is Changing Tactics Before the Fall Election*" (29 marzo 2020), New York Times (edizione digitale)

[90] M. Santarelli, "La Russia Cambia La Strategia Di Disinformazione Politica Negli Usa", (23 marzo 2021), Agenda Digitale

[91] Vedi: R. Rijtano "*Da Nancy Pelosi alla Gioconda mai viste prima: lera dei deepfake è appena cominciata*" (29 maggio 2019), La Repubblica (edizione digitale)

falsificare dichiarazioni o azioni di politici influenti, e creare caos informativo. Deepfake più complessi e convincenti sono stati effettivamente individuati, anche se nessun caso clamoroso ha avuto un impatto devastante quanto si temeva, probabilmente a causa della crescente consapevolezza pubblica e della vigilanza dei media e delle piattaforme digitali e, probabilmente, vista anche l'esperienza del 2016.

Oltre al caso Pelosi, ci sono stati timori che i Deepfake potessero essere utilizzati per creare video falsi di Joe Biden o Donald Trump, per minare la loro credibilità o diffondere false dichiarazioni che avrebbero potuto influenzare il voto[92]. Anche se non ci sono state prove definitive di Deepfake su larga scala che abbiano avuto un impatto diretto sui risultati elettorali del 2020, diverse piattaforme hanno messo in atto misure preventive per rilevare e limitare la circolazione di tali video.

Le piattaforme di social media come Facebook, Twitter e YouTube hanno adottato politiche più severe contro i Deepfake (come discuteremo durante la trattazione del § 7.5), collaborando con ricercatori ed esperti di IA per sviluppare algoritmi in grado di identificare contenuti manipolati. Tuttavia, la difficoltà sta nella capacità dei Deepfake di eludere questi controlli, poiché la tecnologia continua a evolversi rapidamente.

L'intelligenza artificiale ha amplificato le sfide della disinformazione durante le elezioni del 2020, anche se gli scenari più catastrofici temuti con i Deepfake non si sono materializzati. L'ingerenza russa ha continuato a sfruttare le piattaforme digitali per creare divisioni e minare la fiducia nei processi elettorali, ma con una crescente consapevolezza pubblica e sforzi di mitigazione da parte delle piattaforme, molte delle minacce sono state contenute, almeno temporaneamente.

L'attuale presente, però, pone nuove sfide, poiché i Deepfake e altre forme di manipolazione mediatica potrebbero essere usate in modo più preciso e con un impatto più distruttivo se non affrontate con tempestive innovazioni di rilevamento e strategie di sensibilizzazione.

L'ingerenza di stati stranieri nelle elezioni americane del 2024 tramite l'intelligenza artificiale e i Deepfake è un tema di crescente preoccupazione. Paesi come la Russia, la Cina e

[92] Vedi D. O'Sullivan "*Congress to investigate deepfakes as doctored Pelosi video causes stir*" (4 giugno 2019), CNN (edizione digitale)

l'Iran sono attivi in queste operazioni, utilizzando strumenti di AI per generare contenuti falsi destinati a manipolare l'opinione pubblica. La Cina, ad esempio, ha impiegato sofisticati modelli di AI per creare video di ancoraggi di notizie falsi, e la Russia ha condotto campagne di disinformazione mirate, in particolare contro il sostegno americano all'Ucraina. Questi attacchi cercano di sfruttare le divisioni interne negli Stati Uniti e diffondere messaggi polarizzanti[93].

L'uso di Deepfake è particolarmente pericoloso perché può influenzare direttamente le percezioni degli elettori, specialmente quando i contenuti falsificati imitano candidati politici o creano eventi di crisi che non hanno mai avuto luogo. Un esempio è un Deepfake audio che ha imitato la voce di Joe Biden per scoraggiare i suoi sostenitori a votare durante le primarie del New Hampshire nel 2024[94]

O, ancora, ulteriore esempio riguarda diverse fotografie ritraenti gruppi di persone che indossano magliette con la scritta "Swifters for Trump", nonostante le foto fossero, abbastanza, evidentemente alterate, ciò ha fatto scaturire polemica con l'artista (Taylor Swift), la quale da subito ha smentito lo schieramento dichiarato nel post del candidato al quale sono allegale le fotografie di cui sopra. Sembra quasi banale sottolineare l'influenza che quest'artista ha sulle masse e quale sia potere mediatico della stessa, con la conseguente capacità di influenzare l'idea di massa[95].

Come tratteremo nel capitolo dedicato alle legislazioni e alle attività di regolamentazione (§ 7.3 & § 7.5) gli Stati Uniti stanno rispondendo a queste minacce con nuove leggi proposte che mirano a vietare l'uso di Deepfake con fini elettorali. Inoltre, grandi aziende tecnologiche stanno collaborando per etichettare automaticamente i contenuti generati dall'AI e migliorare la trasparenza nelle piattaforme digitali.

In conclusione, mentre l'IA e i Deepfake rappresentano una sfida, il loro impatto diretto sulle elezioni resta ancora incerto, ma, comunque, rimane un rischio significativo soprattutto se

[93] C.Watts, "*Russian US election interference targets support for Ukraine after slow start*" (17 aprile 2024) Microsoft on The Issues

[94] R. Mirza, "*How AI deepfakes threaten the 2024 elections"* (16 febbraio 2024) The journalist's Resource

[95] L. Tremolada "*Dai deepfake di Taylor Swift alle fake news con ChatGpt: come difendersi dalla disinformazione elettorale?*" (21 agosto 2024) Il Sole 24 Ore (edizione digitale)

utilizzati in momenti critici vicino al giorno del voto, quando non v'è tempo per procedere con un "debunking" delle informazioni.

.3 Far Guerra con le informazioni

Le nuove tecnologie da sempre permeano quelli che sono gli aspetti più truci della storia umana, o per meglio dire, spesse volte hanno una relazione così intima con questi da non poter comprendere se l'invenzione è nata per altro scopo se non per il fine bellico. L'intelligenza artificiale non fa troppa eccezione, se nel paragrafo precedente abbiamo potuto osservare quella che potrebbe dirsi una guerra di ingerenze o una nuova guerra fredda, ora si vuole trattare di quello che nell'immaginari comune è più facilmente indicabile come guerra vera e propria.

Come primo aspetto dobbiamo considerare un cambio rispetto a quelli che sono gli strumenti, di fatti le armi fisiche, chimiche o altri agenti che operano esclusivamente nella realtà fisica e materica non sono più sufficienti, molta parte delle battaglie devono essere sorrette dal supporto della popolazione, la capacità di governare quest'elemento è quello che porta civili e forze armate, anche straniere, a prendere una scelta piuttosto che altra. Da quanto abbiamo appena detto si può comprendere che è nato un aspetto digitale della guerra, ottenendo una guerra, per così dire, ibrida.

La guerra ibrida è una forma di conflitto che combina strategie militari convenzionali e non convenzionali con l'uso di tecnologie moderne, come la disinformazione, le operazioni cibernetiche e le campagne di propaganda. Questa modalità di guerra non si limita all'uso di forze armate tradizionali, ma include anche strumenti non militari, come attacchi informatici, atti terroristici, manipolazione delle informazioni e disordini civili, con l'obiettivo di destabilizzare un nemico senza entrare in uno scontro diretto[96]. In questo contesto, l'intelligenza artificiale (IA) e i Deepfake stanno assumendo un ruolo sempre più rilevante, grazie alla loro capacità di influenzare sia i militari che la popolazione civile. Questi strumenti, capaci di generare contenuti falsi o manipolati con un realismo impressionante, vengono utilizzati nelle

[96] F. G. Hoffman, "*Conflict in the 21st Century: The Rise of Hybrid Wars*" (2007) Potomac Institute for Policy Studies, Pp. 18-33.

guerre informative o operazioni psicologiche per seminare confusione, manipolare l'opinione pubblica e condizionare decisioni strategiche di alto livello.

L'utilizzo dell'IA può avere diversi scopi, il primo tra questi è sicuramente uno scopo propagandistico o disinformativo; L'intelligenza artificiale viene impiegata per analizzare enormi quantità di dati, soprattutto dai social media, individuando tendenze e vulnerabilità psicologiche della popolazione. Queste informazioni sono poi utilizzate per creare contenuti mirati, volti a influenzare le percezioni della guerra, a demonizzare il nemico e a generare confusione su cosa sia vero o falso. I Deepfake, in particolare, con la loro capacità di falsificare volti e voci in modo credibile, sono diventati uno strumento estremamente potente in questo ambito.

Da questo deriva, per conseguenza logica, il secondo obiettivo strategico, ovvero erodere quella che è la fiducia della popolazione nell'informazione, limitandone l'accesso o rendendo difficoltoso discernere tra il vero e l'artefatto[97]. In questo quadro, la diffusione di video o messaggi manipolati attraverso Deepfake mina la fiducia delle persone in ciò che vedono o sentono, creando uno stato di confusione generalizzata. Nei conflitti, questa incertezza può destabilizzare sia la popolazione civile che le forze armate, con effetti devastanti. La capacità di confondere e manipolare la percezione della realtà rappresenta un pericolo significativo in contesti di guerra.

Passiamo, ora ad un esempio emblematico dell'uso dei Deepfake in guerra è quello che ha coinvolto il presidente ucraino Volodymyr Zelensky durante l'invasione russa del 2022. In quell'occasione, venne diffuso un video manipolato sui social media e in TV, che mostrava Zelensky mentre chiedeva ai soldati ucraini di arrendersi ai russi. Sebbene il video non fosse

[97] Prendendo le mosse dalla celebre frase attribuita ad Ernesto Che Guevara "Un popolo ignorante è più facile da governare" o, per altre fonti "…da ingannare", non è falsificata nemmeno in questa occasione, è risaputo che meno permetti alla popolazione di prendere coscienza di quanto sta succedendo minori saranno le possibilità che questi si rivoltino verso quello che è un potere centrale opprimente ovvero che si adoperi per contrastare quella che è un'azione contro i propri interessi. Questa è sostanzialmente la condizione in cui sopravvive il popolo russo, tagliato da qualsivoglia contatto con altri stati eccezion fatta per le poche concessioni permesse dal governo stesso. È altrettanto vero che la campagna militare in Ucraina procede in sordina anche grazie alla difficoltà nel reperimento di informazioni veritiere legate alla situazione e non solo a causa di un disinteresse generale per una questione che, per quanto vicina viene considerata distante dai nostri ambienti familiari. Cfr: "*Russia, censura dell'informazione e persecuzione delle proteste contro la guerra. Crescono opposizione alla guerra e repressione del dissenso*" (28 febbraio 2022) amnesty.it; V. Mancini & P. DellOsso "*Focus Ucraina – La censura del Cremlino sulla guerra*" (6 aprile 2022) irpa.eu; G. Asta "*La Russia blocca laccesso a RaiNews.it e altri 80 media dellUnione europea* " (25 giugno 2024) RaiNews.it.

perfetto, era abbastanza realistico da poter ingannare chi non fosse esperto di queste tecnologie[98].

il video faceva parte di un'operazione di disinformazione mirata a demoralizzare le forze armate ucraine e la popolazione civile. Mostrare il proprio leader che si arrendeva avrebbe potuto indebolire il morale dei soldati, facendoli dubitare della missione e facendoli pensare che la guerra fosse già persa.

Anche se il video venne rapidamente smentito da Zelensky stesso e dai media internazionali, l'episodio dimostrò chiaramente la pericolosità di queste tecnologie. In questo caso, la tempestività della reazione ha limitato i danni, ma ha anche messo in evidenza il potenziale distruttivo dei Deepfake in un contesto di guerra.

Altro Deepfake, forse anche più pericoloso vista la possibilità di interagirci direttamente, utilizzato durante la stessa guerra è quello del 2022, Vitaly Klitschko, il sindaco di Kiev, fu vittima di un sofisticato attacco di Deepfake. In questo episodio, truffatori digitali crearono una versione falsificata di Klitschko per condurre videoconferenze con diversi sindaci europei, tra cui quelli di Berlino, Vienna e Madrid. I criminali sfruttarono la tecnologia del Deepfake per far sembrare che Klitschko stesse parlando direttamente con loro, ingannando inizialmente gli interlocutori[99].

Il Deepfake mostrava Klitschko parlare in modo convincente, ma i sindaci e i loro staff notarono alcune incongruenze e argomenti insoliti, come richieste di aiuti militari che sollevarono sospetti. Dopo le prime interazioni, le conversazioni furono interrotte, e l'inganno venne alla luce. Questo evento sollevò serie preoccupazioni sulla sicurezza digitale e le potenziali implicazioni politiche dei Deepfake, in un periodo particolarmente delicato a causa della guerra in Ucraina.

La manipolazione di messaggi da parte dei leader militari attraverso Deepfake o altre forme di IA può causare confusione nei ranghi, rallentare le decisioni tattiche e, nei casi

[98] V. Berra "*Zelensky chiede alle truppe di deporre le armi, ma è un falso: il video deepfake rimosso da Facebook e YouTube*" (17 marzo 22) Open.online

[99] Autore sconosciuto "*Kiev, in videochiamata con un falso Klitschko: i sindaci di Madrid, Berlino e Vienna ingannati dal deepfake*" (25 giugno 2022) La Repubblica (edizione online)

peggiori, portare alla resa. Le forze armate devono quindi essere preparate a distinguere tra ordini legittimi e falsificati, complicando le operazioni in tempo reale[100] [101].

Visto quanto appena detto, il campo di battaglia non è più considerabile come esclusivamente il luogo dove sono attive le forze militari, ma in molti conflitti moderni, le operazioni psicologiche mirano a minare il morale della popolazione o a provocare panico e disorganizzazione. La disinformazione può spingere i cittadini a prendere posizione, sostenendo o opponendosi a determinate decisioni politiche o militari basate su informazioni false. I Deepfake, con la loro capacità di replicare eventi e discorsi in modo credibile, possono essere particolarmente efficaci nel fomentare il caos.

Le risposte a questi attacchi devono essere veloci e ben mirate. Tra le strategie di contrasto più efficaci possiamo trovare lo sviluppo di tecnologie di rilevamento di cyberattacchi, tra cui anche i Deepfake, ovvero utilizzare intelligenze artificiali per identificare manipolazioni delle informazioni così da limitarne la diffusione.

Questo strumento deve essere, però, supportato da una forte sensibilizzazione e educazione del pubblico al fine di renderla più scettica e capace di differenziare quelle che possono essere informazioni vere da quelle manipolate – così com'è successo per le manipolazioni durante le elezioni americane nel 2020, come parlavamo al sottocapitolo precedente (§ 5.2).

Da ultimo, ma non certamente per importanza, la capacità dei leader o delle istituzioni di reagire rapidamente e confutare i falsi può limitare significativamente l'impatto della disinformazione.

l'utilizzo dell'IA e dei Deepfake nei conflitti moderni rappresenta una nuova frontiera nella guerra ibrida, dove le operazioni psicologiche e la disinformazione sono cruciali per influenzare tanto i soldati quanto i civili. Il caso del Deepfake di Zelensky - o quello del sindaco di Kiev - è un chiaro esempio del potere distruttivo di queste tecnologie, ma dimostra anche che

---

[100] Vedi P.W. Singer, E.T. Brooking, "*LikeWar: The weaponization of social media*." (2018) Eamon Dolan Books, Pp. 83-117, 148-180.

[101] Vedi: "L' agenzia ucraina Ukrinform, ripresa da vari media, che riportano una protesta dello stesso Klitschko, che accusa i russi di praticare la guerra su tutti i fronti, "compreso quello della diffusione di disinformazione per mettere politici ucraini in cattiva luce, mettendoli contro i loro partner occidentali con lo scopo di interrompere gli aiuti dellOccidente allUcraina".". redazione Ansa " *Dopo deepfake Zelensky, anche video finto sindaco Kiev*" (28 giugno 2022) Ansa.

una preparazione adeguata e una risposta tempestiva possono contrastarne efficacemente gli effetti.

.4 L'intelligenza della pornografia

Negli ultimi vent'anni, Internet è stato spesso considerato un nuovo strumento per esprimere la sessualità umana[102]. Sebbene contenuti erotici fossero accessibili già a partire dall'invenzione della stampa di Gutenberg, Internet ha reso questo accesso molto più facile e diffuso[103]. Con "attività sessuale online" si intende qualsiasi interazione sessuale svolta sul web, sia che avvenga per svago, per la ricerca di partner, per informazione o per motivi commerciali[104].

Anche se alcuni crimini sessuali avvengono online- di alcuni parleremo nei capitoli successivi (§8) -, la maggior parte delle persone non sembra manifestare problemi patologici o effetti negativi derivanti da tali attività.

Questo detto, le attività legate alla pornografia per mezzo internet sono estremamente permeanti dello strumento, arrivando fino alla citazione, non attribuibile ad un soggetto specifico, ma di comune conoscenza, "l'ultimo sito che rimarrà dopo il crollo di internet sarà un sito porno".

Altra dimostrazione di questo concetto deriva dalle regole di internet. Per quanto il cyberspazio sia universalmente riconosciuto come uno spazio caotico e sregolato, questo soprattutto durante la sua nascita, esistono delle "regole di internet" stilate per la prima volta

---

[102] J. S. Carroll, L. M. Padilla-Walker, L. J. Nelson, C. D. Olson, C. McNamara Barry & S. D. Madsen "*Generation XXX: Pornography acceptance and use among emerging adults.*" (2008) Journal of adolescent research vol. 23(1), Pp. 6 - 30.

[103] Cfr. A. Barak, W. A. Fisher, S. Belfry, & D. R. Lashambe "*Sex, Guys, and Cyberspace: Effects of Internet Pornography and Individual Differences on Men's Attitudes Toward Women*" (1999) Journal of Psychology & Human Sexuality, Vol. 11(1), Pp. 63 - 91. A. Cooper, N. Galbreath, & M.A. Becker, "*Sex on the Internet: Furthering Our Understanding of Men with Online Sexual Problems.*" (2004). Psychology of Addictive Behaviors. Vol. 18(3), Pp. 223 - 230.

[104] A. Cooper, E. Griffin-Shelley, D.L. Delmonico & R.M. Mathy "*Online Sexual Problems: Assessment and Predictive Variables.*" (2001). Sexual Addiction and Compulsivity. Vol 8 (3-4), Pp. 267 - 285.

sulla piattaforma di Blog 4Chan dal 2003, presumibilmente con la cooperazione di alcuni attivisti di Anonymous; Anche per non alterare la natura caotica di internet, queste "regole" non sono indicizzate in maniera univoca, escluse poche eccezioni che sono generalmente accettate.

Una delle prime regole, e anche quella più utile per questa trattazione, è la #34, la quale recita: "se una cosa esiste, allora ne esiste anche la versione porno"[105]. Seguita di lì a poco dalla successiva #35: "Se non c'è nessun porno al momento, esso sarà creato."

L'intelligenza artificiale, nella sua sempre crescente capacità di conoscere e soddisfare le pretese e volontà umana, da che è stata codificata per la generazione o modifica di immagini e video è stata da, quasi da subito, prestata alla realizzazione di materiale pornografico. Questi materiali, se la loro produzione è consenziente non fa sorgere alcun problema. Nel caso in cui il consenso mancasse, si entra negli eventi che nei prossimi paragrafi definiremo e analizzeremo: Deep nude e Deep porn.

I Deep porn sono una tipologia di contenuto sessuale non consensuale creato utilizzando tecniche di intelligenza artificiale, come quelle utilizzate nei Deepfake. Il termine "deep porn" si riferisce a video o immagini in cui il volto di una persona è sovrapposto in modo realistico su un corpo diverso, spesso in contesti pornografici, senza il consenso della persona. Uno degli sviluppi più noti all'interno di questo contesto è stato DeepNude, un applicativo che permetteva di "svestire" digitalmente le donne utilizzando l'intelligenza artificiale.

L'app DeepNude è stata lanciata nel 2019 e utilizzava tecniche di intelligenza artificiale, in particolare le reti neurali generative (GANs, Generative Adversarial Networks, come si è meglio nel capitolo dedicato §3.3), per rimuovere virtualmente i vestiti da immagini di donne, creando nudità realistiche. Il software era in grado di produrre immagini estremamente convincenti, sollevando immediatamente preoccupazioni etiche e legali. Dopo una serie di critiche, l'applicazione è stata ritirata dal mercato dai suoi stessi sviluppatori entro pochi giorni

[105] La prima formulazione risale al 2003, quando Peter Morley-Souter, dopo aver casualmente trovato una versione pornografica di *Calvin & Hobbes*, creò una vignetta che raffigurava un ragazzo seduto davanti a un computer con unespressione sconvolta che esclamava: "Calvin e Hobbes?", aggiungendo poi: "Internet. Distrugge la tua innocenza dal 1996" (F. Lanza "*La storia della regola 34 di Internet*." (6 novembre 2015) linkiesta.it). Questo detto si deve aggiungere che un articolo del The Telegraph (allora The daily telegraph) ne parlò nel 2009 come una delle "top 10" regole e leggi di internet (T. Chivers "Internet rules and laws: the top 10, from Godwin to Poe" (23 ottobre 2009) The Telegraph (ed digitale)), successivamente, sempre in riferimento alla regola 34, è stato pubblicato un articolo su CNN, nel 2013, il quale affermava che questa fosse la regola più famosa del cyberspazio nonostante restasse un "meme" o uno scherzo(T. Leopold "meet the rules of internet" (15 febbraio 2013) CNN business (ed digitale)).

dalla diffusione. Tuttavia, copie pirata e varianti del software continuano a circolare sul web.[106] [107]

DeepNude rappresenta uno dei primi strumenti di facile accesso a chiunque volesse creare contenuti sessualmente espliciti falsificati, aumentando drasticamente il rischio di abuso e diffusione di immagini non consensuali. Il creatore dell'app si è giustificato affermando di non aver previsto il possibile uso improprio [108], ma questo non ha impedito che il software fosse utilizzato per lo scopo di creare deep porn.

L'evoluzione dei deep porn è strettamente collegata allo sviluppo dei Deepfake, dei quali abbiamo estensivamente trattato al capitolo (§4), ovvero tecnologie inizialmente sviluppate per scopi creativi o di intrattenimento. Il primo esempio noto di Deepfake a scopo pornografico risale al 2017, quando su Reddit apparvero i primi video in cui volti di attrici famose venivano sovrapposti a corpi di attrici pornografiche. Questi video venivano realizzati grazie a software di intelligenza artificiale sempre più sofisticati, che permettevano di generare contenuti falsi ma estremamente realistici. Reddit, dopo pressioni da parte di utenti e osservatori esterni, bandì il subreddit dedicato a questi contenuti, ma ormai la tecnologia si era diffusa, e con essa le varianti del fenomeno.

La combinazione della disponibilità di strumenti di AI sempre più accessibili, come quelli sviluppati su piattaforme open-source, e l'uso di GPU sempre più potenti ha reso possibile a molti utenti la creazione di video deep porn con poche competenze tecniche. Nonostante gli sforzi da parte di varie piattaforme per fermare la diffusione di questi contenuti, gli strumenti continuano a essere accessibili su forum e darknet.

Numerosi casi di Deepfake pornografico hanno coinvolto celebrità, attirando l'attenzione sui gravi danni psicologici e reputazionali causati a queste vittime. Taylor Swift, Gal Gadot, Scarlett Johansson e altre donne famose sono state tra le più colpite, con contenuti

---

[106] Come esempio più lampante è possibile trovare una pagina denominata esattamente Deepnude.org che si propone di agire esattamente come l'applicativo in esame.

[107] "Con la tecnologia del deepnude, infatti, i visi delle persone (compresi soggetti minori) possono essere "innestati", utilizzando appositi software, sui corpi di altri soggetti, nudi o impegnati in pose o atti di natura esplicitamente sessuale.". Aa.Vv. "*Deepfake-vademecum*" (28 dicembre 2020) Garante della Privacy.

[108] "*Abbiamo creato questa app per far divertire gli utenti. Non immaginavamo che sarebbe diventata virale e che non saremmo stati capaci di controllare il traffico*", hanno scritto i creatori di DeepNude, lapplicazione in questione, in un post su Twitter. " *La possibilità che le persone ne abusino è troppo alta*". G. Giacobini "*Storia dellapp che genera(va) false foto di nudo femminile* " (28 giugno 2019) Wired <www.wired.it/>

falsi generati e diffusi senza consenso che hanno danneggiato la loro immagine pubblica e privata. Indagini hanno rivelato che quasi 4.000 celebrità sono state vittime di pornografia Deepfake, con la maggior parte delle vittime che sono donne. Una ricerca del 2024 ha confermato l'incremento della produzione e della distribuzione di questi contenuti a causa della facilità con cui strumenti di intelligenza artificiale avanzata sono oggi accessibili al pubblico, rendendo il fenomeno sempre più difficile da arginare.[109]

Diversi individui e aziende sono stati coinvolti in scandali legati ai deep porn. Tra le figure più note possiamo citare attrici come Scarlett Johansson e Gal Gadot.

La prima è una delle vittime più famose di Deepfake pornografici. La sua immagine è stata utilizzata senza consenso in vari video espliciti, portando l'attrice a parlare pubblicamente del danno che tali contenuti causano alla reputazione delle persone coinvolte. Johansson ha affermato che è praticamente impossibile per una vittima ottenere giustizia o prevenire la diffusione di questi contenuti.[110]

Gal Gadot è stata un'altra delle primissime celebrità ad essere vittima del fenomeno dei Deepfake pornografici, in cui il suo volto è stato inserito in video espliciti senza il suo consenso. Questi contenuti sono stati creati utilizzando algoritmi avanzati di intelligenza artificiale per realizzare rappresentazioni realistiche e altamente dettagliate, diffusi su diverse piattaforme online.

La diffusione di Deepfake pornografici con volti di celebrità, come quello di Gal Gadot, ha contribuito a sollevare questioni legali ed etiche riguardanti l'uso dell'immagine personale senza autorizzazione, evidenziando il bisogno urgente di regolamentazioni più stringenti e di strumenti di protezione. Gal Gadot, insieme ad altre celebrità, ha espresso preoccupazione per la mancanza di protezione legale e per le difficoltà incontrate dalle vittime di Deepfake nel far rimuovere tali contenuti, che rimangono online e possono essere facilmente replicati.

L'uso delle tecnologie di Deepfake nel contesto pornografico è destinato a crescere se non vengono attuate regolamentazioni più severe. Al momento, esistono strumenti come il

[109] N. Badshah "*Nearly 4,000 celebrities found to be victims of deepfake pornography*" (21 marzo 2024) The guardian (ed digitale). Crf C Newman "*Exclusive: Hundreds of British celebrities victims of deepfake porn*" (21 marzo 2024) Channel 4.

[110] D.Harwell, "*Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*" (3 dicembre 2018) The Washington Post (ed online)

Project PIMEyes[111], che cercano di combattere il deep porn permettendo agli utenti di rintracciare immagini e video non consensuali attraverso la ricerca per immagine. Tuttavia, la diffusione di strumenti simili ai DeepNude e la facilità con cui possono essere modificati rende ancora difficile arginare il problema.

.5 Applicazioni positive dei Deepfake: innovazioni e benefici

I Deepfake, sebbene spesso associati a usi negativi, ciò anche provato da tutto quanto detto fino a questo momento, è da notare che stanno trovando applicazioni positive in vari settori, dal cinema alla terapia, dall'educazione alle campagne sociali.

Nella cinematografia, i Deepfake hanno trasformato l'approccio agli effetti speciali, risultando efficaci e meno costosi rispetto alle tecniche tradizionali. Un esempio significativo è stato il film Star Wars: Rogue One, in cui il personaggio di Peter Cushing, deceduto anni prima, è stato riportato in scena grazie a tecnologie simili ai Deepfake. Questa scelta ha permesso di rispettare l'integrità narrativa senza compromettere la qualità visiva del film, ovvero, per rimanere nello stesso universo cinematografico, nello sviluppo della serie The Mandalorian 2, volendo utilizzare il personaggio di luke skywalker con le fattezze che possedeva durante la produzione del sesto capitolo della celeberrima saga, è stato realizzato un Deepfake molto elaborato così da sostituire le fattezze dell'attore in scena al momento della ripresa con quelle di Mark Hamill nel 1982.[112]

---

[111] PimEyes è una piattaforma di riconoscimento facciale che consente agli utenti di rintracciare immagini di sé stessi sul web caricando una foto. Creata per aiutare a proteggere la privacy online, viene spesso utilizzata per individuare immagini non autorizzate o potenzialmente dannose. Tuttavia, ha suscitato preoccupazioni etiche: alcuni critici ritengono che possa favorire lo stalking o la sorveglianza non autorizzata. Diverse forze di polizia nel Regno Unito ne hanno limitato l'uso proprio per questi rischi, mentre PimEyes continua a dichiarare il suo impegno nel rispetto della privacy e nell'utilizzo responsabile della tecnologia. M. Wilding, C. Milmo et al. "*Met police computers access 'dangerous' facial recognition search engine*"(10 maggio 2024) liberty investigate <libertyinvestigates.org>. "*PimEyes: The Digital Detective Redefining Cold Case Investigation*" pimeyes.com

[112] Vedi: A. Qobraty "*Star Wars fans love CGI Luke Skywalker, but deepfake implications are dangerous*" (3 marzo 2022) The daily Targum <www.dailytargum.com>. Cfr C. Sernagiotto "*The Mandalorian, Lucasfilm assume lo YouTuber del video deepfake con Luke Skywalker*" (28 luglio 2021) Sky tg 24 <tg24.sky.it>

Nel doppiaggio, i Deepfake stanno migliorando la qualità e l'accessibilità dei contenuti audiovisivi attraverso una sincronizzazione precisa delle labbra, indipendentemente dalla lingua originale. Questo sviluppo rende i doppiaggi più naturali e migliora l'esperienza per gli spettatori di tutto il mondo, riducendo il divario tra lingua originale e traduzione.

Uno degli esempi più noti di utilizzo positivo dei Deepfake è la campagna "Malaria Must Die"[113] è stata un'iniziativa di sensibilizzazione lanciata per aumentare la consapevolezza globale sulla malaria e promuovere azioni per la sua eradicazione. Uno degli aspetti più innovativi di questa campagna è stato l'uso della tecnologia Deepfake, che ha coinvolto la celebrità e attivista David Beckham. Utilizzando l'intelligenza artificiale, Beckham è stato "programmato" per parlare in nove lingue diverse, rendendo il messaggio accessibile a un pubblico globale più ampio. La campagna è stata ideata per creare un impatto emotivo forte e catturare l'attenzione delle persone, grazie all'effetto realistico della tecnologia di sintesi vocale e video.

Beckham, in realtà, non ha pronunciato i discorsi in tutte le lingue utilizzate nel video. Invece, la sua voce è stata generata artificialmente per sembrare autentica, sfruttando tecniche di deep learning e intelligenza artificiale. Il risultato è stato un video efficace e coinvolgente che ha portato il messaggio della lotta alla malaria a milioni di persone in tutto il mondo.

La campagna "Malaria Must Die" rappresenta un esempio positivo dell'uso della tecnologia Deepfake per scopi umanitari e socialmente utili. Ha dimostrato che, se utilizzate in modo etico, queste tecnologie possono amplificare il messaggio di una causa nobile, portando all'attenzione del pubblico questioni cruciali e coinvolgendo un'audience globale in un modo che sarebbe stato difficile ottenere con metodi tradizionali.

Altro caso in cui si può parlare di doppiaggio, e del quale si è già discusso nel paragrafo dedicato alle ingerenze nelle elezioni (§.2), nel febbraio 2020, durante la campagna per le elezioni dell'Assemblea Legislativa di Delhi, il Bharatiya Janata Party (BJP) ha utilizzato la tecnologia Deepfake per ampliare la propria comunicazione elettorale. In particolare, sono stati creati video in cui il leader del partito, Manoj Tiwari, appariva parlare in dialetti locali come l'haryanvi e in inglese, nonostante l'originale fosse in hindi. Questa strategia mirava a

[113] https://malariamustdie.com/

raggiungere elettori di diverse comunità linguistiche, rendendo il messaggio più accessibile e personale.

La realizzazione di questi video è stata affidata alla società di comunicazione indiana The Ideaz Factory, che ha collaborato con l'unità IT del BJP di Delhi per creare "campagne positive" utilizzando Deepfake. Questa collaborazione ha segnato il debutto dei Deepfake nelle campagne elettorali in India.[114]

I video modificati sono stati distribuiti in oltre 5.800 gruppi WhatsApp, raggiungendo circa 15 milioni di persone. L'obiettivo era convincere gli elettori, in particolare quelli di lingua haryanvi, a votare per il BJP, presentando Tiwari come un leader in grado di comunicare direttamente nella loro lingua.[115]

Questa iniziativa, come già detto, ha sollevato preoccupazioni non indifferente riguardo all'uso etico dei Deepfake in politica, poiché la tecnologia può essere utilizzata per manipolare l'opinione pubblica. Tuttavia, i rappresentanti del BJP hanno sottolineato che l'uso dei Deepfake in questo contesto era destinato a scopi positivi, permettendo al partito di avvicinarsi agli elettori in modo più efficace.[116]

Seppur di non particolare rilevanza, è interessante notare che nel giugno del 2024 la popolazione indiana è stata chiamata alle urne per le elezioni parlamentari. In questa occasione sono state centinaia i politici a richiedere o per i quali sono stati prodotti video Deepfake più o meno satirici, o comici, molti riprendevano famose scene di film della casa di produzione di Bollywood così da aumentare il trasporto degli elettori per il candidato. Unico video di cui si vuole far menzione, a titolo puramente esemplificativo, è quello nel quale una raffigurazione leader dell'opposizione, Arvind Kejriwal (questi arrestato a ridosso della campagna elettorale) canta con la chitarra in mano: «Dimenticami, devi vivere senza di me ora», come in una popolare canzone di Bollywood. [117]

---

[114] C. Nilesh "*We've Just Seen the First Use of Deepfakes in an Indian Election Campaign*" (18 febbraio 2020) Vice <www.vice.com>

[115] ibidem

[116] B. Dasgupta "*BJP's deepfake videos trigger new worry over AI use in political campaigns*" (21 settembre 2020) Hindustan Times, New Delhi (edizione digitale)

[117] A. Muglia "*Il re dei deepfake indiani: «Sono inondato dalle richieste dei politici»*" (30 aprile 2024) Corriere della sera (edizione digitale)

Anche in ambito educativo e culturale, i Deepfake stanno rendendo le esperienze più coinvolgenti. Musei e progetti storici utilizzano queste tecnologie per "resuscitare" figure storiche.

Video e audio sintetici di rievocazioni storiche, come l'apparizione di un personaggio del passato, possono aumentare l'impatto e il coinvolgimento degli studenti, diventando strumenti di apprendimento più efficaci[118]. Un esempio è la ricreazione, tramite tecniche IA, del discorso mai pronunciato di John Kennedy per porre fine alla guerra fredda, che è stato riprodotto con la sua voce sintetica e il suo stile. Questo approccio permette agli studenti di esplorare temi storici in modo creativo e immersivo.

Oltre alla storia, la tecnologia sintetica consente di modellare e simulare vari campi di studio, come l'anatomia umana, complessi macchinari industriali, e molto altro ancora, all'interno di ambienti di realtà mista. Queste simulazioni, fruibili attraverso visori di realtà virtuale come il Microsoft Hololens[119], offrono a studenti, futuri medici, forze dell'ordine e altri professionisti la possibilità di svolgere esercitazioni pratiche in modo realistico, efficace e sicuro. L'uso creativo di voce e video sintetici aumenta il successo complessivo e i risultati dell'apprendimento, offrendo esperienze educative ricche e interattive con costi e risorse limitati.

MyHeritage[120], ad esempio, ha lanciato un progetto che anima vecchie fotografie, consentendo agli utenti di vedere i propri antenati o figure storiche prendere vita in modo realistico. Questa applicazione ha avuto un enorme impatto emozionale, trasformando le tradizionali esperienze museali in dinamiche e accessibili.

Uno dei progetti più noti è Deep Nostalgia[121], che utilizza tecniche di intelligenza artificiale per animare le fotografie dei propri antenati, consentendo agli utenti di vedere le immagini dei loro cari prendere vita attraverso espressioni facciali e movimenti realistici.

---

[118] "*Taking medical instruction remote and to mixed reality at Case Western*" Microsoft Customer Stories <www.customers.microsoft.com/en-us >

[119] "*HoloLens 2 x Education*" Microsoft Education. < www.microsoft.com/en-us >

[120] MyHeritage, una piattaforma nota per il suo servizio di genealogia e analisi del DNA, ha introdotto una serie di strumenti avanzati basati sullintelligenza artificiale per rendere le esperienze di ricerca storica e genealogica più interattive e coinvolgenti. <www.myheritage.it >

[121] https://deepnostalgia.ai/

Secondo MyHeritage, Deep Nostalgia si basa su tecnologie di deep learning sviluppate dalla società D-ID, specializzata in sintesi video e animazione facciale.

Lanciato nel 2021, questo strumento ha rapidamente attirato l'attenzione globale, poiché permette di creare una connessione emotiva con immagini di persone che non è possibile conoscere direttamente, come antenati lontani o figure storiche. Come riportato da The Verge, l'applicazione ha riscosso un enorme successo grazie al suo impatto emotivo, consentendo agli utenti di vedere per la prima volta i volti dei loro familiari storici muoversi e sorridere[122]. Ciò ha aumentato il coinvolgimento degli utenti con la piattaforma di genealogia e ha portato molti a vedere la storia della propria famiglia in modo più personale e vivido.

Oltre a Deep Nostalgia, MyHeritage ha integrato strumenti come Photo Enhancer e Photo Repair, che utilizzano l'IA per migliorare la qualità delle vecchie fotografie, rendendole più nitide e dettagliate. Questi strumenti possono ripristinare e colorare le immagini, trasformando vecchie foto in bianco e nero in versioni colorate realisticamente, facendo emergere dettagli che altrimenti andrebbero persi con il tempo. Il successo di queste applicazioni dimostra come l'intelligenza artificiale possa arricchire l'esperienza genealogica e portare un valore aggiunto alle persone interessate a scoprire le proprie radici familiari.[123]

Anche in campo medico queste tecnologie stanno avendo un grande impatto, basti pensare che già dai primi approcci con il progetto Dendral e gli altri algoritmi specializzati (di cui al §3.4) e considerare l'evoluzione che è intervenuta sui sistemi informatici dagli anni '70 ad oggi. Preso coscienza di quelle informazioni, ad oggi, i campi di applicazione di questi sistemi in ambito medico sono molteplici, come potremo parlare a breve e come abbiamo discusso in precedenza (parlando di HoloLens), la possibilità di non agire direttamente su un paziente e poter "provare" interventi su copie artificiali è un grande passo avanti per la sicurezza dei degenti.

---

[122] K. Lyons "*New AI 'Deep Nostalgia' brings old photos, including very old ones, to life*" (28 febbraio 2021) The verge <www.theverge.com >

[123] Certo non è tutto oro quel che luccica, le problematiche legate a questi strumenti, seppur separate e distinte dallo scopo per le quali sono stati creati, esistono e non sono di poco conto. Per accedere a questi strumenti vengono richiesti numerosi dati estremamente sensibili, primi tra tutti dati atti all'identificazione personale e dati bancari; questi dati passano, però, in secondo piano nel momento in cui si pensa che questa società tratta dati genetici, li analizza e li utilizza per profilare i propri clienti, essendo questo lo scopo primario dell'azienda, ciò non di meno si comprende la sensibilità di queste informazioni e ciò è stato molto più accentuato nel 2018, anno nel quale v'è stata una breccia nella sicurezza dell'azienda. Cfr F. Michetti "*Deep Nostalgia non è un giochetto, ma la nuova arma della post-verità*" (19 marzo 2021) Agenda digitale <agendadigitale.eu>

Nonostante questa possibilità, con gli aggiornamenti del Regolamento Generale sulla Protezione dei Dati (GDPR)[124] nell'UE, il flusso libero di dati è stato limitato per garantire il consenso e l'anonimato dei pazienti. Anche i dati anonimizzati non possono essere condivisi tra gruppi di ricerca in diversi Paesi, poiché combinando alcune variabili di un dataset si potrebbe risalire all'identità di una persona. Per qualsiasi trasferimento di dati sanitari, è quindi necessario che ogni paziente fornisca un consenso informato.

Tuttavia, la medicina personalizzata necessita di grandi set di dati sanitari aperti e accessibili pubblicamente per migliorare le soluzioni di machine learning nel campo medico. Ed è qui che i Deepfake possono offrire un contributo significativo: è possibile generare "pazienti artificiali" partendo dai dati reali, alleggerendo così i problemi di privacy e permettendo la condivisione di dati sintetici con altri gruppi di ricerca. Un esempio recente di utilizzo dei Deepfake in medicina è la creazione di dati sintetici di elettrocardiogrammi realistici. Un recente studio[125] ha dimostrato come questa tecnologia possa generare dati clinici utili per la ricerca, evitando al contempo le problematiche legate alla privacy.

Oltre alla generazione di dati sintetici, i Deepfake aprono nuove possibilità per altre applicazioni in medicina. Per molti anni è stato possibile far parlare un computer digitando del testo, ma ora, con i Deepfake vocali, si può riprodurre la voce specifica di una persona anche senza che questa abbia mai pronunciato determinate parole. Questa tecnologia sta già trasformando la vita di persone che hanno perso la capacità di parlare, come i pazienti colpiti da ictus o affetti da malattie progressive, come la sclerosi laterale amiotrofica (SLA). Il linguaggio sintetico permette loro di comunicare con la propria voce, permettendo di interagire con i propri cari anche dopo aver perso la possibilità di parlare autonomamente, come mostrato nel documentario Gleason del 2016.[126]

Un ulteriore possibile utilizzo, anche delle applicazioni di deep nostalgia, sotto la supervisione di un terapista, consente alle persone di avere conversazioni video realistiche con un Deepfake di una persona cara deceduta, offrendo così conforto e aiuto per superare la perdita. Questa applicazione terapeutica dei Deepfake dimostra come tali tecnologie possano non solo

[124] Art. 9 par. 2 lett. A)

[125] V. Thambawita, J.L. Isaksen, S.A. Hicks, et al. "*DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine.*" (9 novembre 2021) Nature, vol. 11, art. n. 21896.

[126] Vedi: Gleason (2016) Imdb.com <https://www.imdb.com/title/tt4632316/>

proteggere la privacy ma anche migliorare il benessere emotivo dei pazienti e delle loro famiglie.[127]

Questi casi illustrano come, se usati eticamente e responsabilmente, i Deepfake possano aprire nuove strade e offrire soluzioni innovative in ambiti culturali, terapeutici e sociali.

[127] Z. Yang "*I deepfake dei cari defunti sono un business cinese in piena espansione*" (7 maggio 2024) MIT Technology Review <technologyreview.it >

6 risvolti commerciali

.1 introduzione alle pratiche commerciali influenti sulla volontà

L'intelligenza artificiale (IA) sta rivoluzionando il marketing e l'influenza sui consumatori, utilizzando tecniche sofisticate per orientare le decisioni di acquisto e i comportamenti. L'analisi predittiva consente alle aziende di raccogliere e interpretare enormi quantità di dati relativi al comportamento dei consumatori, come la cronologia degli acquisti e le interazioni sui social media. In questo modo, è possibile prevedere i bisogni futuri e fornire suggerimenti mirati che aumentano la probabilità di conversione.[128] La personalizzazione dei contenuti è un'altra tecnica fondamentale, in cui piattaforme come Netflix e Amazon sfruttano algoritmi di machine learning per offrire contenuti in linea con gli interessi specifici degli utenti, aumentando il coinvolgimento e influenzando le scelte.[129] [130]

La pubblicità comportamentale (behavioral advertising) è uno degli strumenti più potenti, sfruttando dati dettagliati sulle attività online per mostrare annunci mirati. Gli algoritmi di IA analizzano sia le caratteristiche demografiche che i comportamenti di navigazione per fornire messaggi promozionali che risuonano con l'utente. Anche i chatbot e gli assistenti virtuali, programmati per guidare gli utenti durante il processo di acquisto e suggerire prodotti, giocano un ruolo cruciale. Questi strumenti rendono l'interazione con il brand più fluida e personalizzata, incentivando acquisti spesso non pianificati.[131] [132]

Un'altra tecnica è il nudging[133] digitale, teoria secondo la quale le interfacce devono essere progettate in modo tale da orientare le scelte dei consumatori, alterando

---

[128] E. Sigel "*Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* di Eric Siegel" (Febbraio 2016), John Wiley & Sons. Pp. XXI-XXIIX.

[129] Ivi Pp. 185-206

[130] Questo fenomeno è stato anche descritto come Hyper personalization in una ricerca di McKinsey. L'applicazione di questo strumento porta evidentissimi vantaggi per le aziende che lo applicano, arrivando ad un dimezzamento dei costi per l'acquisizione della clientela e un aumento del ritorno degli investimenti in marketing fino al 40% rispetto a investimenti simili senza l'applicazione di misure di personalizzazione dell'esperienza. Questo vantaggio, però obbliga le aziende al continuo aggiornamento e un'elevatissima consapevolezza degli strumenti digitali. N. Arora et alt. "*The value of getting personalization right—or wrong—is multiplying*" (12 Novembre 2021) McKinsey & company <mckinsey.com>

[131] G. D'Amico "*Behavioural Marketing e implicazioni privacy: di cosa parliamo e il ruolo delle parti in gioco*" (12 ottobre 2020) Cyber Security 360 < cybersecurity360.it/legal >

[132] Cfr Adobe Communications Team "*Behavioral targeting - what it is, why it's important, and how to do it*" (15 novembre 2022) Adobe <business.adobe.com>; per un approfondimento rispetto a quella che è la normativa europea a questi riguardi vedi: IPOL "*Regulating targeted and behavioural advertising in digital services*" (luglio 2021) parlamento europeo, su richiesta del comitato JURI.

[133] Il "nudging" è una tecnica di persuasione comportamentale che mira a influenzare le decisioni delle persone in modo prevedibile senza limitarne la libertà di scelta. Il termine è stato reso popolare dagli economisti

significativamente l'implementazione teorizzata da Thaler, il quale prevedeva un utilizzo in aiuto alla scelta etica senza privare della libertà e non lo sfruttamento dei bias di attenzione del consumatore al fine di veicolarne la scelta. Certo è che grazie alle capacità delle IA di analizzare grandissime quantità di dati, permettono di giungere prima a realizzare lo scopo più capitalistico, offrendo interfacce che rendano prominenti determinati contenuti, proponendo scelte di default o, comunque, riducendo quello che è lo sforzo mentale del consumatore.[134] [135]

In parallelo, il principio del "social proof" è rafforzato dall'IA attraverso l'analisi delle recensioni e dei feedback, suggerendo che, se molte persone scegliessero un certo prodotto, anche il singolo consumatore dovrebbe farlo. Questa logica di imitazione, supportata da contenuti che appaiono autentici e socialmente condivisi, può influenzare fortemente le decisioni.[136]

L'analisi del sentiment consente all'IA di interpretare i testi sui social media e i commenti online per cogliere le emozioni degli utenti e adattare la comunicazione di marketing per fare leva su tali emozioni, rendendo il messaggio più efficace. Gli algoritmi possono anche effettuare A/B testing automatico per ottimizzare in tempo reale la versione di un sito o una pubblicità che funziona meglio, garantendo un targeting preciso e una risposta immediata ai feedback degli utenti.[137]

Tecniche più avanzate, come i Deepfake e la generazione automatica di contenuti, possono creare messaggi personalizzati sotto forma di video o immagini che parlano direttamente ai bisogni e alle preferenze degli utenti. Queste tecnologie, benché controverse, sono estremamente efficaci nel catturare l'attenzione e rendere i messaggi pubblicitari particolarmente convincenti, si pensi anche a quanto detto durante la trattazione del § 5, con

---

Richard Thaler e Cass Sunstein nel loro libro *Nudge: Improving Decisions About Health, Wealth, and Happiness*. R. H. Thaler e C. R. Sunstein "*Nudge: Improving Decisions About Health, Wealth, and Happiness*" (2008) Penguin Books (riedito 2009).

[134] M. Jesse, D. Jannach "*Digital nudging with recommender systems: Survey and future directions*" (13 gennaio 2021) Computer in Human Behavior, report 3, <sciencedirect.com>

[135] Vedi: L. Saulle "*Che cos'è il Digital Nudging: la persuasione digitale*" (10 Dicembre 2020) PXR Italy <pxritaly.com/it >. F. Pozzi "*Digital Nudge. Larchitettura delle scelte nei contesti digitali* " (luglio 2022) Ledizioni.

[136] K. Kawon, K. Woo Gon e L. Minwoo "*Impact of dark patterns on consumers' perceived fairness and attitude: Moderating effects of types of dark patterns, social proof, and moral identity*" (ottobre 2023) Elesevier, vol 98 edizione online.

[137] Vedi: Young, Scott W. H "*Improving Library User Experience with A/B Testing: Principles and Process*" (Agosto 2014). Weave: Journal of Library User Experience. Vol. 1 (1). Kohavi, Ron; Xu, Ya; Tang, Diane "*Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*." (22 ottobre 2021) Cambridge University Press

riferimento all'utilizzazione dell'immagine di persone care ovvero di persone di spicco, come politici o attori.

Infine, la gamification[138], integrata nelle app e nei siti web grazie all'IA, incoraggia l'interazione con i consumatori attraverso sistemi di premi e progressi, spingendoli a compiere azioni specifiche, come l'acquisto o la registrazione.

A complemento di queste tecniche, il neuromarketing introduce una nuova prospettiva nel comprendere i consumatori, combinando neuroscienze e marketing per studiare le risposte inconsce agli stimoli pubblicitari. Attraverso strumenti come la risonanza magnetica funzionale (fMRI) e l'elettroencefalografia (EEG), il neuromarketing permette di analizzare l'attività cerebrale, individuando immagini, suoni o messaggi che attivano aree del cervello legate alle emozioni e alla memoria.[139] Questa disciplina si propone di andare oltre le dichiarazioni esplicite dei consumatori, fornendo agli esperti di marketing una visione più profonda delle preferenze inconsce, e si affianca così alle tecniche basate su IA per creare approcci di marketing estremamente mirati e persuasivi, che sfruttano i processi cognitivi ed emotivi umani.

Se le tecniche appena citate risultano particolarmente invasive e impossibili da eseguire senza un evidente ed esplicito consenso o una violenza, altre tecniche possono assumere vesti molto meno evidenti, si sta parlando delle tecniche di Eye tracking, Biometria del viso e Analisi del ritmo cardiaco e della respirazione. Di cui parleremo meglio nel prossimo sottocapitolo (§ 6.3).

Queste tecniche, come approfondiremo in seguito rappresentano un'arma a doppio taglio, di fatti, se da un lato migliorano l'esperienza del consumatore offrendo contenuti

---

[138] In italiano risulterebbe "Ludicizzazione", per quanto particolare questa espressione fa riferimento allo sfruttamento di meccaniche di gioco in contesti non ludici, questo tema è stato reso noto al grande pubblico nel febbraio 2010 grazie alla conferenza che Jesse Schell, game-designer americano, tenne in occasione del "*D.I.C.E. Summit*" di Las Vegas. Questo principio trae la sua forza da quello che è l'interattività connessa ai moderni mezzi di comunicazione e di e-commerce, permettendo di avere una maggiore risposta attiva rispetto al contenuto realizzato. L'attività rispetto al contenuto permette una maggiore internalizzazione dello stesso e di conseguenza una maggiore assimilazione del messaggio che si vuole trasmettere. Ovviamente, il messaggio da veicolare può essere dei più disparati, ma la tecnica è tendenzialmente simile, ovvero: l'ottenimento di punti, ricompense, raggiungere livelli ovvero ottenere emblemi o distintivi da esibire. Per approfondire ulteriormente si consiglia di consultare il sito: www.gamification.it/, ovvero, la consultazione de: J. McGonigall "*La realtà in gioco: Perché i giochi ci rendono migliori e come possono cambiare il mondo*" (marzo 2011) Apogeo.

[139] E. Harrell "*Neuromarketing: what you need to know*" (23 gennaio 2019) Harvard business review <hbr.org>

pertinenti e interazioni personalizzate, dall'altro sollevano questioni etiche riguardo alla manipolazione della volontà e alla protezione della privacy.

.2 Nudging

Il "nudging" è una tecnica derivata dall'economia comportamentale che si propone di influenzare le decisioni delle persone attraverso interventi che alterano il comportamento in maniera sottile e non coercitiva. Questa metodologia si basa sull'idea di modificare l'ambiente di scelta in modo tale da guidare gli individui verso decisioni che possono migliorare il loro benessere, senza però restringere la gamma di opzioni disponibili o modificare significativamente gli incentivi economici.

Il concetto di nudging è stato ampiamente diffuso e teorizzato da Richard H. Thaler e Cass R. Sunstein nel loro libro "Nudge: Improving Decisions about Health, Wealth, and Happiness" pubblicato nel 2008. Gli autori sostengono che attraverso piccoli cambiamenti nel modo in cui le scelte vengono presentate, è possibile 'spingere' le persone a prendere decisioni che ritengono essere nel loro migliore interesse, senza che queste si sentano costrette a una specifica scelta.

Ad esempio, nel contesto di un supermercato, posizionare frutta e verdura in posizione prevalente può naturalmente incoraggiare scelte alimentari più salutari. In ambito finanziario, impostare la partecipazione automatica ai piani di risparmio pensionistico può aumentare i tassi di risparmio tra i lavoratori. Analogamente, i promemoria per appuntamenti medici o per il rinnovo di prescrizioni sono semplici strategie di nudging che possono migliorare l'aderenza ai trattamenti sanitari raccomandati.

Nonostante la sua efficacia, il concetto di nudging non è esente da critiche, soprattutto per quanto riguarda le implicazioni etiche. I critici sollevano preoccupazioni riguardo la manipolazione sottile delle scelte individuali, interrogandosi su chi debba decidere cosa è nel

‘miglior interesse’ degli individui e in che misura questo approccio possa limitare la libertà di scelta.

Nel contesto del nudging, alcune delle questioni etiche principali riguardano la manipolazione, l’autonomia e la dignità degli individui. Mentre i sostenitori del nudging, come i paternalisti libertari, ritengono che sia moralmente permissibile influenzare le decisioni delle persone in modi prevedibili e non coercitivi, vi sono preoccupazioni significative riguardo al potenziale di manipolazione sottile che queste tecniche possono comportare.

La questione della manipolazione è cruciale, infatti, sebbene il nudging preservi la libertà di scelta, l’intenzione di influenzare le decisioni può essere vista come manipolativa e quindi eticamente discutibile. Tuttavia, sostenitori del nudging argomentano che, con trasparenza e rendicontazione adeguata, il rischio di manipolazione può essere mitigato. La chiave è che ogni intervento di nudging dovrebbe essere aperto al controllo pubblico e non nascosto.[140]

Secondo quanto scritto da Cass R. Sunstein le tecniche di nudging dovrebbero essere visibili e comprensibili per coloro che potrebbero essere influenzati da esse, consentendo così una scelta più consapevole. Questo approccio si propone di mantenere la libertà di scelta individuale mentre si cerca di guidare le persone verso decisioni che possano migliorare il loro benessere.[141]

Tuttavia, la critica al nudging si concentra sul suo potenziale paternalistico, poiché, nonostante l’intenzione positiva, potrebbe limitare l’autonomia individuale influenzando sottilmente le preferenze e le scelte senza un consenso esplicito. Questo solleva questioni etiche significative su chi decide cosa sia “migliore” per gli individui e se tali decisioni debbano essere guidate da entità governative o commerciali. Di fronte a questo contesto etico complesso, la sfida consiste nel bilanciare l’uso del nudging come strumento per il miglioramento del benessere sociale con il rispetto e la promozione dell’autonomia e della dignità individuale, attraverso una regolamentazione più rigorosa e un maggiore coinvolgimento informato delle persone interessate.[142]

---

[140] C. R. Sunstein “*The Ethics of Nudging*” (25 novembre 2021), Yale Jurnal on Reguation, Vol. 32(2), Pp. 428-429 & p.450

[141] Ivi pp. 442-449

[142] Cfr ivi pp. 449-450

È cruciale riconoscere le distinzioni tra i vari tipi di nudging per valutare le loro implicazioni etiche. Da un lato, alcuni nudging sono progettati per correggere disinformazioni o errori cognitivi, fornendo informazioni che aiutano gli individui a prendere decisioni più informate. Questi sono generalmente percepiti come meno problematici perché mirano a rafforzare l'autonomia dell'individuo migliorando la qualità delle scelte senza imporre direzioni predeterminate.[143]

D'altra parte, esistono pratiche di nudging che potrebbero essere viste come manipolative, in quanto sfruttano le vulnerabilità comportamentali degli individui, come l'inclinazione a seguire la via di minor resistenza o la suscettibilità a suggerimenti sottili senza un miglioramento effettivo della loro capacità di decisione. Questi interventi possono alterare il comportamento in modi che l'individuo potrebbe non scegliere se fosse pienamente informato o se operasse in un contesto privo di tali influenze. Questo solleva preoccupazioni etiche significative, in quanto tali tecniche potrebbero limitare l'autonomia piuttosto che promuoverla, inducendo le persone a fare scelte che non riflettono necessariamente i loro valori o obiettivi personali.

La sfida nel campo del nudging è quindi quella di bilanciare questi strumenti in modo che promuovano un'autentica autonomia e benessere, piuttosto che manipolare sottilmente le decisioni. Ciò richiede un attento esame delle intenzioni dietro ogni intervento di nudging e delle sue implicazioni etiche, oltre alla trasparenza e alla responsabilità nell'implementazione di tali strategie. In definitiva, la legittimità del nudging come strumento di policy dipende dalla sua capacità di rispettare e potenziare la capacità decisionale degli individui in modi che siano sia eticamente difendibili sia effettivamente orientati al miglioramento del benessere individuale e collettivo.

In sintesi, il dibattito sulle questioni etiche del nudging è complesso e multifaccettato, con opinioni divergenti su quanto tali tecniche possano essere giustificate nel contesto di politiche pubbliche e pratiche commerciali. L'importanza di valutare ogni caso specifico di nudging alla luce dei suoi effetti sull'autonomia e il benessere degli individui è fondamentale per determinare la sua accettabilità etica.

[143] Ivi pp.424-246

Tuttavia, quando usato responsabilmente, il nudging rappresenta uno strumento potente per le politiche pubbliche e la gestione aziendale, offrendo un modo per guidare le persone verso decisioni migliori con minore resistenza e maggiore accettazione. Questi interventi, se ben progettati, possono essere un modo efficace e rispettoso per migliorare le decisioni individuali e collettive in vari ambiti della vita quotidiana.

.3 Neuro Marketing un'introduzione alla teoria e le applicazioni con intelligenze artificiali

.1 Introduzione

Per dare una prima definizione, Herbert Simon,[144] introdusse il modello della "teoria della razionalità limitata", ribaltando le teorie razionalistiche precedenti che assumevano decisioni totalmente massimizzanti da parte degli individui. Secondo Simon, la razionalità delle persone è vincolata da limitazioni esterne come le informazioni disponibili, il tempo a disposizione e le capacità cognitive. Di conseguenza, il processo decisionale del consumatore non segue logiche perfettamente razionali, ma è spesso guidato dall'irrazionalità e dall'istinto. Le emozioni e la ragione vengono stimolate simultaneamente e la loro interazione influisce sul comportamento. Basandosi su questa visione, le moderne teorie di marketing riconoscono che l'uomo non è un essere completamente razionale ma le sue scelte sono spesso dettate dall'istinto.[145]

Il neuromarketing, una disciplina che nasce dalle fondamenta teoriche sopra descritte, integra le regole del marketing tradizionale con la medicina neurologica e le scienze

[144] Herbert Simon ha introdotto il concetto di "razionalità limitata" nel 1957, nel suo libro intitolato "*Models of Man: Social and Rational*". In questa opera, Simon sfida lidea della razionalità perfetta nellambito economico e decisionale, proponendo invece che la capacità di prendere decisioni razionali è limitata dalle informazioni disponibili, dalla capacità cognitiva dellindividuo e dal tempo disponibile per prendere la decisione. Questa teoria ha avuto un impatto profondo non solo in economia, ma anche in psicologia, scienze decisionali e nel marketing.

[145] R. Viale "*La razionalità limitata "embodied" alla base del cervello sociale ed economico.*" (2019) Il Mulino

psicologiche cognitive. Questo campo è stato definito per la prima volta nel 2002 dall'olandese Ale Smidts, che attraverso ricerche innovative ha esplorato come il cervello umano risponde inconsciamente agli stimoli commerciali.

In sostanza, questa materia di studio è un'area interdisciplinare che fonde neuroscienza e marketing per studiare come il cervello umano reagisce agli stimoli pubblicitari e come queste reazioni influenzano le decisioni di acquisto. Scoprire le risposte subconscie e le motivazioni di fondo che guidano i comportamenti dei consumatori è fondamentale, dato che le tecniche tradizionali come sondaggi e focus group spesso non riescono a catturare questi aspetti profondi.[146][147]

Il neuromarketing mira a comprendere i meccanismi inconsci dietro le decisioni di acquisto, studiando come immagini, suoni, colori e forme interagiscono con le emozioni per influenzare il comportamento del consumatore. Questa comprensione aiuta le aziende a sviluppare campagne e strategie di marketing più efficaci, progettate per attirare e influenzare i clienti, guidando così le decisioni di acquisto. David Ogilvy, un pioniere nel campo del marketing, riassume il concetto sottolineando un principio fondamentale nella pratica del neuromarketing: che le persone "non pensano quello che sentono, non dicono quello che pensano, non fanno quello che dicono".

In neuromarketing, si esplorano le emozioni che gli stimoli di marketing evocano, poiché sono cruciali nel guidare le decisioni di acquisto. Ad esempio, le emozioni positive possono aumentare la probabilità di un acquisto. Misurare l'attenzione rivolta agli elementi specifici di una campagna può rivelare cosa realmente cattura l'interesse dei consumatori, mentre l'analisi della memoria può indicare quali aspetti di un prodotto o annuncio saranno più facilmente ricordati.

Per ottenere queste informazioni, come si è già accennato nel paragrafo introduttivo al capitolo, il neuromarketing si avvale di tecniche come l'eye tracking, l'elettroencefalografia (EEG), l'imaging a risonanza magnetica funzionale (fMRI), la risposta galvanica della pelle (GSR) e la biometria del viso. Queste tecnologie scientifiche permettono di osservare l'attività

[146] Sharad Agarwal "*Neuromarketing for Dummies*" (2014), Journal of Consumer Marketing, Vol. 31 (4) Pp. 330 - 331.

[147] Vedi: P. Renvoisé & C. Morin "*Neuromarketing: Understanding the Buy Buttons in Your Customers Brain.*" (2007) Thomas Nelson Inc.

cerebrale e le risposte fisiologiche in maniera non invasiva e accurata, offrendo una presa diretta sui processi inconsci e fornendo dati preziosi per una migliore analisi.

L'utilizzo di tecniche come l'EEG, l'fMRI e la GSR nel neuromarketing offre alle aziende strumenti avanzati per comprendere le reazioni subconscie dei consumatori agli stimoli di marketing. [148] [149]

L'EEG è apprezzato per la sua capacità di registrare l'attività elettrica del cervello con alta risoluzione temporale, catturando le fluttuazioni dell'attività cerebrale che si verificano in frazioni di secondo durante l'esposizione a messaggi pubblicitari o marchi. Questa tecnica è utilizzata per valutare l'engagement cognitivo e le risposte emotive, misurando specifiche onde cerebrali che indicano stati come attenzione, eccitazione o relax. Questi dati aiutano i marketer a comprendere quali aspetti di un annuncio suscitano maggiore interesse o coinvolgimento emotivo, permettendo di ottimizzare la creatività e il contenuto delle campagne pubblicitarie.

L'fMRI, invece, offre una visualizzazione maggiormente dettagliata dell'attività cerebrale misurando i cambiamenti nel flusso sanguigno nelle diverse aree del cervello, indicando quali regioni sono coinvolte quando i consumatori prendono decisioni o reagiscono a stimoli specifici. Questo strumento è usato per analizzare come i consumatori reagiscono a livello cerebrale a prodotti, branding, pubblicità e decisioni d'acquisto, ma è meno utilizzato rispetto all'EEG a causa del suo alto costo, del disagio causato dall'utilizzo del macchinario e della necessità di un ambiente controllato per la scansione.

La GSR misura le variazioni minime nella conduttività della pelle, influenzate dallo stato emotivo della persona. Un aumento della traspirazione può indicare una maggiore eccitazione emotiva. Nel neuromarketing, la GSR è spesso utilizzata insieme all'EEG per fornire una misura completa delle reazioni emotive ai messaggi di marketing, valutando non solo l'attenzione e l'elaborazione cognitiva, ma anche l'intensità della risposta emotiva evocata da un annuncio, un prodotto o un'esperienza di marca.

Queste tecnologie offrono ai professionisti del marketing la possibilità di raffinare le loro strategie, personalizzando gli annunci in modo che risuonino più profondamente sul piano emotivo e cognitivo con il loro pubblico. La capacità di misurare le risposte inconsce aiuta a

[148] Id. nota 128

[149] R. Riccardi "*Neuromarketing e AI: Strategie per la Mente del Consumatore*" (30 maggio 2023) Università del Marketing < www.universitadelmarketing.it>

creare messaggi pubblicitari che non solo attirano l'attenzione, ma che sono anche emotivamente coinvolgenti, aumentando la probabilità di memorizzazione e di azione, come l'acquisto. Tuttavia, è essenziale utilizzare questi strumenti con responsabilità etica, garantendo trasparenza e consenso nel loro impiego per rispettare la privacy e l'integrità dei consumatori.

Se le tecniche appena citate risultano particolarmente invasive e impossibili da eseguire senza un evidente ed esplicito consenso o una violenza, altre tecniche possono assumere vesti molto meno evidenti, si sta parlando delle tecniche di Eye tracking, Biometria del viso. Queste due tecniche offrono la possibilità di avere una visione rispetto alle reazioni spontanee e alle emozioni dei consumatori di fronte a vari stimoli di marketing, consentendo alle aziende di ottimizzare le loro strategie per essere più efficaci ed emotivamente coinvolgenti.

L'eye tracking traccia il movimento degli occhi per determinare dove e per quanto tempo una persona guarda, rivelando quali elementi di un annuncio attirano più l'attenzione e quali vengono ignorati. Questi dati indicano come i consumatori leggono e interpretano le informazioni, aiutando a perfezionare la disposizione visiva degli annunci per massimizzare l'impatto e l'efficacia. Il fattore più preoccupante di questa tecnica è la semplicità di implementazione, di fatti, seppur con strumentazione realizzata ad hoc si possa avere un maggior grado di precisione, anche la fotocamera degli smart device che utilizziamo quotidianamente possiedono le capacità tecniche per supportare uno studio di questo tipo.

Riguardo alla seconda delle tecniche meno invasive, la biometria del viso utilizza software avanzati per analizzare le espressioni facciali e identificare le emozioni esposte in risposta a stimoli specifici. Queste espressioni possono indicare reazioni emotive precise, come gioia, sorpresa, o disapprovazione, e sono utili per testare l'efficacia emotiva degli annunci pubblicitari. Misurare la risposta emotiva può rivelare se un messaggio è in grado di generare l'emozione desiderata, legata a una maggiore probabilità di ricordo del brand e alla decisione di acquisto.

L'impiego combinato di queste tecniche fornisce una comprensione dettagliata di come i consumatori vedono e reagiscono agli stimoli di marketing, permettendo di creare campagne che non solo catturano l'attenzione, ma che sono anche emotivamente coinvolgenti, trasformando il modo in cui le aziende comunicano con i loro consumatori e portando a una maggiore soddisfazione del cliente e lealtà al brand. Questi strumenti, se usati in modo responsabile e etico, possono avere un impatto significativo sulle strategie di marketing.

Tuttavia, l'impiego del neuromarketing solleva questioni etiche significative, che analizzeremo in maniera più approfondita nel prossimo paragrafo (§ 6.4). Una preoccupazione maggiore è quella che possa manipolare i consumatori influenzando il loro comportamento in modi di cui non sono consapevoli. Ad esempio, il potenziale per le aziende di influenzare in modo eccessivo le decisioni dei consumatori è un punto focale del dibattito etico, poiché interroga il confine tra persuasione e manipolazione. La preoccupazione principale sta nel fatto che queste tecniche possono essere talmente incisive da ridurre l'autonomia decisionale del consumatore, creando una sorta di consumatori non pienamente consapevoli delle loro decisioni di acquisto.

.2 applicazione alle intelligenze artificiali

L'intelligenza artificiale (IA) ha rivoluzionato molti settori, e il neuromarketing non fa eccezione. L'integrazione dell'IA nel neuromarketing ha aperto nuove possibilità per strategie di marketing personalizzate ed efficaci.

Una delle principali applicazioni dell'IA nel neuromarketing è l'analisi dei dati e il riconoscimento dei modelli. Grazie alla possibilità di elaborare enormi quantità di dati raccolti attraverso attività online, interazioni sui social media e cronologia degli acquisti, gli algoritmi di IA possono identificare modelli complessi e preferenze non immediatamente evidenti agli analisti umani; questo, evidentemente permette ai marketer di creare campagne altamente mirate, adattate alle esigenze e ai comportamenti individuali dei consumatori.

Il riconoscimento delle espressioni facciali e la rilevazione delle emozioni rappresentano un altro ambito chiave in cui l'IA sta facendo passi da gigante nel neuromarketing. Le reti neurali convoluzionali (CNN)[150] sono al centro delle innovazioni nella visione artificiale, consentendo un'interpretazione più accurata ed efficiente delle risposte emotive agli stimoli di marketing. Questa tecnologia offre informazioni preziose su come le

[150] Le reti neurale convoluzionali sono reti che permettono l'utilizzo di filtri per estrarre caratteristiche dalle immagini attraverso un processo chiamato, appunto, convoluzione. Questa tecnica permette la riduzione delle immagini mantenendo le informazioni di maggiore importanza, così da permettere alla macchina di compiere scelte più o meno complesse, dal riconoscimento del contenuto alla categorizzazione delle stesse. V. Infodata, "*Il funzionamento delle reti neurali convoluzionali spiegate bene in un video*", (11 novembre 2023) il sole 24 ore, edizione digitale < www.infodata.ilsole24ore.com>

persone reagiscono emotivamente a diversi messaggi pubblicitari, aiutando le aziende a progettare esperienze più coinvolgenti e di impatto.[151]

Anche l'analisi della personalità sta diventando sempre più importante nel neuromarketing guidato dall'IA. Il modello OCEAN (Apertura mentale, Coscienziosità, Estroversione, Amicalità, Nevroticismo) viene spesso utilizzato per prevedere le risposte dei consumatori agli annunci pubblicitari. Analizzando i tratti della personalità di un individuo, i marketer possono creare messaggi persuasivi che risuonino con segmenti di pubblico specifici. Questo livello di personalizzazione consente un targeting più efficace e un aumento dell'efficacia delle campagne[152].

L'integrazione dell'IA nel neuromarketing offre numerosi vantaggi, a partire da una migliore comprensione del comportamento dei consumatori.

Grazie agli strumenti basati sull'IA, è possibile analizzare grandi quantità di dati per scoprire modelli nascosti e preferenze che i metodi tradizionali spesso non riescono a rilevare. Inoltre, l'IA consente una personalizzazione in tempo reale, permettendo alle aziende di adattare le loro strategie di marketing ai singoli clienti, creando esperienze più pertinenti e coinvolgenti. Un altro aspetto fondamentale è la capacità di migliorare l'analisi predittiva, poiché gli algoritmi di IA possono prevedere con elevata precisione il comportamento dei consumatori, aiutando le aziende ad anticipare le tendenze di mercato e prepararsi ad affrontarle. L'IA rappresenta anche una soluzione economica, poiché offre alternative meno costose rispetto a strumenti neuroscientifici come le scansioni fMRI, garantendo comunque informazioni di valore. Infine, i sistemi di IA assicurano una straordinaria scalabilità, permettendo di elaborare e analizzare dati su larga scala e di applicare tecniche di neuromarketing a un pubblico globale in modo rapido ed efficiente.[153]

---

[151] Cfr R. Montinaro "*Riconoscimento Delle Emozioni E Marketing Personalizzato*", Persona e mercato, 2024, n. 3, Pp. 847 - 853.

[152] Questo processo forma quella che viene definita come Digital Profile, ovvero un aggregato di dati che pertiene al collettivo, permettendo di realizzare insiemi di individui in ragione delle loro caratteristiche sociografiche, psicografiche, e alle loro abitudini di consumo. Più correttamente bisognerebbe fare riferimento a "rappresentazioni del profilo" – in quanto raggruppamento di rappresentazioni individuali – sotto forma di dati e informazioni. La formazione e l'aggiornamento di questi profili è automatico e tendenzialmente segue processi sconosciuti alla maggior parte degli utenti, a differenza di quella che è definita come Digital Persona, ovvero quell'insieme di dati che si diffonde volontariamente, nonostante il livello di attenzione a questo rilascio possa essere di gran lunga differente in base al fatto che si discuta di tracce GPS o contenuti su piattaforme Social. A. Roosendaal "*Digital Personae and Profiles as Representations of Individuals"*, in M. Bezzi et al. "*Privacy and Identity Management for Life*, Berlin" (2010) Springer Science & Business Media.

[153] Idem nota 131

Diverse sono le aziende che hanno implementato strategie di neuromarketing con una componente di intelligenza artificiale, per portare degli esempi, la piattaforma Netflix analizzando quella che è la cronologia di visualizzazione suggerisce quelli che potrebbero essere programmi di futuro interesse. Similmente, ma con un quantitativo di informazioni molto maggiore Amazon, sfruttando avanzati algoritmi riesce a suggerire contenuti e prodotti correlati all'interno di tutto il suo ecosistema, quindi non limitandosi al settore Prime Video, ma applicando il profilo digitale realizzato per l'utente, in modo trasversale tanto all'audiovisivo, quanto all'e-commerce, permettendo una migliore capacità predittiva rispetto a quello che potrebbe fare una piattaforma ristretta a quello che è un singolo campo di applicazione.[154]

L'adozione dell'IA rappresenta anche una soluzione economica. Mentre strumenti neuroscientifici come le scansioni fMRI possono essere costosi, le alternative basate sull'IA offrono modi più accessibili per raccogliere informazioni simili. Ad esempio, Amazon applica l'IA generativa in attività come la sintesi delle recensioni dei prodotti, aiutando i clienti a prendere decisioni di acquisto migliori e consentendo ai partner di vendita di aumentare le vendite e ridurre i resi.[155]

L'integrazione dell'IA nel neuromarketing rappresenta un cambiamento significativo nell'industria, offrendo opportunità senza precedenti per strategie di marketing personalizzate ed efficaci. Tuttavia, come per qualsiasi tecnologia potente, emergono sfide e preoccupazioni etiche che devono essere affrontate.

Man mano che le aziende continuano a sfruttare il potenziale delle intuizioni guidate dall'IA è fondamentale stabilire quadri normativi che diano priorità ai diritti dei consumatori, favoriscano la trasparenza e garantiscano un uso responsabile dei dati. Bilanciando l'innovazione con le considerazioni etiche, le aziende possono creare esperienze di marketing più coinvolgenti, affidabili e, in definitiva, di maggiore successo per i consumatori.

.4 Pratiche scorrette di influenza e dubbi etici

Come discusso in precedenza, il neuromarketing, che fonde neuroscienza e marketing, suscita un acceso dibattito etico sin dalla sua nascita. Questo particolare campo ha, inoltre, visto

[154] ibidem

[155] Staff di About amazon "*Amazon introduce nuovi strumenti basati sull'IA generativa per i partner di vendita europei*" (20 giugno 2024) about amazon < www.aboutamazon.it>

una rivoluzione con l'introduzione dell'intelligenza artificiale (IA), fatto che sicuramente non ha ridotto le questioni e i dubbi sollevati in precedenza; tutt'altro, l'implementazione di queste tecnologie è stato percepito come una preoccupante invasione della privacy e un rischio per il libero arbitrio e determinazione dei consumatori.

Queste tecnologie, infatti, possono analizzare dettagliatamente impressionanti quantità di dati biometrici e neuropsicologici, offrendo sguardi approfonditi senza precedenti sul comportamento dei consumatori. Viste le impressionanti potenzialità di tali tecnologie, emergono anche serie preoccupazioni etiche, seppur alcuni allarmismi possano essere considerati, alle volte, eccessivi.

In questo paragrafo si vogliono meglio esplorare le problematiche chiave legate alla privacy, manipolazione, disuguaglianze, impatti sociali, e la mancanza di trasparenza nel neuromarketing AI, sottolineando la necessità di un intervento regolamentare e etico rigoroso.

Il primo argomento che si vuole trattare riguarda le questioni sollevate rispetto la privacy e il consenso. Di fatti la capacità di analizzare le reazioni subconscie ai prodotti pubblicitari senza un consenso pienamente informato minaccia la confidenzialità dei dati personali. Le implicazioni di tali pratiche sono profonde, poiché la raccolta di dati senza un adeguato consenso espone gli individui a rischi di privacy non consapevoli e potenzialmente dannosi.[156]

Di seguito a quanto appena detto, supponendo che la violazione dei dati sia in atto, è facilmente comprensibile come l'intelligenza artificiale applicata alle teorie di neuromarketing possa identificare e sfruttare le vulnerabilità psicologiche umane così da modellare subconsciamente comportamenti e decisioni, comprimendo, o per assurdo impedendo, l'espressione del principio dell'autodeterminazione della persona.

Questo solleva ovvie, seppur tendenti all'allarmismo ingiustificato, questioni etiche sulla giustizia delle pratiche di marketing e sulla protezione dell'autonomia individuale. La

[156] Questa evenienza è stata presa in gran contro anche dai legislatori europei, di fatti, come si parlerà in seguito, le questioni relative alla privacy e al consenso informato sono state trattate estensivamente all'interno dell'GDPR e in via traversa anche nell'Ai Act. Cfr E. R. Murphy, J. Illes & P.B. Reiner "*Neuroethics of neuromarketing.*" (2008) Journal of Consumer Behaviour.

manipolazione subconscia minaccia il concetto di libero arbitrio, ponendo in dubbio la legittimità di tali tecniche sotto il profilo etico. [157]

Il neuromarketing, quando associato a messaggi subliminali, solleva ulteriori preoccupazioni riguardo alla violazione del consenso informato e alla manipolazione del comportamento del consumatore. I messaggi subliminali, per loro stessa natura, operano al di sotto della soglia di coscienza dell'individuo, bypassandone la capacità di elaborazione consapevole e di critica, e risultano particolarmente insidiosi nel contesto del neuromarketing, dove l'obiettivo è influenzare le decisioni di acquisto senza che il consumatore sia pienamente consapevole di essere manipolato. L'AI Act vieta espressamente l'uso di tecniche subliminali oltre la coscienza di una persona quando l'obiettivo o l'effetto è di distorcere materialmente il comportamento di un individuo, limitandone la capacità di prendere decisioni informate e causandogli un danno significativo. Questo divieto sottolinea la gravità della manipolazione che tali tecniche possono esercitare sulla volontà del consumatore, minando il principio fondamentale del consenso informato.[158]

Il consenso informato presuppone che l'individuo sia pienamente consapevole delle implicazioni delle proprie scelte e delle informazioni a cui acconsente. Nel caso di messaggi subliminali, questa consapevolezza è assente, poiché il consumatore non è in grado di percepire o comprendere il contenuto del messaggio, né tantomeno la sua influenza sul proprio comportamento. Di conseguenza, qualsiasi forma di consenso rilasciato in tali circostanze non può essere considerata valida o genuina, in quanto viziata dalla mancanza di informazione e dalla manipolazione occulta.

Il neuromarketing, impiegando strumenti che misurano le risposte fisiologiche e neurali, mira a comprendere le reazioni inconsce ai messaggi di marketing. Quando queste tecniche sono combinate con messaggi subliminali, il rischio di una distorsione comportamentale aumenta significativamente. Il consumatore diventa un bersaglio vulnerabile, incapace di difendersi da una forma di persuasione che opera al di fuori della sua consapevolezza. Questa combinazione di tecniche può portare a una manipolazione inaccettabile della volontà, in cui le

[157] E. R. Murphy, J. Illes & P.B. Reiner "*Neuroethics of neuromarketing*." (2008) Journal of Consumer Behaviour vol: 7(4-5) Pp. 293 – 302.

[158] E. Tuccari "*Neuromarketing: Un'asistematica Disciplina … Oltre Il Consenso?* "(14 ottobre 2024) Persona E Mercato, vol. 2 Pp. 511 - 537

decisioni di acquisto non sono più il frutto di una scelta libera e informata, ma di una risposta condizionata a stimoli non percepiti a livello cosciente.

La problematica si estende alla difficoltà di individuare il confine tra persuasione accettabile e manipolazione inaccettabile. Mentre il marketing tradizionale mira a influenzare il comportamento attraverso la comunicazione, il neuromarketing con messaggi subliminali rischia di superare tale confine, compromettendo la libertà di scelta del consumatore e violando il principio di libero arbitrio. La mancanza di trasparenza, tipica dei messaggi subliminali, aggrava ulteriormente questa problematica, rendendo difficile per il consumatore esercitare i propri diritti e tutelarsi da pratiche commerciali sleali. La profilazione degli utenti, ottenuta con l'analisi delle loro reazioni fisiologiche e neurali, può essere utilizzata per inviare messaggi subliminali personalizzati che sfruttino le loro vulnerabilità, aumentando il rischio di manipolazione e violazione del consenso.[159]

La focalizzazione del targeting permesso dal neuromarketing rinforzato dall'utilizzo di AI può esacerbare le disuguaglianze sociali, influenzando in modo sproporzionato individui o gruppi vulnerabili. Queste tecnologie potrebbero, essere utilizzate per indirizzare prodotti o servizi a persone con minore capacità critica o maggiore suscettibilità, intensificando le disuguaglianze esistenti e manipolando i gruppi più a rischio [160], ovvero riaffermare quella che è una convinzione del soggetto proponendo conferme della propria teoria senza permettere "facile" accesso a quelle che possono essere punti di vista in contrasto o che possano contraddire una tesi, limitando quello che è il pensiero critico e lo sviluppo dell'idea.

La maggiore preoccupazione rispetto a questo punto è da rilevare nell'eventualità che l'intelligenza artificiale perpetri e alimenti pregiudizi o Bias esistenti nei dati utilizzati per addestramento del modello, portando a un targeting ingiusto o alla discriminazione.

---

[159] La direttiva 2005/29/CE ha introdotto per la prima volta una distinzione giuridica tra influenza comportamentale (ammessa se esercitata con "diligenza professionale") e distorsione comportamentale (vietata in quanto contraria a tale diligenza). In particolare, l'articolo 5 della direttiva sancisce un divieto generale di pratiche commerciali che alterino il comportamento economico dei consumatori, qualificandole come sleali.

Nel contesto del neuromarketing, questa distinzione apre diverse questioni: non esiste infatti una risposta univoca per stabilire quando l'impiego di tali tecniche superi il confine tra semplice influenza e vera e propria distorsione del comportamento. La valutazione va condotta caso per caso, per comprendere se, in una data circostanza, un determinato strumento o prodotto finisca col violare i principi stabiliti dalla direttiva. Proprio a sostegno di questa valutazione, la stessa 2005/29/CE elenca alcune pratiche commerciali sempre considerate sleali, che si aggiungono al divieto di carattere generale.

[160] A. Smidts, A. T. H. Pruyn & C. B. M. Van Riel, "*The impact of consumer advertising on product awareness and perception*" (2001) Academy of Management Journal vol: 49(5) Pp.1051-1062.

In aggiunta a tutto ciò la focalizzazione sull'efficacia del neuromarketing potrebbe portare le aziende a privilegiare gli investimenti in marketing a scapito dell'innovazione di prodotto o della responsabilità sociale. Questa tendenza potrebbe avere effetti distorsivi sui valori sociali, promuovendo un modello di consumo non sostenibile e eticamente discutibile.[161]

La complessità delle AI impiegate nel neuromarketing rende difficile per i consumatori, ma anche gli utenti in generale, e ai regolatori comprendere come i dati vengano utilizzati o come le decisioni vengano prese. Questo deficit di trasparenza impedisce un efficace controllo regolamentare e riduce la fiducia del pubblico nelle pratiche di mercato.[162] [163]

Le questioni sollevate dall'uso del neuromarketing e dell'intelligenza artificiale necessitano di un'urgente attenzione regolamentare e di un dibattito pubblico approfondito. È imperativo implementare leggi più rigorose, standard etici elevati, e garantire una maggiore trasparenza per proteggere i diritti e la dignità dei consumatori. Solo attraverso un impegno congiunto di aziende, legislatori, esperti etici e il pubblico, sarà possibile bilanciare i benefici del neuromarketing con le necessarie garanzie etiche e sociali.

In definitiva, il futuro dell'IA nel neuromarketing dipenderà dalla capacità di bilanciare l'avanzamento tecnologico con la responsabilità etica. Mentre questo campo continua a evolversi, è essenziale dare priorità alla trasparenza, alla privacy dei consumatori e alla concorrenza leale per garantire che il neuromarketing guidato dall'IA migliori l'esperienza del cliente senza compromettere l'autonomia individuale o i valori sociali.

[161] G. Zaltman "*How customers think: Essential insights into the mind of the market.*" (2003). Harvard Business Press.

[162] R. M. S. Wilson, J. Gaines & R. P. Hill "*Neuromarketing and consumer free will*." (2008) Journal of Consumer Affairs. vol: 42(3) Pp.389-410

[163] Ivi nota 149

7 Legislazione riguardante l'intelligenza artificiale e i Deepfake

.1 Fonti europee

Considerata la novità della materia l'Unione Europea, nel corso degli anni ha intrapreso un percorso ambizioso e articolato per regolamentare e sviluppare l'intelligenza artificiale (IA). La spinta proattiva delle leggi e regolamenti europei, nonostante ancora poco prevedano rispetto alle intelligenze artificiali generative, già permettono di poter applicare le stesse normative anche a questo particolare e avanzato campo.

Questo impegno si concretizza attraverso una serie di iniziative legislative e strategiche, come il Regolamento sull'IA (AI Act)[164], le Linee Guida Etiche per un'IA Affidabile, il Regolamento Generale sulla Protezione dei Dati (GDPR)[165] e il Piano Coordinato sull'Intelligenza Artificiale aggiornato in ultima battuta nel 2021.

Questi strumenti normativi e strategici, interconnessi tra loro, hanno l'obiettivo di assicurare che l'IA venga sviluppata e utilizzata rispettando i diritti fondamentali, promuovendo la trasparenza e garantendo la responsabilità.

Il percorso dell'UE per regolamentare le IA generative è parte integrante di questa strategia più ampia. Attraverso il GDPR, l'AI Act e le Linee Guida Etiche, l'Unione affronta le sfide specifiche poste da queste tecnologie, che includono la capacità di creare contenuti autonomamente. L'obiettivo è promuovere un uso sicuro e responsabile dell'IA, prevenendo i rischi etici e garantendo la protezione dei diritti umani.

Uno dei principali elementi di questo approccio è la Proposta di Regolamento sull'Intelligenza Artificiale (AI Act), che comprende anche le IA generative. Questo regolamento, che a breve analizzeremo più nel dettaglio, mira non solo a regolamentare l'uso sicuro dell'IA, ma anche a creare un ambiente che favorisca l'innovazione, la trasparenza e la protezione dei diritti fondamentali. A supporto di queste normative, le Linee Guida Etiche per

[164] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024, entrato in vigore il 1° agosto 2024

[165] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio

un'Intelligenza Artificiale Affidabile offrono un quadro di principi per assicurare che tutte le tecnologie IA, comprese quelle generative, siano sviluppate in modo etico e trasparente.

Le normative europee sull'IA generativa sono ancora in evoluzione, ma fanno parte di un quadro normativo più ampio che regola l'intelligenza artificiale nel suo complesso, concentrandosi su temi cruciali come trasparenza, etica, responsabilità e la protezione dei diritti fondamentali. Tra gli strumenti chiave, troviamo l'AI Act, le Linee Guida Etiche e il GDPR, tutti documenti che costituiscono una solida base normativa applicabile anche alle IA generative. Queste tecnologie, capaci di generare contenuti autonomamente, rappresentano sfide significative, soprattutto per quanto riguarda i rischi etici, la trasparenza e i possibili impatti sociali, rendendo ancora più importante un approccio regolamentare ed etico ben strutturato.

Orbene, vista tutta la premessa introduttiva a questo paragrafo pare il caso di analizzare più nello specifico quello che è il regolamento europeo di maggior impatto rispetto al campo dell'intelligenza artificiale: il cosiddetto AI act.

Il Regolamento sull'Intelligenza Artificiale (AI Act), proposto dalla Commissione Europea nel 2021, affronta le sfide poste dalle IA generative all'interno di un quadro più ampio di regolamentazione delle tecnologie di intelligenza artificiale. Questo regolamento introduce una classificazione basata sui rischi, suddividendo i sistemi di IA in base al loro potenziale impatto su aspetti fondamentali come i diritti umani, la sicurezza e la società.

In particolare, le IA generative, utilizzate per automatizzare processi complessi o creare contenuti informativi autonomamente, possono essere classificate come sistemi ad alto rischio. Ciò accade soprattutto quando queste tecnologie vengono applicate in ambiti critici che possono influenzare profondamente la vita delle persone, come la gestione di servizi essenziali o la produzione di informazioni.

L'AI Act si distingue per essere la prima proposta legislativa in Europa che tenta di regolamentare in modo quasi completo l'uso e la commercializzazione di tecnologie IA, comprese le IA generative. Introduce una classificazione del rischio che distingue tra sistemi a rischio inaccettabile, alto, limitato e minimo, con l'intento di garantire che le IA, soprattutto quelle ad alto impatto, siano soggette a normative più rigorose.

Le IA generative, proprio per la loro capacità di influenzare la privacy e la sicurezza, rientrano spesso tra i sistemi ad alto rischio, specialmente quando impiegate in settori che possono incidere sui diritti fondamentali.

Il regolamento mira a stabilire un quadro giuridico uniforme per garantire che lo sviluppo e l'uso dell'intelligenza artificiale in Europa siano sicuri, responsabili e rispettosi dei valori etici e delle normative sulla protezione dei dati, come il GDPR.

Questo detto l'AI Act introduce una serie di obblighi specifici per i sistemi di IA generativa, con l'obiettivo di garantire trasparenza, supervisione umana e una valutazione rigorosa dei rischi etici. I punti principali riguardano una serie Obblighi per i Sistemi di IA Generativa nell'AI Act, ovvero:

1. L'identificazione dell'IA: Tutti i sistemi di IA generativa devono essere chiaramente identificabili. Questo significa che, quando interagiscono con gli esseri umani o generano contenuti, come testi, immagini o video, devono essere segnalati come prodotti da un'IA. L'obiettivo è garantire che gli utenti siano consapevoli di quando stanno interagendo con un contenuto generato artificialmente, riducendo così il rischio di disinformazione o manipolazione.
2. La valutazione del rischio: Gli sviluppatori di IA generativa sono tenuti a condurre valutazioni dettagliate sui rischi etici e sociali associati all'uso del loro sistema, come la creazione di Deepfake o altri contenuti manipolativi. Devono prevedere meccanismi di monitoraggio continuo per identificare e mitigare i rischi emergenti durante l'uso del sistema.
3. Supervisione umana: Uno dei requisiti fondamentali è che le IA generative siano sempre soggette a supervisione umana. Questo è particolarmente importante in settori critici come la salute, la finanza e la giustizia, dove le decisioni prese dai sistemi IA possono avere un impatto diretto sulla vita delle persone. L'intervento umano serve a garantire che le decisioni automatizzate siano corrette e non causino danni o ingiustizie.

Visti gli obblighi che il regolamento prevede, si rende necessario individuare e classificare i sistemi IA in base al loro livello di rischio. Per quello che a noi concerne le IA generative possono essere considerate particolarmente critiche, ovvero, per utilizzare le parole del legislatore, sistemi ad alto rischio, è obbligatorio, quindi, condurre una valutazione

d'impatto etico che identifichi i rischi legati alla disinformazione, alla creazione di contenuti manipolativi (come i Deepfake) e alla possibilità di amplificare pregiudizi esistenti. Questa valutazione serve a garantire che i sistemi generativi rispettino gli standard di sicurezza e privacy europei ed è di particolare rilevanza in settori come: l'occupazione (IA utilizzate per automatizzare il reclutamento del personale o per gestire decisioni sull'impiego), la finanza (IA usate per valutazioni creditizie o per la concessione di prestiti), la sanità (IA impiegate per diagnosticare malattie o raccomandare trattamenti). In tutti questi casi, le IA generative richiedono rigorosi controlli per evitare errori o pregiudizi che possano avere impatti negativi sulla vita delle persone.

Oltre alla previsione di questa classificazione e valutazione, uno degli obblighi di sistema principali imposti dall'AI Act riguarda la trasparenza. Gli utenti devono sapere chiaramente quando stanno interagendo con un sistema di IA generativa. Ad esempio, se un articolo di notizie o un'immagine è stato generato da un'IA, deve essere segnalato in modo evidente. Questo obbligo aiuta a prevenire l'uso improprio della tecnologia, come la creazione di contenuti ingannevoli o falsi.

Dal punto precedente deriva direttamente quello della spiegabilità dei processi decisionali dietro alle decisioni prese dai sistemi di IA generativa, di fatti questi devono essere comprensibili agli utenti. Questo significa che gli sviluppatori devono essere in grado di spiegare in modo chiaro come il sistema è stato addestrato, quali dati sono stati utilizzati e su quali criteri si basano le decisioni o i contenuti generati.

Da ultimo il regolamento richiede che le decisioni prese da sistemi di IA generativa siano sempre sottoposte a supervisione umana, in particolare in settori critici come la giustizia e la sanità. Questo è essenziale per evitare errori o decisioni che potrebbero causare danni o essere ingiuste. La supervisione umana serve a garantire che, anche se il sistema IA genera contenuti o prende decisioni in modo autonomo, un essere umano possa intervenire quando necessario.

In generale, come già detto, le IA generative sono considerate sistemi ad alto rischio se utilizzate per automatizzare decisioni che possono influenzare direttamente la vita delle persone, come le assunzioni, la concessione di prestiti o la valutazione del merito creditizio. Per queste tecnologie, l'AI Act impone obblighi di trasparenza, supervisione umana e valutazione del rischio per prevenire abusi o danni derivanti da un uso improprio.

Conclusa l'analisi di questo regolamento, veniamo a trattare di un ulteriore documento prodotto dal legislatore europeo, ovvero: le linee guida per un'intelligenza artificiale affidabile.[166]

Le Linee Guida Etiche per un'Intelligenza Artificiale Affidabile, pubblicate nel 2019 dal Gruppo di esperti ad alto livello sull'IA della Commissione Europea, delineano un quadro di riferimento per lo sviluppo di IA, inclusa quella generativa, che rispetti i diritti umani e promuova trasparenza, equità e responsabilità.

Meglio analizzando questi principi fondamentali troviamo nuovamente caratteri come la trasparenza e spiegabilità, concetti che sono stati integralmente ripresi dall'AI act, ma anche elementi distinti che non sono entrati a far parte di quel regolamento, come la non discriminatorietà delle AI e la responsabilità degli sviluppatori rispetto a quelli che sono eventi dannosi o scopi illeciti per i quali l'intelligenza potrebbe essere utilizzata, quindi la supervisione umana e la gestione dei rischi etici restano elementi centrali per evitare abusi della tecnologia, punto, questo, fortemente rimarcato dal successivo regolamento.

Riguardo al primo punto "nuovo", anche in ossequio a quanto prescritto dalla carta dei diritti fondamentali dell'uomo, le IA devono essere addestrate su dataset diversificati ed equi, per evitare la perpetuazione di pregiudizi e discriminazioni. Questo è fondamentale, poiché modelli addestrati su dati non rappresentativi possono produrre risultati distorti o discriminatori, soprattutto nelle decisioni automatizzate che influenzano anche molto significativamente la vita delle persone.

In sintesi, queste linee guida forniscono un quadro solido per assicurare che le IA, anche quelle generative, siano progettate e utilizzate in modo etico, trasparente e inclusivo. Attraverso la trasparenza, la non discriminazione e la responsabilità, la Commissione Europea stabilisce che le IA devono essere strumenti affidabili e sicuri, capaci di promuovere benefici senza compromettere i diritti fondamentali.

Ancora più risalente è il GDPR, entrato in vigore nel 2018, è uno dei regolamenti europei più rilevanti per la protezione dei dati personali, con un impatto significativo sull'IA. Poiché queste tecnologie utilizzano grandi quantità di dati per creare contenuti autonomi, l'uso

[166] Aa. Vv., "*Ethics guidelines for trustworthy ai*" (8 aprile 2019) Commissine europea

di dati personali deve rispettare rigorosi requisiti di protezione della privacy. Gli aspetti chiave del GDPR per le IA generative includono:

1. Diritto di accesso: Gli utenti possono sapere se e come i loro dati personali vengono utilizzati per addestrare modelli di IA generativa e possono richiederne l'accesso.
2. Diritto alla cancellazione: Gli utenti possono richiedere la cancellazione dei propri dati dai dataset usati per addestrare modelli generativi, in conformità con il diritto all'oblio (articolo 17 del GDPR).
3. Decisioni automatizzate: Gli individui hanno il diritto di non essere soggetti a decisioni automatizzate basate solo su trattamenti algoritmici, inclusi quelli delle IA generative, quando tali decisioni hanno effetti significativi o legali[167].

Il GDPR richiede inoltre il consenso esplicito per l'uso di dati personali nei modelli di IA generativa. Questo consenso deve essere chiaro e preventivo, soprattutto per i sistemi che generano immagini o contenuti testuali basati su dati personali. In questo contesto, il, già ampiamente trattato, principio di trasparenza è cruciale: le organizzazioni devono fornire informazioni chiare e accessibili sul trattamento dei dati, spiegando agli utenti come e perché i loro dati sono utilizzati.

Ultimo prodotto dell'attività europea di cui si intende trattare è il Piano Coordinato sull'Intelligenza Artificiale. Nonostante sia stato aggiornato l'ultima volta nel 2021 resta una strategia chiave adottata dall'Unione Europea in collaborazione con gli Stati membri per promuovere una gestione condivisa e sicura dello sviluppo dell'IA, con particolare attenzione alle tecnologie emergenti.

Il piano mira a creare una sinergia tra le politiche nazionali e le risorse europee per accelerare l'adozione di IA etica e responsabile in tutta Europa. Di seguito i punti principali:

1. Collaborazione tra Stati membri: Il piano favorisce la cooperazione tra gli Stati membri, incentivando lo scambio di conoscenze e la condivisione delle migliori pratiche. Ciò è cruciale per lo sviluppo sicuro ed etico dell'IA generativa, evitando frammentazioni normative che potrebbero ostacolare l'innovazione. La creazione di una rete di esperti europei è centrale per promuovere un approccio coerente all'adozione di queste tecnologie.

[167] Articolo 22 del GDPR

2. Norme etiche comuni: Uno degli obiettivi centrali del piano è l'adozione di standard etici condivisi, specialmente per affrontare le sfide dell'IA generativa in termini di trasparenza e responsabilità. Il piano sottolinea l'importanza di garantire che tali sistemi rispettino i diritti umani e che il loro funzionamento possa essere spiegato in modo comprensibile e trasparente.
3. Infrastrutture comuni: Il piano prevede lo sviluppo di infrastrutture comuni, come piattaforme di test e sperimentazione, per fornire un ambiente regolato dove ricercatori e sviluppatori possano testare le tecnologie IA, comprese quelle generative. Questo ecosistema di supporto consente di sperimentare in sicurezza, in conformità con gli standard etici e tecnici dell'UE.
4. Settori strategici: Il piano si concentra su settori cruciali come la sanità, la ricerca scientifica e la gestione pubblica, dove l'IA generativa può avere un impatto significativo, ad esempio nella creazione di modelli predittivi per la ricerca medica o nell'automazione di processi amministrativi. L'UE incoraggia lo sviluppo di IA in questi ambiti, sempre nel rispetto delle normative sulla protezione dei dati e dei principi etici fondamentali.
5. Finanziamenti e incentivi: Per supportare l'attuazione del piano, l'UE ha destinato fondi e incentivi finanziari, come il programma Horizon Europe, che prevede finanziamenti specifici per progetti sull'IA generativa e altre tecnologie emergenti. Questo programma promuove la ricerca e l'innovazione, facilitando lo sviluppo di soluzioni IA etiche e sicure.

In sintesi, il Piano Coordinato sull'Intelligenza Artificiale promuove la collaborazione tra gli Stati membri e l'adozione di norme etiche comuni, garantendo che l'IA generativa venga sviluppata e utilizzata in modo sicuro, etico e conforme ai valori europei. Il piano sostiene anche la creazione di infrastrutture comuni e incentivi finanziari, rendendo l'Europa un leader nello sviluppo di tecnologie IA affidabili e trasparenti.

In conclusione, di questa prima parte del paragrafo, L'Unione Europea sta adottando un approccio regolamentato e olistico nei confronti dell'intelligenza artificiale generativa, combinando strumenti come il Regolamento sull'IA, le Linee Guida Etiche e il Piano Coordinato sull'IA. L'obiettivo principale è garantire che lo sviluppo e l'uso di queste tecnologie avvengano nel rispetto dei diritti fondamentali, della trasparenza e della sicurezza. Le IA generative, pur offrendo un enorme potenziale innovativo, devono essere soggette a

rigorosi controlli per prevenire disinformazione, pregiudizi e manipolazione dei contenuti, mantenendo sempre il controllo umano.

L'UE ha creato un quadro normativo rigoroso per regolamentare le IA, e di conseguenza le IA generative, assicurando trasparenza, responsabilità e rispetto dei diritti fondamentali, mirando a garantire l'uso responsabile delle IA generative, assicurando la protezione dei dati personali e il rispetto della trasparenza. Inoltre, il Piano Coordinato sull'IA incoraggia la collaborazione tra Stati membri e istituzioni per creare un ecosistema in cui le IA generative possano svilupparsi senza compromettere la sicurezza e la fiducia del pubblico.

Complementari alla normativa predetta sono il "Digital Markets Act"[168] e il "Digital Services Act"[169]; questi, sono due regolamenti dell'Unione Europea che affrontano le sfide del mercato digitale.

Il "Digital Markets Act" si concentra sul contrastare il potere eccessivo delle grandi piattaforme online, chiamate "gatekeeper", per garantire un mercato digitale equo e contendibile. Questo regolamento si preoccupa principalmente di pratiche commerciali che potrebbero soffocare la concorrenza, l'utilizzo dei dati in modo sleale e la creazione di barriere all'uscita per gli utenti.

Il "Digital Services Act" mira a creare un ambiente online più sicuro e trasparente, affrontando la diffusione di contenuti illegali e dannosi e proteggendo i diritti fondamentali degli utenti. Questo regolamento si applica a un'ampia gamma di prestatori di servizi intermediari, dalle piattaforme di social media ai siti di e-commerce, e stabilisce una serie di obblighi di diligenza per garantire un ambiente online più responsabile.

Entrambi i regolamenti riconoscono l'importante ruolo dell'intelligenza artificiale (IA) nel plasmare l'ambiente digitale. L'IA può essere utilizzata sia per scopi positivi, come la personalizzazione dei servizi e la lotta ai contenuti illegali, sia per scopi negativi, come la manipolazione del comportamento e la diffusione di disinformazione.

In sintesi, il "Digital Markets Act" e il "Digital Services Act" rappresentano un passo importante verso la regolamentazione del mercato digitale e la creazione di un ambiente online

[168] Regolamento (UE) 2022/1925 del parlamento europeo e del consiglio del 14 settembre 2022
[169] Regolamento (UE) 2022/2065 del parlamento europeo e del consiglio del 19 ottobre 2022

più equo, trasparente e sicuro. L'obiettivo è promuovere l'innovazione e la crescita economica, garantendo al contempo la protezione dei diritti fondamentali degli utenti, compresa la protezione dalla manipolazione e l'accesso a un'informazione libera e pluralistica.

.2 ONU, UNESCO e OCSE: l'impegno delle associazioni internazionali

Volendo rivolgere lo sguardo a quanto ci riferiscono le associazioni internazionali le IA generative rappresentano una delle sfide etiche più complesse e urgenti nel panorama tecnologico attuale. Sono sicuramente da citare la risoluzione ONU A/RES/73/27, adottata nel 2018, e la raccomandazione UNESCO riguardo l'etica di utilizzo dell'intelligenza artificiale[170].

La risoluzione ONU e la Raccomandazione UNESCO si intrecciano nei loro obiettivi di garantire che lo sviluppo e l'uso delle tecnologie emergenti, come l'intelligenza artificiale generativa, avvengano in un quadro etico e di sicurezza globale. comprese le IA generative.

Sebbene la Raccomandazione UNESCO sull'etica dell'intelligenza artificiale non sia specificamente focalizzata sulle IA generative, essa pone le basi per affrontare le sfide etiche e di sicurezza che derivano dall'uso delle tecnologie avanzate, sottolineando vari principi etici che si applicano anche a queste tipologie di IA. Sollevando questioni che toccano direttamente la natura dell'innovazione tecnologica e la sua relazione con la responsabilità umana e sociale.

La risoluzione ONU A/RES/73/27, intitolata "Developments in the field of information and telecommunications in the context of international security", affronta la crescente preoccupazione per l'uso delle tecnologie avanzate, tra cui l'IA, che potrebbero minacciare la sicurezza internazionale. Mentre la risoluzione sottolinea la necessità di misure di fiducia, cooperazione internazionale, e sviluppo di norme per garantire che le tecnologie non siano

[170] AA.VV., "*Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence*", (2023), UNESCO.

utilizzate per scopi malevoli, come la destabilizzazione delle infrastrutture critiche o la diffusione di disinformazione.

Rispetto alle sfide etiche, come evidenziato nella prefazione del documento rilasciato dall'UNESCO, l'IA generativa è stata spinta nel mercato globale da interessi commerciali senza una valutazione adeguata dei rischi etici. Questo sviluppo ha portato alla proliferazione di sistemi che amplificano pregiudizi e discriminazioni, e diffondono informazioni potenzialmente false o dannose, fenomeni che sono particolarmente gravi nel contesto delle IA generative, capaci di creare contenuti falsi che sembrano autentici, come i Deepfake.

La mancanza di una regolamentazione chiara e di un'adeguata valutazione d'impatto etico ha contribuito all'adozione incontrollata di IA generativa, con implicazioni per la privacy e la fiducia degli utenti. È per questo che la Raccomandazione UNESCO insiste sulla necessità di applicare i principi etici lungo l'intero ciclo di vita di tali tecnologie, con un'attenzione particolare alle fasi di progettazione, sviluppo, distribuzione e monitoraggio dei sistemi di IA.

Uno dei pilastri della Raccomandazione è la valutazione dell'impatto etico, che è particolarmente rilevante per l'IA generativa. L'UNESCO richiede che le organizzazioni conducano valutazioni approfondite prima di implementare tecnologie IA, includendo l'identificazione di rischi come discriminazione, bias nei dati, violazione della privacy e la possibilità che il sistema generi contenuti dannosi o ingannevoli. Questi rischi sono evidenti nei modelli di IA generativa che apprendono dai dati esistenti, spesso perpetuando bias già presenti in essi.

Visto quanto detto per l'impatto etico, ulteriore aspetto cruciale, fosse anche solo per conseguenza, riguarda la trasparenza; data la complessità dei modelli di IA generativa e le discriminazioni che potrebbe mettere in atto, è fondamentale che questi sistemi compiano scelte comprensibili e spiegabili, evitando quindi intelligenze, come già accennato nel sottoparagrafo precedente, prenda scelte per le quali non sia valutabile il processo logico. Gli utenti devono essere in grado di sapere come il sistema arriva ai suoi risultati e quali dati sono stati utilizzati per addestrarlo. L'UNESCO spinge affinché vengano sviluppati strumenti e metodologie che permettano di tracciare il processo decisionale dell'IA, in modo da garantire la fiducia del pubblico.

La Raccomandazione UNESCO sottolinea anche l'importanza della supervisione umana su tutti i sistemi di IA, inclusi quelli generativi. Le decisioni prese dall'IA devono sempre rimanere sotto la responsabilità di persone fisiche o giuridiche, che devono essere in grado di monitorare e correggere eventuali errori o abusi. Questo principio è essenziale per garantire che le IA generative non vengano utilizzate in modi che potrebbero causare danni irreparabili o manipolare informazioni sensibili.

Al fine di garantire questi pilastri, la Raccomandazione invita, caldamente, all'uso di dataset diversificati e inclusivi. Di fatti un IA generativa, se addestrata su dati limitati o non rappresentativi, rischia di perpetuare e amplificare stereotipi e discriminazioni alimentando quelli che sono bias cognitivi preesistenti ovvero nati da mancanze nelle informazioni che le erano state fornite per l'addestramento.

La Raccomandazione insiste su un uso responsabile dei dati, incoraggiando pratiche che garantiscano l'equità e l'inclusione, fattori fondamentali per evitare pregiudizi sistematici; questa insistenza nasce da sviluppi recenti, i quali dimostrano l'urgenza di queste raccomandazioni. I Deepfake, come tratteremo nel prossimo capitolo (§ 5 sono già stati utilizzati per influenzare opinioni politiche e minare la fiducia del pubblico nelle informazioni. Ciò detto anche le IA generative che producono testi o immagini basate su dati preesistenti possono essere utilizzate per diffondere disinformazione, rafforzando la necessità di una supervisione etica e regolamentare.

Volendo fare un breve sunto di quanto appena detto l'approccio delineato dall'UNESCO nella sua Raccomandazione e nel documento sull'impatto etico è un chiaro richiamo all'importanza di una regolamentazione responsabile dell'IA generativa. La trasparenza, la responsabilità e l'inclusione sono principi che devono guidare lo sviluppo di queste tecnologie, affinché i benefici che l'IA può offrire non siano compromessi dai rischi etici e sociali che essa porta con sé. Il documento dell'UNESCO fornisce un quadro di riferimento prezioso per affrontare queste sfide e garantire che l'IA generativa sia sviluppata e utilizzata a beneficio di tutta l'umanità.

Fonti derivate da associazioni internazionali come l'UNESCO, l'OCSE[171] e altre organizzazioni multilaterali evidenziano l'importanza di questo approccio globale e olistico alla

[171] Vedi il sito ufficiale dell'OCSE nella sezione dedicata ai principi riferibili alle intelligenze artificiali (https://oecd.ai/en/ai-principles)

regolamentazione dell'IA generativa, con un'enfasi particolare sulla protezione dei diritti umani, la promozione dell'equità e la trasparenza, al fine di garantire uno sviluppo ed il successivo uso di queste tecnologie che possa essere sicuro e responsabile, ovvero l'adozione di strumenti che possano contribuire al benessere generale e alla sicurezza e stabilità dell'informazione.

.3 Stati uniti: Integrazione tra le leggi statali e federali in materia di IA e IA generativa

Negli Stati Uniti, la regolamentazione dell'intelligenza artificiale sta evolvendo rapidamente per affrontare le sfide e i rischi connessi a tecnologie emergenti come l'IA generativa. Le leggi della California e del Colorado, delle quali parleremo in maniera più ampia rispetto ad altre fonti, sono parte di un quadro più ampio che include leggi federali come il National AI Initiative Act e altre leggi statali rilevanti. Vediamo come queste normative si integrano e completano l'approccio legislativo verso un uso etico e trasparente dell'IA generativa.

L'Assembly Bill 2013 approvato in California, il 31 gennaio 2024, è un punto di riferimento nella regolamentazione dell'IA generativa. Essa impone obblighi di trasparenza senza precedenti, richiedendo che gli sviluppatori di IA pubblichino dettagli completi sui dati utilizzati per addestrare i sistemi. Questa legge specifica anche i requisiti per la documentazione relativa ai dati, come le fonti, la descrizione dei dati e l'eventuale presenza di materiali protetti da diritto d'autore, marchi o brevetti.

Come già riportato l'AB 2013 richiede che gli sviluppatori pubblichino documentazione dettagliata sui dati di addestramento utilizzati per costruire i sistemi di IA generativa. Questo obbligo risponde a una necessità critica di trasparenza, consentendo al pubblico di comprendere meglio come funzionano questi sistemi e quali tipi di dati influenzano i risultati prodotti. Le preoccupazioni relative ai pregiudizi (bias) nei dati di addestramento e all'uso delle informazioni personali sono centrali in questo contesto.

Procediamo, quindi, con un'analisi di quelli che sono i dettagli della Documentazione che la legge specifica. Gli elementi chiave che devono essere inclusi nella documentazione: il primo elemento riguarda, sicuramente, le origini e la descrizione dei dati: gli sviluppatori devono fornire dettagli sulle fonti dei dati, la descrizione del numero di punti dati utilizzati, i tipi di dati impiegati (come le etichette) e il periodo di raccolta dei dati. Queste informazioni sono essenziali per valutare se il sistema di IA possa essere influenzato da pregiudizi incorporati nei dati.

Altro punto fondamentale richiesto dalla legge riguarda gli aspetti legali: l'AB 2013 richiede anche la divulgazione se i dati includono materiali protetti da diritto d'autore, marchi o brevetti, se sono di dominio pubblico o concessi in licenza, e se includono informazioni personali. Questo requisito si allinea alle preoccupazioni più ampie di protezione della privacy e rispetto delle leggi sui dati personali, come il California Consumer Privacy Act (CCPA)[172].

Per bilanciare l'intrusività di queste prescrizioni sono state previste dei casi che esulano dagli obblighi previsti, di fatti la legge include alcune eccezioni all'obbligo di trasparenza.

Le esenzioni riguardano i sistemi di IA sviluppati esclusivamente per: la sicurezza e integrità, ovvero i sistemi di IA utilizzati per prevenire attacchi o garantire la sicurezza informatica, l'aviazione, quindi i sistemi di IA utilizzati per gestire operazioni di volo in spazi aerei e da ultimo per la difesa e sicurezza nazionale, quindi i sistemi di IA progettati per scopi militari o di sicurezza, protetti da norme federali.

Questa legge riflette la crescente preoccupazione che le IA generative siano utilizzate in modo irresponsabile, bilanciando l'innovazione tecnologica con la necessità di proteggere i diritti degli individui e la sicurezza delle informazioni

L'integrazione con leggi federali come l'Algorithmic Accountability Act[173] è evidente; di fatti, la legge federale, al pari di quella della California, richiede alle aziende di condurre

[172] Il California Consumer Privacy Act (CCPA) è una delle leggi più importanti negli Stati Uniti in materia di privacy dei dati. Il CCPA mira a fornire ai consumatori californiani un maggiore controllo sulle informazioni personali raccolte dalle aziende, con lobiettivo di aumentare la trasparenza e la responsabilità nelluso dei dati. I diritti fondamentali garantiti dal CCPA includono: il diritto di accesso, il diritto di cancellazione, il diritto di opt-out e il diritto alla non discriminazione. In caso di violazioni, il CCPA prevede sanzioni che possono arrivare fino a 7.500 dollari per ogni infrazione intenzionale. Inoltre, le disposizioni del California Privacy Rights Act (CPRA), che ha ampliato il CCPA, hanno ulteriormente rafforzato i diritti dei consumatori e stabilito una nuova agenzia di controllo, la California Privacy Protection Agency.

[173] S.3572 - Algorithmic Accountability Act of 2022, 117th congress, USA

valutazioni d'impatto sugli algoritmi che utilizzano IA, con particolare attenzione ai pregiudizi e ai bias algoritmici e agli effetti negativi che queste tecnologie potrebbero avere sui consumatori. Insieme, queste leggi creano un quadro completo che protegge i diritti dei consumatori e garantisce che i sistemi di IA generativa vengano sviluppati e utilizzati in modo trasparente e responsabile.

Passando, quindi, al Senate Bill 24-205 approvato in Colorado, pur non focalizzandosi direttamente sull'IA generativa amplia la regolamentazione includendo le decisioni consequenziali, ossia decisioni che possono avere un impatto significativo su ambiti vitali come istruzione, lavoro e finanza.

Il concetto di decisioni consequenziali è centrale nella SB 24-205. Queste sono decisioni che hanno effetti legali o sostanziali in aree chiave come l'istruzione, l'occupazione, i servizi finanziari, la salute, l'alloggio e i servizi legali. L'IA generativa potrebbe avere un impatto diretto in questi ambiti se utilizzata per automatizzare decisioni critiche, come la concessione di prestiti o l'assegnazione di posti di lavoro.

Approfondendo questo discorso l''IA generativa può realizzare contenuti che influenzano queste decisioni, come la produzione automatica di lettere di rifiuto per un impiego o la valutazione dell'idoneità di un individuo per ottenere un prestito. Questi utilizzi possono amplificare pregiudizi già presenti nei dati di addestramento o nelle regole programmate per il sistema. Risulta, quindi, evidente come un intelligenza artificiale che non permette la visione o non giustifica il proprio processo decisionale sia un enorme rischio per coloro che devono sostenere il peso di queste scelte, in particolare sono da rilevare le difficoltà date dall'utilizzo indiscriminato di AI dal dubbio addestramento, o che anche ammettendo un addestramento esaustivo possono essere soggette, come già citato nei paragrafi precedenti, a bias decisionali a causa di carenze ovvero dati errati..

Al fine di mitigare questi rischi la legge impone ai sviluppatori di IA ad alto rischio di implementare meccanismi di gestione del rischio per evitare discriminazioni algoritmiche e di fornire documentazione ai deployers (coloro che implementano i sistemi). Allo stesso tempo, i deployers devono adottare politiche per mitigare i rischi e informare i consumatori sull'utilizzo di sistemi di IA in decisioni consequenziali. Sostanzialmente la legge richiede che tali sistemi siano soggetti a processi di audit e valutazioni d'impatto, in modo simile alle disposizioni della AB 2013, per garantire che non siano discriminatori o opachi.

Questa normativa si allinea con i principi delineati nel "National AI Initiative Act", di cui parleremo di seguito, la quale promuove la ricerca e lo sviluppo dell'IA a livello federale, con particolare attenzione alla creazione di standard etici. La legge del Colorado completa questo approccio, garantendo che le tecnologie IA utilizzate in contesti critici siano soggette a rigorosi controlli etici e a valutazioni sui loro potenziali impatti social.

Un altro esempio di regolamentazione statale che ha un impatto sull'IA generativa è l' "Illinois Artificial Intelligence Video Interview Act". Questa legge richiede trasparenza nell'uso dell'IA durante le interviste video per l'assunzione. Quindi le aziende che utilizzano IA per analizzare interviste video dovranno informare i candidati sull'uso dell'IA e ottenere il loro consenso. Questo si allinea con il principio di trasparenza già delineato durante la discussione dell'AB 2013 della California.

La Local Law 144 di New York City, approvata nel 2021 e successivamente emendata fino al maggio 2023, entrata in vigore il 5 luglio 2023, richiede audit obbligatori per verificare che i sistemi di IA utilizzati per la selezione del personale non siano discriminatori. Questa legge introduce il concetto di responsabilità algoritmica, similmente a quanto previsto dalla Senate Bill 24-205 del Colorado per le decisioni consequenziali. In entrambi i casi, i legislatori stanno cercando di affrontare il rischio che i sistemi di IA generativa possano influenzare decisioni critiche senza una supervisione umana adeguata.

In un'ottica più ampia, a livello federale è stato approvato, nel 2020, il National AI Initiative Act (NAIIA, come parte del National Defense Authorization Act per l'anno fiscale 2021, con l'obiettivo di stabilire una strategia nazionale per la ricerca, lo sviluppo e l'adozione dell'intelligenza artificiale (IA) negli Stati Uniti. Questa legge rappresenta un passo cruciale per garantire che gli Stati Uniti mantengano la loro leadership nell'innovazione tecnologica dell'IA, obiettivo inserito come uno dei pilastri fondanti rispetto alla legge stessa, promuovendo un uso responsabile ed etico di tali tecnologie. Gli obiettivi principali riguardano lo stabilire una struttura per il coordinamento delle politiche sull'IA attraverso agenzie governative, il settore privato e il mondo accademico.

In estrema sintesi, il National AI Initiative Act rappresenta una base legale per supportare lo sviluppo e l'implementazione di sistemi di intelligenza artificiale negli Stati Uniti, con un focus su ricerca, etica, responsabilità stabilendo una piattaforma di *"governance nazionale dell'IA"*, porgendo un occhio a quella che deve rimanere uno dei punti focali

dell'implementazione di queste sistemi, ovvero la migliore integrazione con la forza lavoro anche attraverso alla formazione di personale qualificato e competente, oltre che strutturare collaborazioni intersettoriali considerando anche programmi federali e senza violare quelli che sono i diritti dei consumatori già consolidati e le norme etiche.

In conclusione ed estrema sintesi l'integrazione tra le leggi statali, come la AB 2013 e la SB 24-205, e le iniziative federali come il National AI Initiative Act e l'Algorithmic Accountability Act, porta alle prime fasi della creazione di un quadro giuridico robusto e multilivello per regolamentare l'intelligenza artificiale generativa negli Stati Uniti, non solo frammentario ma generando una sorta di base federale che dia un indirizzo unitario sul quale procedere con successive specificazioni e approfondimenti in base alla volontà giuridica del legislatore del singolo stato.

Rispetto ai casi analizzati, è facilmente comprensibile come sarebbe potuto essere frammentario lo sforzo legislativo se fosse venuto meno lo strumento federale. Ad esempio, la California, si concentra principalmente sulla trasparenza dei dati di addestramento e la protezione dei consumatori; il Colorado affronta principalmente i rischi legati alle decisioni consequenziali automatizzate, cercando di prevenire discriminazioni e garantire un uso responsabile in ambiti vitali; prospettata una così vasta disomogeneità le iniziative federali, come il NAIIAe l'Algorithmic Accountability Act, diventano fondamentali per supportano una visione più ampia per la governance etica dell'IA, fungendo da tessuto fondante della legislazione di materia e promuovendo standard condivisi e collaborazioni tra settore pubblico e privato.

Come già detto, però queste leggi rappresentano solo l'inizio del percorso verso un quadro regolatorio completo, che dovrà evolversi man mano che le tecnologie di IA generativa continuano a progredire.

.4 L'impegno delle grandi aziende di social network

I grandi social network hanno implementato una serie di politiche e tecnologie per contrastare il fenomeno dei Deepfake, dati i rischi che tali contenuti rappresentano per la disinformazione e la fiducia degli utenti. Di seguito una panoramica dettagliata delle strategie e delle collaborazioni introdotte dalle principali piattaforme.

Per quanto riguarda la piattaforma Reddit, "luogo" che ha dato i natali a questo fenomeno, ha adottato diverse misure per contrastare il fenomeno dei Deepfake sulla sua piattaforma. Nel febbraio 2018, ha bandito il "subreddit" (r/deepfakes), noto per la condivisione di video pornografici non consensuali generati tramite Deepfake, citando la violazione della politica contro la pornografia involontaria.[174]

Successivamente, nel gennaio 2020, Reddit ha aggiornato le sue politiche per vietare l'uso ingannevole dell'immagine di altri, includendo esplicitamente i Deepfake. Questa decisione è stata presa per prevenire la diffusione di contenuti manipolati che potrebbero fuorviare gli utenti, specialmente in contesti sensibili come le elezioni.[175]

Meta, che controlla Facebook, Instagram e Threads, è stata una delle prime aziende a reagire al fenomeno dei Deepfake. Nel 2020, ha aggiornato le sue politiche, vietando esplicitamente i contenuti multimediali manipolati che possano indurre in errore gli utenti su questioni di sicurezza pubblica, elezioni o argomenti di forte impatto sociale. Questi video alterati in modo avanzato, solitamente con tecniche di intelligenza artificiale, vengono monitorati e rimossi se risultano pericolosi per la percezione della realtà degli utenti.[176]

[174] L. Keolin "*Reddit bans deepfake porn videos*" (8 febbraio 2018) BBC <bbc.com>
[175] J. Peters "*Reddit bans impersonation on its platform / But impersonation that's satire will still be allowed*" (9 gennaio 2020) The Verge <www.theverge.com>
[176] M. Bicket "*Enforcing Against Manipulated Media*" (6 gennaio 2020) Meta <about.fb.com>

Meta collabora inoltre con organizzazioni di fact-checking esterne in tutto il mondo[177]. Questo approccio permette alle piattaforme di integrare l'analisi umana ai sistemi automatizzati di rilevamento, migliorando l'accuratezza nella segnalazione dei contenuti manipolati.[178]

Oltre a questo approccio, Meta ha lanciato un programma di fact-checking, La Deepfake Detection Challenge (DFDC). L'obiettivo principale della DFDC è stimolare la comunità di ricerca e sviluppo a creare strumenti efficaci per rilevare e contrastare i Deepfake.[179]

Quando un contenuto viene identificato come Deepfake o manipolato, viene etichettato come tale, riducendone la visibilità e avvisando l'utente che interagisce con esso.

Anche Twitter (ora X) ha adottato una posizione chiara contro i Deepfake, introducendo una politica specifica per i "media sintetici e manipolati". Questa politica, similmente a quanto previsto da Meta, prevede che i contenuti multimediali falsificati o distorti in modo significativo siano etichettati, in modo che gli utenti siano consapevoli della loro natura. In casi di contenuti particolarmente dannosi o ingannevoli, Twitter può, unilateralmente, decidere di rimuoverli completamente.[180]

Oltre a fornire contesto ai contenuti manipolati, Twitter incoraggia la trasparenza e ha adottato un sistema di avvisi per i tweet che potrebbero contenere contenuti non autentici, rendendo più difficile la diffusione virale di Deepfake ingannevoli.

YouTube, piattaforma appartenente al colosso Alphabet, al quale appartiene, tra le altre, anche Google ha introdotto un approccio simile, seppur più restrittivo, vietando contenuti che

---

[177] Meta collabora con organizzazioni di fact-checking indipendenti, certificate dalla rete indipendente International Fact-Checking Network (IFCN) o dallEuropean Fact -Checking Standards Network (EFCSN), per individuare, verificare e adottare provvedimenti riguardo ai contenuti che favoriscono la disinformazione. "*Informazioni sul fact-checking su Facebook, Instagram e Threads*" centro assistenza per le aziende di meta <www.facebook.com/business/help/ >

[178] Cfr. A. Martin "*Chinese law enforcement linked to largest covert influence operation ever discovered*" (29 agosto 2023) The record <therecord.media>

[179] C. Carton Ferrer, B. Dolhansky, B. Pflaum, J. Bitton, J. Pan & J Lu, "*Deepfake Detection Challenge Results: An open initiative to advance AI*" (12 giugno 2020), Meta <ai.meta.com>

Si veda anche M. Bickert "*Our Approach to Labeling AI-Generated Content and Manipulated Media*" (5 aprile 2024) Meta <about.fb.com/news>

[180] Twitter ha lavorato a lungo ad un aggiornamento della sua politica sui media manipolati (di fatti divenendo l'ultima azienda tra i grandi social network a implementare tali modifiche) e, nellottobre 2020, ha proposto dei cambiamenti invitando il pubblico a partecipare a un sondaggio. Ha ricevuto oltre 6.500 risposte e consultato esperti e organizzazioni civiche. Il 90% dei partecipanti ha richiesto l'etichettatura dei tweet alterati e un avviso prima della condivisione, ma vi è divisione sulleventuale ri mozione di questi contenuti: il 55% degli intervistati negli Stati Uniti è favorevole, mentre altri temono un impatto negativo sulla libertà di espressione e la censura. S. Ghaffary "*Twitter is finally fighting back against deepfakes and other deceptive media*" (4 febbraio 2020) Vox <Vox.com>

utilizzano tecniche di manipolazione per distorcere la realtà, soprattutto quelli che possono influenzare in modo scorretto il pubblico su temi di grande rilevanza, come elezioni e salute pubblica. I Deepfake che risultano particolarmente pericolosi per l'integrità sociale o che possono causare danni sono prontamente rimossi dalla piattaforma.

YouTube utilizza una combinazione di algoritmi di machine learning e revisione umana per scovare e gestire questi contenuti. Inoltre, per tutelarsi contro i Deepfake più sofisticati, YouTube collabora con partner esterni specializzati nel rilevamento di media falsificati.[181]

Anche TikTok, piattaforma particolarmente popolare tra i giovani, ha aggiornato le sue linee guida per affrontare il fenomeno dei Deepfake. La piattaforma ha vietato la pubblicazione di contenuti manipolati che possano causare danni o ingannare gli utenti su eventi reali. TikTok si avvale di tecnologie avanzate di rilevamento, insieme a collaborazioni con fact-checker, per monitorare e gestire efficacemente i contenuti manipolati.

Questa piattaforma adotta un approccio preventivo, impedendo la circolazione di Deepfake prima che possano raggiungere un pubblico ampio, un modello che si è rivelato efficace per limitare l'impatto di queste manipolazioni sulla sua vasta base di utenti.

Quando un contenuto è sospettato di essere un Deepfake o mostra evidenti segni di manipolazione, viene segnalato per una revisione umana. In molti casi, questi video vengono bloccati prima di essere completamente caricati o resi visibili al pubblico. Se il video non viene immediatamente rimosso, TikTok può applicare delle limitazioni alla sua visibilità: ad esempio, il video potrebbe essere reso visibile solo a una piccola parte del pubblico, limitandone la possibilità di diventare virale.[182]

Le piattaforme social non agiscono solo individualmente; spesso collaborano anche tra loro e con organizzazioni esterne. Un esempio è la Coalizione per l'integrità dei media, un'iniziativa che vede il coinvolgimento di aziende tecnologiche e organizzazioni per standardizzare le metodologie di rilevamento dei Deepfake. L'Unione Europea ha anche

[181] Dallinizio del 2024, più di 200 artisti, inclusi nomi noti come Billie Eilish e Katy Perry, hanno sottoscritto una lettera aperta in cui definiscono limitazione generata dallIA senza autorizzazione come un "attacco alla creatività umana". YouTube sta, anche, rafforzando le misure per impedire lo scraping non autorizzato della sua piattaforma, un metodo impiegato da varie aziende tecnologiche per addestrare i propri sistemi di intelligenza artificiale. A. Genco "*YouTube si prepara a contrastare i deepfake con nuovi strumenti di protezione per i creator*" (6 settembre 2024) HD Blog < www.hdblog.it >

[182] S. Mlot "*TikTok Bans Deepfakes, Expands Fact-Checking to Fight Election Misinformation*" (6 agosto 2020) PC MAG <pcmag.com>

proposto normative che spingono le piattaforme a distinguere chiaramente i contenuti sintetici o manipolati, specialmente in periodi elettorali, per evitare che influiscano sul processo democratico.[183]

In generale, la collaborazione tra le piattaforme e con enti governativi o organizzazioni di fact-checking rappresenta un importante passo avanti nella lotta ai Deepfake. Tuttavia, la continua evoluzione delle tecniche di manipolazione richiede aggiornamenti costanti sia dal punto di vista tecnologico sia normativo.

[183] "*La Commissione pubblica orientamenti nellambito della legge sui servizi digitali per lattenuazione dei rischi sistemici online per le elezioni*" (26 marzo 2024) Rappresentanza in Italia <italy.representation.ec.europa.eu>

8 l’applicazione specifica della regolamentazione IA

.1 ingerenze nelle questioni politiche

Come si è potuto osservare durante questa trattazione l’intelligenza artificiale ha le capacità di influire in modo significativo sulla politica. Ciò ha fatto sì che venisse impiegata in diversi modi che possono tanto supportare quanto interferire con i processi politici.

Gli strumenti di IA vengono utilizzati per creare e diffondere propaganda e disinformazione su larga scala. Come si è visto durante la trattazione delle vicende americane e ucraine (§ 5) l’utilizzo di intelligenze artificiali e bot sui maggiori canali informativi e su social media possono amplificare messaggi politici mirati, mentre le tecnologie Deepfake possono generare video o audio falsi che appaiono estremamente realistici veicolando messaggi che possono essere molto distanti dalla volontà del soggetto rappresentato ovvero, come nel caso di Nancy Pelosi, facendo percepire un soggetto in stati di confusione, minando la solidità di quanto questa vuole rappresentare. Allo stesso tempo, governi e grandi aziende di tutto il mondo adottano l’IA per “sorvegliare” le persone, giustificando tale uso con la necessità di migliorare la sicurezza, sollevando preoccupazioni per la privacy e il rispetto dei diritti umani, specialmente in regimi autoritari.

Queste problematiche, però, non devono lasciare l’idea per la quale l’utilizzo di queste tecnologie può essere solo in danno; di fatti, l’IA può ottimizzare le campagne elettorali attraverso l’analisi dei dati per identificare gli elettori indecisi o per personalizzare gli annunci politici, sollevando questioni di trasparenza e equità. (§ 5.2)

Come si è visto al § 7 alcuni paesi stanno introducendo leggi per regolamentare l’uso dell’IA, influenzando chi ha accesso e controllo su queste tecnologie e come possono essere impiegate. Infine, attraverso l’uso di algoritmi che personalizzano le informazioni visualizzate online, l’IA può influenzare sottilmente le opinioni politiche, indirizzando gli utenti verso particolari punti di vista o ideologie (§ 5.2 & § 5.5). Questo panorama in continua evoluzione presenta sfide etiche e legali che richiedono un’attenzione costante.

L’ordinamento legale italiano ed europeo ha implementato diverse misure per contrastare le potenziali ingerenze e abusi legati all’uso dell’intelligenza artificiale (IA) nella politica. Come si è già detto a livello europeo, il Regolamento generale sulla protezione dei dati

(GDPR), introdotto nel 2018, è uno dei pilastri fondamentali nella protezione dei dati personali. Questa normativa rafforza il controllo degli individui sui loro dati personali, limitando il modo in cui questi possono essere raccolti e utilizzati, anche in contesti politici e di campagne elettorali.

Inoltre, nel 2021, la Commissione Europea ha proposto un nuovo regolamento sull'IA che mira a stabilire norme per lo sviluppo e l'uso sicuro dell'IA, categorizzando i sistemi di IA in base al rischio. Questo regolamento cerca di prevenire l'uso di sistemi di IA ad alto rischio che potrebbero manipolare il comportamento umano, minacciando i diritti civili. Parallelamente, l'UE ha introdotto un codice di buona pratica volontario per le piattaforme online, che incoraggia la trasparenza negli annunci politici e la chiarezza sull'uso di bot e AI nella diffusione di informazioni, mirando a ridurre la disinformazione.

In particolare, relativamente al caso di specie in analisi:

L'articolo 5 comma 1 lettera a) vieta l'immissione sul mercato, la messa in servizio o l'uso di un sistema di intelligenza artificiale che utilizzi tecniche subliminali al di fuori della coscienza di una persona o tecniche manipolative o ingannevoli, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di una persona in modo da causare o poter causare a tale persona o ad un'altra persona un danno fisico o psicologico o una perdita significativa di controllo sul proprio comportamento.[184]

Mentre *l'allegato III*, che elenca i sistemi di intelligenza artificiale ad alto rischio, include sistemi di intelligenza artificiale destinati ad essere utilizzati per influenzare l'esito di elezioni o referendum o il comportamento di voto delle persone naturali nell'esercizio del loro voto in elezioni o referendum. Sono esclusi i sistemi di intelligenza artificiale il cui output non è direttamente esposto alle persone fisiche, come gli strumenti utilizzati per organizzare, ottimizzare o strutturare le campagne politiche dal punto di vista amministrativo o logistiche.[185]

[184] Ivi nota 131
[185] ibidem

A livello nazionale, l'Italia ha recepito le norme del GDPR attraverso il Garante per la protezione dei dati personali ciò nel D. lgs 196/2003[186], che fornisce direttive specifiche sull'uso dei dati durante le campagne elettorali.

L'Italia promuove anche iniziative di sensibilizzazione e formazione per educare i cittadini e le istituzioni sui rischi associati all'uso dell'IA in contesti politici. Inoltre, vi è una stretta collaborazione con enti europei e internazionali per allinearsi agli standard globali nella regolamentazione dell'IA, assicurando che le misure adottate siano efficaci e congruenti con quelle dei partner internazionali. (§ 10)

Queste iniziative riflettono un impegno significativo, sia a livello italiano che europeo, per prevenire l'abuso dell'IA in ambito politico e mantenere un equilibrio tra innovazione tecnologica e protezione dei diritti civili.

.2 la legislazione del neuromarketing

Come abbiamo meglio discusso nel § 6 (.3 & .4) il neuromarketing è un campo che utilizza tecniche di neuroscienza per studiare le reazioni del cervello alle campagne di marketing, con l'obiettivo di capire meglio cosa spinge le decisioni di acquisto dei consumatori. A causa della sua natura intrusiva, il neuromarketing ha suscitato preoccupazioni in termini di privacy e etica, portando a discussioni su una possibile regolamentazione.

In Italia, al momento non esistono leggi specifiche che regolino il neuromarketing. Tuttavia, le attività di neuromarketing devono comunque conformarsi al Regolamento Generale sulla Protezione dei Dati (GDPR), il quale impone il consenso informato e la trasparenza nell'uso dei dati personali.

[186] Il d. lgs 196/2003 ha modificato il precedente Codice della privacy (istituito con legge 31 dicembre 1996, n. 675), introducendo quelle che sono le disposizioni in materia di protezione dei dati contenute nel GDPR e serve come pilastro della disciplina, che è in continua evoluzione e influenzata dallambito europeo.

La normativa italiana sulla privacy e la protezione dei dati personali segue dunque i dettami europei, richiedendo che le aziende divulghino le loro metodologie e ottenere il consenso esplicito prima di raccogliere dati biometrici o neurologici.

A livello europeo, il GDPR è la principale normativa che incide sul neuromarketing. Il testo del regolamento non menziona esplicitamente il termine "neuromarketing". Tuttavia, discutono come l'IA possa essere utilizzata per profilare gli individui e influenzare il loro comportamento attraverso la manipolazione, concetti che possono essere correlati al neuromarketing.

La ricerca pubblicata dall'EPRS[187] ha messo in luce alci campi di particolare importanza e rilevanza nell'applicazione della regolamentazione del GDPR[188]:

- Sezione 2.3.4: Descrive come l'IA può essere usata per prevedere la propensione di una persona a reagire a determinati stimoli. Questo è un concetto chiave nel neuromarketing, che mira a capire come il cervello risponde a stimoli di marketing.
- Sezione 2.3.5: Il caso di Cambridge Analytica illustra come i profili psicologici basati sui dati degli utenti di Facebook siano stati utilizzati per inviare messaggi politici mirati, influenzando potenzialmente il comportamento degli elettori. Questo è un esempio di come tecniche di profilazione avanzata, che potrebbero includere anche aspetti di neuromarketing, possono essere impiegate per influenzare le decisioni.
- Sezione 2.3.7: Si parla dell'uso di messaggi mirati nel dominio politico, sollevando preoccupazioni sul fatto che la personalizzazione, che potrebbe attingere a principi del neuromarketing, possa minare la formazione di opinioni politiche ponderate.

Sebbene lo studio non esplori il neuromarketing in modo approfondito, questi esempi suggeriscono come l'IA, con la sua capacità di profilare e influenzare il comportamento, possa essere utilizzata in modi che si sovrappongono a questo campo. La mancanza di trasparenza e

[187] European Parliamentary Research Service < www.europarl.europa.eu >
[188] G. Sartor, F. Lagioia *"The impact of the General Data Protection Regulation (GDPR) on artificial intelligence"* (2020). European Parliament, Brussels.

la potenziale manipolazione sollevano questioni etiche importanti, evidenziando la necessità di un'attenta regolamentazione dell'uso dell'IA in questi contesti.

Oltre a questo primo studio, il Parlamento Europeo e altre istituzioni dell'UE hanno iniziato a mostrare interesse per il potenziale impatto delle neurotecnologie e dell'intelligenza artificiale sulla società. Ad esempio, il Gruppo Europeo di Etica nelle Scienze e nelle Nuove Tecnologie ha pubblicato diversi pareri che toccano temi correlati, evidenziando la necessità di garantire che tali tecnologie siano usate in maniera etica e responsabile.

Nel 2018 è stato avviato un ulteriore progetto di ricerca giustappunto a riguardo dell'etica dell'utilizzazione dell'intelligenza artificiale; da questo è derivata la pubblicazione riguardo gli orientamenti etici per un'IA affidabile[189]**.**

In questo documento all'articolo 9, paragrafo 1, dell'Allegato III si affronta l'uso dei sistemi di intelligenza artificiale (IA) per la "manipolazione subliminale", che può essere correlata al campo del neuromarketing.

L'articolo 9, paragrafo 1, dell'Allegato III afferma: "i sistemi di IA destinati a distorcere il comportamento di una persona in modo da causare o poter causare a tale persona o ad un'altra persona un danno fisico o psicologico o una perdita significativa di controllo sul proprio comportamento" sono considerati sistemi di IA ad alto rischio.

Sebbene non sia esplicitamente menzionato il neuromarketing, la manipolazione subliminale potrebbe essere considerata una tecnica utilizzata in questo campo. Il neuromarketing si basa sulla comprensione delle risposte del cervello agli stimoli di marketing per influenzare il comportamento dei consumatori. L'uso di tecniche subliminali, che operano al di sotto del livello di coscienza, potrebbe essere visto come un modo per raggiungere questo obiettivo.

A proposito delle pubblicità ingannevoli o subliminali abbiamo altre fonti. Infatti, questo campo è protetto dal Decreto Legislativo 2 agosto 2007, n. 145[190] il quale attua l'articolo 14 della direttiva 2005/29/CE che modifica la direttiva 84/450/CEE sulla pubblicita' ingannevole. Il decreto, all'art 5 vieta "ogni forma di pubblicità subliminale".

[189] N. Smuha et alt. "*orientamenti etici per unIA affidabile* " (8 aprile 2019) Commissione europea
[190] Decreto Legislativo 2 agosto 2007, n. 145

Tornando al discorso precedente, sebbene il termine "neuromarketing" non compaia nelle fonti, l'articolo 9, paragrafo 1, dell'Allegato III affronta una tecnica (manipolazione subliminale) che potrebbe essere utilizzata in questo campo, evidenziando le potenziali preoccupazioni etiche e la necessità di una regolamentazione specifica.

Si prevede che nel futuro possano essere sviluppate normative più specifiche per il neuromarketing, sia a livello nazionale che europeo, soprattutto all'aumentare dell'uso delle tecnologie di rilevazione e analisi dei dati neurali. L'obiettivo dovrebbe essere quello di bilanciare le opportunità offerte dal neuromarketing con la necessità di proteggere i diritti fondamentali dei consumatori.

In sintesi, il quadro legislativo attuale incentrato sul GDPR e poche altre fonti normative europee; eppure fornisce già una certa protezione. Nonostante ciò, l'evoluzione continua delle tecnologie di marketing e neuroscienza potrebbe necessitare aggiustamenti normativi mirati.

In questo campo, seppur si continui a non far riferimento diretto, tornano ad essere di particolare rilievo le direttive complementari di cui abbiamo parlato ne capitolo precedente (§ .1); di fatti il "Digital Markets Act" e il "Digital Services Act" sono due regolamenti dell'Unione Europea che affrontano le sfide del mercato digitale con obiettivi complementari.

Come già discusso nel sotto capitolo precedente (§ .1), il "Digital Markets Act" si concentra sul contrastare il potere eccessivo delle grandi piattaforme online, chiamate "gatekeeper", per garantire un mercato digitale equo e contendibile. Questo regolamento si preoccupa principalmente di pratiche commerciali che potrebbero soffocare la concorrenza, come l'autopreferenza, l'utilizzo dei dati in modo sleale e la creazione di barriere all'uscita per gli utenti. Introduce quindi obblighi specifici per i gatekeeper, come la necessità di garantire l'interoperabilità con i servizi di terzi[191], la trasparenza nelle condizioni di accesso e l'equità nel trattamento dei propri prodotti e servizi rispetto a quelli offerti da altri utenti commerciali[192]. La Commissione Europea ha il potere di intervenire tempestivamente per imporre questi obblighi, anche in situazioni di emergenza, al fine di prevenire danni gravi e irreparabili alla concorrenza[193].

[191] Regolamento (UE) 2022/1925 considerando 64
[192] ivi considerando 52
[193] Ivi considerando 27, 83 e 84

D'altro canto, il "Digital Services Act", sempre come rapidamente discusso nel capitolo precedente (§ 5.1), mira a creare un ambiente online più sicuro e trasparente, affrontando la diffusione di contenuti illegali e dannosi e proteggendo i diritti fondamentali degli utenti. Questo regolamento si applica a un'ampia gamma di prestatori di servizi intermediari (a differenza del precedente che è pensato solo per le grandi realtà), dalle piattaforme di social media ai siti di e-commerce, e stabilisce una serie di obblighi di diligenza per garantire un ambiente online più responsabile. Tra questi obblighi, troviamo la necessità di implementare meccanismi di segnalazione e rimozione dei contenuti illegali[194], di fornire informazioni chiare e accessibili agli utenti sulle loro politiche di moderazione dei contenuti[195] e di istituire sistemi interni di gestione dei reclami equi ed efficienti. [196]

Come discusso nelle nostre conversazioni precedenti, entrambi i regolamenti riconoscono l'importante ruolo dell'intelligenza artificiale (IA) nel plasmare l'ambiente digitale. L'IA può essere utilizzata sia per scopi positivi, come la personalizzazione dei servizi e la lotta ai contenuti illegali, sia per scopi negativi, come la manipolazione del comportamento e la diffusione di disinformazione.

Ad esempio, l'Articolo 3 del "Digital Services Act" definisce i Deepfake come contenuti generati o manipolati dall'IA che possono apparire falsamente autentici[197], e gli articoli 15, 27 39, 42 e 49, introducono obblighi di trasparenza per i sistemi di IA, richiedendo agli utenti di essere informati quando interagiscono con un sistema di IA e di identificare chiaramente i Deepfake.

Sebbene il termine "neuromarketing" non sia esplicitamente menzionato nei documenti, sono diversi gli articoli evidenziano la preoccupazione dell'Unione Europea per l'uso dell'IA per influenzare il comportamento degli utenti, un aspetto centrale del neuromarketing. Ad esempio, l'Articolo 34 del "Digital Services Act", classifica come sistemi di IA ad alto rischio quelli che possono manipolare il comportamento delle persone causando danni fisici o psicologici.

---

[194] Regolamento 2022/2065 considerando 22 e 52

[195] Ivi considerando 49

[196] Ivi considerando 58

[197] Testualmente: "*"deep fake" indica un contenuto di immagini, audio o video generato o manipolato dallIA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona.*" Art 3 punto 60 Regolamento (UE) 2024/1689

In sintesi, il “Digital Markets Act” e il “Digital Services Act” rappresentano un passo importante verso la regolamentazione del mercato digitale e la creazione di un ambiente online più equo, trasparente e sicuro. L’obiettivo è promuovere l’innovazione e la crescita economica, garantendo al contempo la protezione dei diritti fondamentali degli utenti, compresa la protezione dalla manipolazione e l’accesso a un’informazione libera e pluralistica.

.3 Gli effetti sul Copyright

La violazione del copyright mediante lo sfruttamento dell’intelligenza artificiale è oggi un tema di grande rilevanza, poiché coinvolge una serie di profili sia tecnologici che giuridici. L’uso sempre più diffuso di sistemi di intelligenza artificiale, in particolare di quelli cosiddetti “generativi” (tra cui i modelli linguistici di grandi dimensioni e le reti di deep learning), solleva interrogativi delicati riguardo ai diritti di proprietà intellettuale sulle opere protette da diritto d’autore. Il fulcro delle questioni è duplice: da un lato, capire in quali casi e con quali limiti sia lecito utilizzare opere coperte da copyright ai fini dell’addestramento di algoritmi; dall’altro, stabilire se e come proteggere e regolare i contenuti prodotti autonomamente dall’IA, nonché individuare chi risponda in caso di violazioni.

Il procedimento di addestramento di un modello di intelligenza artificiale, come si è discusso durante la trattazione dell’addestramento delle intelligenze Deepfake (§4.2), in genere richiede una quantità notevole di dati: testi, immagini, suoni e altri contenuti che possono essere protetti dalla normativa sul diritto d’autore.

In molti casi, questi dati vengono estratti da fonti pubblicamente accessibili, come i siti internet, i database online o i social network. L’elemento controverso consiste nel fatto che, per apprendere le regole linguistiche, musicali o visive, un sistema di intelligenza artificiale può analizzare miliardi di frammenti di opere che, di per sé, sono tutelate da diritto d’autore, e che – al di fuori di specifiche eccezioni – non dovrebbero essere riprodotte né elaborate senza il consenso dell’autore o del titolare dei diritti.

Se negli Stati Uniti l'uso di tali contenuti per finalità di apprendimento e analisi dati può talvolta beneficiare della dottrina del "fair use", che riconosce alcune forme di utilizzo non autorizzato purché ricorrano determinate condizioni (tra cui l'uso trasformativo e scopi di ricerca o critica), negli ordinamenti giuridici europei questa dottrina non ha un'esatta corrispondenza. L'Unione Europea, tuttavia, ha cercato di fornire strumenti normativi per allineare i vari Paesi a un sistema comune di eccezioni e limitazioni, in particolare con la direttiva sul diritto d'autore e sui diritti connessi nel mercato unico digitale (cosiddetta Direttiva Copyright). [198]

L'appena citata direttiva (UE) 2019/790, del 17 aprile 2019, mira ad un'armonizzazione del diritto d'autore e dei diritti connessi nel mercato unico digitale, con particolare attenzione agli utilizzi digitali e transfrontalieri dei contenuti protetti. La direttiva introduce norme su eccezioni e limitazioni al diritto d'autore, sull'agevolazione delle licenze e sul funzionamento del mercato per lo sfruttamento delle opere.

Per quanto riguarda le eccezioni e le limitazioni al diritto d'autore, la direttiva introduce una eccezione obbligatoria per università, organismi di ricerca e istituti di tutela del patrimonio culturale per l'estrazione di testo e dati a fini di ricerca scientifica, consentendo la conservazione delle copie realizzate per tali scopi.[199] I titolari dei diritti possono applicare misure per la sicurezza dei loro sistemi, ma viene anche introdotta un'eccezione o limitazione per l'estrazione di testo e dati per altri scopi, a condizione che i titolari dei diritti non si siano espressamente riservati tali diritti.[200] Inoltre, è prevista un'eccezione o limitazione per l'utilizzo di opere a scopo illustrativo per l'insegnamento digitale, inclusi contesti online e transfrontalieri, sotto la responsabilità degli istituti di istruzione, con la possibilità per gli Stati membri di escludere tale eccezione se sono disponibili licenze adeguate.[201] L'utilizzo per l'insegnamento online è considerato avvenire nello Stato membro in cui ha sede l'istituto, e gli Stati membri possono prevedere un equo compenso per i titolari dei diritti. Infine, è introdotta un'eccezione per gli istituti di tutela del patrimonio culturale per la riproduzione di opere presenti nelle loro raccolte ai fini della conservazione. [202]

[198] Direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale (c.d. Direttiva Copyright). Modificativa delle precedenti direttive: 96/9/CE e 2001/29/CE

[199] Ivi art 3

[200] Ivi art 4

[201] Ivi art 5

[202] Ivi art 6

Per migliorare le procedure di concessione delle licenze e garantire un più ampio accesso ai contenuti, la direttiva prevede che gli organismi di gestione collettiva possano concedere licenze non esclusive per opere fuori commercio agli istituti di tutela del patrimonio culturale per scopi non commerciali, con la possibilità per gli istituti di rendere disponibili online tali opere, a condizione che non esistano organismi di gestione collettiva. I titolari dei diritti possono escludere le proprie opere da tali meccanismi. Viene definito il concetto di opera fuori commercio e stabilito che sia necessario uno sforzo ragionevole per valutarne la disponibilità. Le licenze per opere fuori commercio possono consentirne l'utilizzo in tutti gli Stati membri, e l'utilizzo di opere in base all'eccezione per opere fuori commercio è considerato avvenire nello Stato membro in cui ha sede l'istituto. [203] Per le opere audiovisive su piattaforme di video on-demand, gli Stati membri devono fornire assistenza per la negoziazione di licenze attraverso organismi imparziali o mediatori. [204]

Per garantire il buon funzionamento del mercato per il diritto d'autore, la direttiva riconosce agli editori di giornali un diritto connesso per l'utilizzo online delle loro pubblicazioni da parte di prestatori di servizi della società dell'informazione, escludendo gli utilizzi privati e non commerciali, i collegamenti ipertestuali e le singole parole o estratti molto brevi.[205] Gli autori di opere incluse in pubblicazioni giornalistiche devono ricevere una quota adeguata dei proventi. Gli Stati membri possono stabilire che gli editori abbiano diritto a una quota del compenso per gli utilizzi di opere in virtù di eccezioni o limitazioni.[206] I prestatori di servizi di condivisione di contenuti online sono responsabili per la comunicazione al pubblico di opere protette caricate dagli utenti e devono ottenere l'autorizzazione dai titolari dei diritti.[207] Tali prestatori sono responsabili se non dimostrano di aver compiuto i massimi sforzi per ottenere l'autorizzazione e per impedire la disponibilità di contenuti non autorizzati, e sono previsti meccanismi di reclamo e ricorso per gli utenti. L'applicazione di tali misure non deve pregiudicare le eccezioni e le limitazioni al diritto d'autore e i prestatori di servizi non hanno un obbligo generale di sorveglianza. Autori e artisti hanno diritto a una remunerazione adeguata e proporzionata e devono ricevere informazioni trasparenti sullo sfruttamento delle loro opere, con la previsione di un meccanismo di adeguamento contrattuale se la remunerazione iniziale si rivela sproporzionatamente bassa, e una procedura di risoluzione alternativa delle

[203] Ivi art 8
[204] Ivi art 13
[205] Ivi art 15
[206] Ivi art 16
[207] Ivi art 17

controversie.[208] Inoltre, è previsto un diritto di revoca in caso di mancato sfruttamento dell'opera o esecuzione.[209]

In sintesi, questa direttiva cerca di modernizzare il diritto d'autore nell'era digitale, promuovendo un equilibrio tra i diritti dei titolari e le esigenze della società dell'informazione e della ricerca.

Il recepimento di queste norme nei diversi Paesi membri, fra cui l'Italia, non ha eliminato dubbi e incertezze, specialmente in merito all'uso dei contenuti a scopo di addestramento commerciale; di fatti, seppur lo scopo fosse proteggere i diritti degli autori di opere originali, la direttiva resta completamente estranea la pressante tematica dell'intelligenza artificiale, argomento che venne trattato solo in seguito.

In Italia, il D.Lgs. 177/2021[210] ha modificato la Legge sul diritto d'autore (n. 633/1941 [211]) per adeguarla alla Direttiva Copyright, ma non esiste ancora un quadro del tutto chiaro circa la liceità dell'impiego di materiali protetti per la formazione di algoritmi di intelligenza artificiale che mirino a fornire servizi, prodotti o strumenti commerciali. In tale scenario si sono aperte trattative tra società titolari di grandi cataloghi di contenuti e sviluppatori di IA, che chiedono licenze o corrispettivi economici per consentire l'uso di canzoni, libri o immagini.

Un secondo fronte problematico riguarda la natura dei contenuti generati dall'IA. Se un algoritmo, dopo l'addestramento, è in grado di creare musiche, testi, disegni o addirittura veri e propri romanzi, ci si domanda a chi spettino i diritti su queste opere.

La tradizione giuridica italiana ed europea assegna la tutela del diritto d'autore a opere che siano frutto di creatività umana e originalità.

L'articolo 1 della Legge n. 633/1941 parla di "opere dell'ingegno di carattere creativo" e la giurisprudenza ha ribadito che occorre un apporto di tipo personale, che dia all'opera un tratto distintivo, frutto di un processo cognitivo e di scelte consapevoli.[212] Se una creazione è il risultato esclusivo di un calcolo algoritmico, privo di qualunque intervento o controllo umano,

---

[208] Ivi art 18

[209] Ivi art 22

[210] Decreto Legislativo 8 novembre 2021, n. 177, recante "*Attuazione della direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale*".

[211] Legge 22 aprile 1941, n. 633 (Legge sul diritto d'autore).

[212] *Infra multis* : Cass. civ., Sez. I, Ordinanza, 19/01/2023, n. 1674

non rientra nella definizione tradizionale di opera protetta, almeno secondo la maggior parte delle interpretazioni dottrinali e giurisprudenziali finora emerse. Da qui sorge una zona grigia che comprende, ad esempio, i casi in cui l'utente fornisce istruzioni all'algoritmo, seleziona le impostazioni, o compie operazioni di rifinitura e correzione. In tali casi, potrebbe sostenersi che esista un apporto umano sufficiente da conferire originalità e che dunque l'opera sia proteggibile dal diritto d'autore, con conseguente riconoscimento dell'utente-sviluppatore come autore o coautore. La valutazione è fortemente casistica e, poiché le tecnologie avanzano rapidamente, ogni caso concreto potrebbe presentare peculiarità difficili da ricondurre agli schemi consolidati.

Altra questione fondamentale è quella della responsabilità per i contenuti generati dall'IA, soprattutto se questi contenuti riproducono, in modo parziale o totale, opere protette senza autorizzazione, configurando così un plagio o una violazione del copyright. La tradizionale impostazione giuridica, sia in Italia sia in Europa, tende ad attribuire responsabilità all'autore o al soggetto che materialmente distribuisce l'opera illecita, ma di fronte a un sistema capace di elaborare dati e creare output quasi in autonomia, la catena di responsabilità si fa più articolata. Chi ha programmato o addestrato il modello, fornendo determinati dataset, potrebbe avere un ruolo, se non ha introdotto filtri o meccanismi di prevenzione della generazione di contenuti illeciti. Anche l'utente che inserisce il prompt o l'istruzione all'IA potrebbe essere corresponsabile, se la finalità perseguita è chiaramente quella di ottenere una contraffazione o un plagio. Infine, le piattaforme che ospitano o condividono i contenuti creati da algoritmi di IA sono tenute a rispettare gli obblighi di sorveglianza, rimozione e blocco di contenuti illeciti secondo quanto stabilito, a livello europeo, dal Digital Services Act[213] e, in Italia, dal D.Lgs. 70/2003 di recepimento della Direttiva sul commercio elettronico[214]. Tuttavia, non è ancora del tutto chiaro se tali obblighi siano destinati a rafforzarsi, in futuro, alla luce della crescente complessità dei modelli di IA e del fatto che questi possono generare infinite varianti di contenuti potenzialmente protetti.

[213] Ivi nota 168

[214] Decreto Legislativo 9 aprile 2003, n. 70 "*Attuazione della direttiva 2000/31/CE relativa a taluni aspetti giuridici dei servizi della società dellinformazione nel mercato interno, con particolare riferimento al commercio elettronico*". Da ultimo modificato dal Decreto Legislativo 25 marzo 2024, n. 50, "*Disposizioni integrative e correttive del decreto legislativo 8 novembre 2021, n. 208*", recante il testo unico dei servizi di media audiovisivi in considerazione dellevoluzione delle realtà del mercato, in attuazione della direttiva (UE) 2018/1808 di modifica della direttiva 2010/13/UE. (24G00067). GU n.90 del 17-04-2024

Il legislatore dell'Unione Europea sull'argomento del copyright attraverso le intelligenze artificiali si è espresso attraverso il Regolamento generale sull'Intelligenza Artificiale (Artificial Intelligence Act).[215]

Questo regolamento, come si è già ripetuto più volte, mira a porre regole di trasparenza, tracciabilità e responsabilità per i sistemi di IA a seconda del loro livello di rischio. L'intenzione è di assicurare che tali tecnologie rispettino i valori fondamentali e i diritti dei cittadini europei. Sebbene questo regolamento non riguardi in modo diretto e specifico la tutela del diritto d'autore, appare evidente come i suoi principi di trasparenza, accountability e tracciabilità potrebbero influire anche sulle modalità di addestramento e di utilizzo dei modelli IA, fornendo un quadro giuridico più chiaro per stabilire quando vi sia un uso scorretto di opere protette.[216]

Allo stesso tempo, le istituzioni europee e l'Organizzazione Mondiale della Proprietà Intellettuale (OMPI) sono consapevoli della necessità di aggiornare, o quantomeno di reinterpretare, le norme esistenti in materia di diritto d'autore per far fronte ai cambiamenti in atto. L'OMPI, in particolare, ha avviato consultazioni e pubblicato rapporti volti a individuare possibili strade che i legislatori nazionali e regionali potrebbero intraprendere, ad esempio stabilendo nuovi tipi di licenze che consentano l'uso di opere protette per l'addestramento a fronte di equi compensi, oppure chiarendo i criteri per la tutela (o il mancato riconoscimento di tutela) dei contenuti prodotti dall'IA.

Nel panorama italiano, il dibattito è vivo e tocca vari livelli. Da un lato, le società di gestione collettiva dei diritti, come la SIAE, si stanno attrezzando per comprendere meglio come negoziare compensi o licenze con i fornitori di sistemi di IA che, grazie a un uso massiccio di opere coperte, migliorano i propri servizi. Dall'altro, autori e artisti sollevano dubbi sul fatto che una parte consistente dei guadagni possa derivare da elaborazioni di dati e opere protette senza un'esplicita autorizzazione. L'incertezza giuridica spinge alcune piattaforme e sviluppatori di IA a regolarsi da soli, adottando policy interne che talvolta vietano la generazione di contenuti che riproducano materiali protetti e introducendo strumenti di riconoscimento e filtraggio. Non mancano, tuttavia, coloro che utilizzano la tecnologia per realizzare contenuti contraffatti o potenzialmente illeciti, sfruttando l'assenza di controlli stringenti o di una normativa adeguata alla realtà attuale e in continuo mutamento.

[215] Ivi nota 169
[216] Ivi, artt. 16, 17, 20, 21, 24, 53 c 1 l c), 53 c 1 l d) e 56 c 2 l b

La situazione europea e, in particolare, quella italiana, mostra dunque come si stia cercando di trovare un bilanciamento tra l'esigenza di proteggere i diritti degli autori, garantendo loro un giusto compenso e una difesa contro lo sfruttamento incontrollato delle proprie opere, e la necessità di non ostacolare la ricerca e l'innovazione.

La Direttiva Copyright, con le sue previsioni sul text and data mining, rappresenta un tentativo di conciliare tali interessi, ma la sua efficacia pratica dipenderà molto dall'interpretazione e dall'attuazione che gli Stati membri sceglieranno di darle. Al contempo, le disposizioni introdotte dal Digital Services Act e i futuri aggiornamenti normativi previsti nell'AI Act potrebbero riflettersi sulle modalità di generazione e di diffusione dei contenuti da parte dei sistemi di IA, introducendo obblighi di trasparenza, di responsabilità e di controllo più rigorosi.

Resta il fatto che la produzione di opere da parte dell'IA, soprattutto quando non è chiaro l'apporto umano, mette in crisi alcune certezze radicate nel diritto d'autore e impone di ridefinire concetti come "originalità", "creazione intellettuale" e "autore". La giurisprudenza italiana ed europea, nei prossimi anni, dovrà fornire risposte a casi concreti che, inevitabilmente, emergeranno con sempre maggiore frequenza. Sarà compito dei giudici stabilire se un output generato da un sistema di IA debba essere considerato copiato da opere preesistenti o rappresenti una rielaborazione sufficientemente originale, e dovranno anche individuare i soggetti responsabili delle eventuali violazioni. È ipotizzabile che si svilupperà una responsabilità a "cerchi concentrici", in cui ciascun attore (dall'utente che immette il prompt, allo sviluppatore dell'algoritmo, alle piattaforme di hosting) potrà essere chiamato a rispondere in funzione del proprio contributo alla creazione e alla diffusione dell'opera illecita.

In conclusione, l'espansione dell'intelligenza artificiale nel campo della produzione di contenuti sta offrendo opportunità enormi e, nello stesso tempo, aprendo scenari inediti di conflitto con la normativa sul diritto d'autore. Il tentativo dei vari ordinamenti giuridici di "mettere pezza" alle nuove problematiche è in pieno svolgimento, ma le soluzioni non sono univoche e richiedono un'opera di coordinamento e di aggiornamento costante.

Il recepimento della Direttiva Copyright, l'elaborazione del Digital Services Act, e dell'AI Act rappresentano i tasselli di un mosaico normativo in continua evoluzione. L'auspicio è che, attraverso questo processo, si riesca a tracciare un quadro chiaro che tuteli i diritti degli autori, eviti abusi sistematici e, allo stesso tempo, non soffochi la creatività e l'innovazione rese

possibili dall'IA. La posta in gioco non è solo la salvaguardia delle opere tradizionali, ma anche la creazione di un ambiente digitale trasparente, responsabile e aperto alla sperimentazione, capace di garantire un equo equilibrio tra gli interessi dei titolari di diritti, la libertà di ricerca e il progresso tecnologico.

9 La pericolosità e i rischi dell'immagine artificiale

.1 Violazione del diritto d'immagine e il deep-porn

Negli ultimi anni, l'utilizzo di strumenti di intelligenza artificiale capaci di creare o alterare immagini in modo estremamente realistico, come nel caso dei cosiddetti "Deepfake", ha spinto governi e autorità nazionali e sovranazionali a interrogarsi su come tutelare i diritti fondamentali della persona, fra cui il diritto all'immagine, alla reputazione e alla privacy.

Il quadro normativo, in continuo aggiornamento, si è evoluto non solo attraverso nuove iniziative legislative volte a regolamentare queste tecnologie, ma anche grazie all'applicazione estensiva di norme già esistenti, adattate al contesto dell'IA.

A livello internazionale ed europeo, si assiste all'impiego di regolamentazioni come il GDPR, che, pur non facendo riferimento esplicito all'intelligenza artificiale, tutela i dati personali, fra cui rientra anche l'immagine di una persona, considerata un dato identificativo.[217] La creazione e la diffusione di immagini alterate senza consenso possono dunque costituire una violazione della normativa sulla protezione dei dati, comportando conseguenze di carattere civile o amministrativo. Nel quadro europeo, si applicano inoltre i tradizionali diritti di personalità e le norme sulla diffamazione, che forniscono strumenti per perseguire chi utilizza contenuti manipolati con IA con l'obiettivo di danneggiare la reputazione altrui. Paesi come l'Italia già dispongono di leggi che tutelano il diritto all'immagine, come l'articolo 10 del Codice Civile e le disposizioni della legge sul diritto d'autore (L. 633/1941), le quali, benché risalenti, possono essere invocate anche contro la pubblicazione di contenuti generati dall'IA senza il consenso dell'interessato.

L'Unione Europea ha inoltre intrapreso sforzi significativi per aggiornare le proprie norme, come dimostrato dall'elaborazione del già citato AI Act, il quale mira a classificare e regolamentare i sistemi di intelligenza artificiale in base al loro livello di rischio, includendo le tecnologie Deepfake fra quelle ad alto rischio, con l'intento di introdurre obblighi di trasparenza e limitazioni più stringenti.

[217] Cfr GDPR, nell'articolo 4, numero 1

In particolare, dando una prima definizione all'art 3 punto 60[218], successivamente all'articolo 50, paragrafo 4, discute gli obblighi di trasparenza per i fornitori e gli utilizzatori di sistemi di IA che generano Deepfake:

> *"Gli utilizzatori di un sistema di IA che genera o manipola contenuti di immagini, audio o video che costituiscono un Deepfake, devono rivelare che il contenuto è stato generato o manipolato artificialmente."*

A completare questo panorama troviamo il Digital Services Act e il Digital Markets Act, che impongono alle grandi piattaforme online di adottare misure di moderazione, prevenendo la diffusione di contenuti illegali, fra cui le immagini manipolate con IA che violano la privacy o i diritti d'immagine.

In particolare, il Digital Services Act (DSA) affronta concetti correlati alla diffamazione e alla manipolazione illecita di immagini, come la diffusione di contenuti illegali online e la disinformazione, che potrebbero essere applicati a questi problemi specifici.

Il DSA definisce i contenuti illegali come qualsiasi informazione che viola la legge dell'Unione Europea o di uno Stato membro. Quindi, se la diffamazione o la manipolazione dell'immagine sono considerate illegali secondo la legge applicabile, potrebbero rientrare nell'ambito di applicazione del DSA.

Il DSA, similmente a quanto detto anche in un precedente paragrafo (§ .2) introduce obblighi per i prestatori di servizi di memorizzazione di informazioni, come le piattaforme di social media, per stabilire meccanismi di segnalazione e azione. Questi meccanismi consentono agli utenti di segnalare contenuti potenzialmente illegali, compresi quelli che potrebbero essere diffamatori o che coinvolgono immagini manipolate.[219] Inoltre, il DSA affronta l'abuso di questi meccanismi, ad esempio presentando segnalazioni manifestamente infondate.

Oltre ai contenuti illegali, il DSA si concentra anche sulla manipolazione online, che include le pratiche che distorcono o compromettono la capacità degli utenti di prendere

[218] Ivi nota 186
[219] Regolamento 2022/2065 considerando 50 e 52

decisioni autonome e informate.[220] La manipolazione dell'immagine per scopi di disinformazione potrebbe rientrare in questa categoria.

Il DSA, similmente a quanto previsto dall'AI act richiede anche ai fornitori di piattaforme online di grandi dimensioni di valutare i rischi sistemici, tra cui la disinformazione, e di adottare misure di mitigazione appropriate.

In Italia, oltre alle norme già citate, il Codice della Privacy (D.Lgs. 196/2003) e la sua integrazione con il GDPR offrono ulteriori tutele, consentendo alla persona lesa di chiedere la rimozione dei contenuti illeciti e un eventuale risarcimento del danno. Un altro esempio è la cosiddetta legge "Codice Rosso" (L.69/2019), pensata principalmente per contrastare quelli che sono i reati di violenza, atti persecutori e maltrattamenti.

Tra i reati appena citati, rientrano il "Revenge porn" e gli atti persecutori, ovvero quella condotta per la quale si pubblichi o diffonda materiale sessualmente esplicito senza un consenso espresso del rappresentato, questione della quale parleremo meglio nel prossimo sottocapitolo.[221]

Fuori dall'Europa, gli Stati Uniti registrano interventi normativi a livello statale, con la promulgazione di leggi specifiche in California[222], Texas[223] o Virginia[224] volte a combattere l'uso di materiali a fini pornografici non consensuali, e conseguentemente anche volte a combattere il fenomeno del Deep porn.

In Cina sono state adottate regole mirate al cosiddetto "deep synthesis content", che obbligano i fornitori a etichettare chiaramente i contenuti generati da IA per limitare la disinformazione e proteggere i diritti all'immagine e all'identità.[225] Nel Regno Unito e in altri

[220] Ivi considerando 67

[221] Caso molto discusso fu quello riguardante piattaforme che permettono la diffusione di materiale esplicito (come applicazioni di incontri per adulti o anche piattaforme con OnlyFans o similari), la cassazione ci ha permesso di comprendere che la diffusione di materiale giunto tramite queste piattaforme costituisce fattispecie integrante del reato di revenge porn. Cfr Cass. pen. Sez. V, Sent., (ud. 05/03/2024) 27/06/2024, n. 25516

[222] California ab 2013/24

[223] Texas Penal code title 5-chapter 21 sec 21.16. Unlawful Disclosure or Promotion of Intimate Visual Material

[224] Code of Virginia title 18.2-chapter 8 article 5 Unlawful dissemination or sale of images of another

[225] Questo sforzo legislativo è stato uno dei primissimi sforzi mondiali per normare le IA generative, da un lato assicurando la protezione degli utenti da una diffusione incontrollata dei Deepfake, dall'altro, però, aumenta e rinforza ulteriormente quello che è il controllo governativo sui contenuti di internet, questione che solleva non poche questioni a livello etico, soprattutto vista l'attuale posizione repressiva dell'ordinamento

Paesi sono in corso valutazioni e proposte per introdurre nuove fattispecie di reato che colmino le lacune, come quelle previste dall'Online Safety Act, che ha trovato approvazione in ottobre 2023.[226]

In conclusione, il sistema normativo internazionale, europeo e italiano non dispone ancora di un corpus unitario dedicato esclusivamente all'abuso dell'IA nella manipolazione dell'immagine. Tuttavia, è evidente una tendenza verso l'aggiornamento del quadro esistente, l'introduzione di nuove disposizioni e l'applicazione estensiva di leggi già in vigore. Le future misure legislative puntano a stabilire obblighi di trasparenza, a etichettare i contenuti creati con IA e a responsabilizzare le piattaforme online, in modo da mitigare i rischi, proteggere la reputazione e l'immagine delle persone e garantire un uso etico e responsabile di queste tecnologie avanzate.

.2 Cyberviolenza e Revenge porn

La cyber-violenza[227] è una forma di violenza realizzata tramite strumenti informatici o telematici, come social network, app di messaggistica, forum, siti web o e-mail, e comprende tutte quelle condotte aggressive, intimidatorie o moleste che sfruttano le potenzialità della Rete per colpire, controllare, diffamare o isolare la vittima.

Sebbene i comportamenti violenti o molesti esistano da sempre, la diffusione di internet e dei dispositivi digitali ha amplificato enormemente le possibilità di perpetrare abusi in modo anonimo o transnazionale, rendendo più difficile la prevenzione e il contrasto.

---

cinese. M Maratona e R. Savella "*AI generativa, la Cina mette i paletti: luci e ombre delle nuove regole*" (10 febbraio 2023) Agenda Digirale <agendadigitale.eu>

[226] Online Safety Act 2023, UK Government Bill, Originated in the House of Commons, Sessions 2021-22, 2022-23

[227] Cfr: F. Di Tano "*I reati informatici e i fenomeni del cyberstalking, del cyberbullismo e del revenge porn*" in Th. Casadei, S. Pietropaoli, "*Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*", Wolters-Kluwer, Cedam giuridica, Milano, 2021, Pp. 165-178

I fenomeni che rientrano nella cyber-violenza sono molteplici: dal cyberbullismo, in cui la vittima viene presa di mira con insulti, minacce o diffusione di contenuti lesivi della reputazione, fino allo stalking informatico, caratterizzato dal controllo ossessivo di e-mail, profili social o telefonate, e dalla reiterazione continua di minacce o pressioni volte a ledere la serenità della persona. Altre forme di cyber-violenza particolarmente insidiose sono il doxxing, ossia la diffusione non autorizzata di dati personali e sensibili per mettere in pericolo o ricattare la vittima, l'istigazione all'odio (hate speech), che prende di mira specifiche categorie di individui o minoranze, e la cosiddetta "pornografia non consensuale" (revenge porn), in cui immagini o video espliciti vengono pubblicati senza il consenso della persona ritratta a scopo di vendetta, ricatto o umiliazione.[228] [229]

La forza della cyber-violenza risiede, oltre che nella potenziale anonimizzazione dell'aggressore, nella capacità di riprodurre, condividere e far circolare i contenuti offensivi in tempi rapidissimi e su larga scala. Ciò può causare un impatto psicologico devastante sulle vittime, che si vedono ripetutamente esposte allo stesso abuso, spesso impossibilitate a bloccare del tutto la diffusione del materiale online.

In Italia, il revenge porn è disciplinato dal Codice penale grazie alla Legge 19 luglio 2019, n. 69, nota come "Codice Rosso", il quale ha inserito l'articolo 612-ter c.p. sanzionando chiunque divulghi, consegni, ceda, pubblichi o diffonda contenuti espliciti destinati a rimanere privati, senza il consenso dei soggetti ritratti, nonché chi, avendone già la disponibilità, diffonda ulteriormente tali contenuti al fine di recare danno. [230]

Le pene previste vanno da uno a sei anni di reclusione e da 5.000 a 15.000 euro di multa, con possibili aggravamenti nel caso di reati commessi dal coniuge, anche separato o divorziato, o da persona con un legame affettivo con la vittima, o ancora se l'atto è perpetrato mediante strumenti informatici e telematici.

Questa fattispecie è facilmente applicabile anche alla fattispecie dei Deepfake a contenuto sessuale o sensuale, ciononostante non vi sia una presa diretta del materiale, ma la

[228] N. Lomba, C. Navarra & M. Fernandes "*Combating gender-based violence: Cyber violence*" (marzo 2021) European Parliamentary Research Service (EPRS). PE 662.621

[229] Cfr: Stolle, Lien. "*The need for protection of environmental defenders from digital intimidation: an analysis of Article 3(8) of the Aarhus Convention*" (agosto 2023) Opolskie Studia Administracyjno-Prawne. Volume 21 (1) Pp. 199-219.

[230] Infra multis Cass. pen., Sez. V, Sent., (data ud. 01/03/2024) 20/06/2024, n. 24379

falsificazione di un video precedentemente diffuso e raffigurante soggetti differenti, o la realizzazione materiale completamente artificiale.

La lettera del codice prevede che la fonte del materiale possa derivare da un'acquisizione qualsiasi; [231] non è rilevante quale sia la fonte, se non per aggravare il reato come descritto nei commi 3 e 4 dell'articolo.

Altri reati che possono configurarsi come forme di cyber violenza sono lo stalking (art. 612-bis c.p.), che copre gli atti persecutori anche commessi online, la diffamazione (art. 595 c.p.), passibile di aggravamento se commessa via Internet, e l'accesso abusivo a un sistema informatico (art. 615-ter c.p.), riconducibile ai casi in cui si acceda illegalmente ai dispositivi altrui per impossessarsi di materiale o diffonderlo illecitamente.

A livello europeo non esiste ancora una normativa specifica dedicata interamente al revenge porn, ma la lotta alla violenza di genere e alle sue forme digitali è promossa da strumenti internazionali come la Convenzione di Istanbul[232], che invita gli Stati membri a proteggere le donne da tutte le forme di violenza, comprese quelle perpetrate tramite le nuove tecnologie.

Diversi atti e raccomandazioni dell'Unione Europea riguardano la protezione dei dati personali, la tutela dei minori e il contrasto alla violenza online, sebbene manchi una direttiva ad hoc sul revenge porn.

La Strategia dell'UE per la Parità di Genere 2020-2025 e la recente Proposta di Direttiva sulla violenza contro le donne e la violenza domestica[233], pubblicata nel 2022, evidenziano la necessità di inquadrare la cyber violenza in una cornice legislativa uniforme, di modo che gli Stati membri adottino disposizioni comuni e più efficaci.

[231] Cfr: Cass. pen. Sez. V, Sent., (ud. 22/02/2023) 07/04/2023, n. 14927

[232] La convenzione è stata approvata dal Comitato dei ministri del Consiglio d'Europa il 7 aprile 2011 e aperta alla firma l'11 maggio 2011. È stato firmato da 45 stati, ma da marzo 2021 il conteggio si è ridotto al 44 vista la revoca della firma Turca. In Italia la convenzione è stata ratificata con la legge n. 77/2013

[233] Proposta di direttiva del Parlamento Europeo e del consiglio sulla lotta alla violenza contro le donne e alla violenza domestica COM/2022/105 final

In parallelo, il Digital Services Act (DSA)[234], entrato in vigore nel 2022, introduce obblighi per le grandi piattaforme digitali in relazione alla rimozione tempestiva di contenuti illeciti, compresa la diffusione non consensuale di materiale a sfondo sessuale.

L'Italia, con la Legge n. 69/2019, ha dunque affrontato il revenge porn in modo piuttosto specifico, garantendo maggiore tutela alle vittime, ma, considerata l'intrinseca dimensione transnazionale della rete, è fondamentale un'armonizzazione a livello europeo per contrastare con più efficacia la cyber violenza.

A tal fine, occorre non solo delineare un quadro normativo omogeneo ma anche investire nell'educazione, nella sensibilizzazione e nei servizi di assistenza psicologica e legale, rendendo più semplice segnalare i reati e far rimuovere i contenuti offensivi.

Le piattaforme digitali, dal canto loro, hanno cominciato a dotarsi di procedure più efficaci di segnalazione e rimozione dei contenuti illeciti, incentivate anche dai nuovi regolamenti comunitari, come il Digital Services Act, che introduce obblighi più stringenti per i fornitori di servizi online.

Sulla stessa linea di gravità rispetto al reveng porn si può parlare di "sextortion"; [235] [236] ovvero una forma di estorsione a sfondo sessuale che si verifica quando qualcuno utilizza immagini o video intimi di un'altra persona per ricattarla. Generalmente, il ricattatore minaccia di diffondere o inviare a terzi (familiari, amici, datori di lavoro) il materiale compromettente se la vittima non acconsente a soddisfare determinate richieste, che possono variare dal pagamento di una somma di denaro all'invio di ulteriori immagini o video espliciti, fino a richieste di natura sessuale.

Spesso, il materiale utilizzato per la sextortion viene ottenuto tramite chat private, videochiamate, social network o app di incontri, anche perché, in molti casi, la vittima condivide inizialmente immagini o video in un contesto di fiducia o intimità. In altri casi, il

[234] Ivi nota 168

[235] Il primo utilizzo italiano di questo termine avviene a cura di Marta Serafini, nell'articolo "*L'hacker che rubava le vite via webcam*" de il Corriere.it, sez. Esteri il 15/3/2012; accademia della Crusca, sezione parole nuove < accademiadellacrusca.it.>

[236] Intervento di Guido Scorza, componente del Garante per la protezione dei dati personali "*Contrasto alla pornografia non consensuale: istruzioni per l'uso*" (5 dicembre 2022) HuffPost

materiale è frutto di hacking (ad esempio, furto di credenziali di accesso a un account cloud) o derivato dall'installazione di malware o spyware sui dispositivi personali.

La sextortion può causare gravi ripercussioni sul piano psicologico ed emotivo: ansia, stress, senso di colpa, vergogna e paura delle conseguenze sociali o professionali di una possibile diffusione del materiale. A questo si aggiunge il timore di ritorsioni da parte del ricattatore in caso di mancata adesione alle richieste. Per tali ragioni, molte vittime esitano a denunciare e possono rimanere intrappolate in un circolo vizioso di ricatti crescenti.[237]

Dal punto di vista legale, nei vari ordinamenti si puniscono l'estorsione e la diffusione illecita di contenuti privati, compresi quelli a sfondo sessuale, in quanto violazioni della privacy e della dignità personale. In Italia, ad esempio, la Legge 69/2019 (il cosiddetto "Codice Rosso") ha introdotto l'art. 612-ter c.p. per punire chi diffonde o minaccia di diffondere video e immagini sessualmente espliciti senza il consenso della persona ritratta, configurando uno specifico reato di "revenge porn" che può estendersi alle forme di ricatto.

Tuttavia, la natura transnazionale della Rete e la velocità di propagazione delle informazioni rendono necessarie sia un'armonizzazione delle norme, sia una continua attività di formazione e sensibilizzazione, affinché gli utenti acquisiscano maggiore consapevolezza sui rischi, sui comportamenti da adottare e sugli strumenti disponibili per reagire. Per prevenire e contrastare la cyber-violenza, inoltre, è fondamentale che le vittime trovino un ambiente sicuro in cui denunciare, un sostegno legale e psicologico adeguato e procedure chiare e rapide per la rimozione dei contenuti offensivi. Soprattutto, occorre promuovere una cultura del rispetto e dell'empatia digitale, in grado di rendere gli spazi virtuali più sicuri e inclusivi, contrastando i comportamenti aggressivi e prevenendo quella spirale di odio e violenza che può degenerare in fenomeni di vera e propria persecuzione online.[238]

---

[237] Y. Ruvalcaba & A. Eaton "*Nonconsensual Pornography Among U.S. Adults: A Sexual Scripts Framework on Victimization, Perpetration, and Health Correlates for Women and Men. Psychology of Violence*" (febbraio 2019).

[238] Nonostante in questa trattatazione ci si sia concentrati maggiormente sugli aspetti legali più vicini, ovvero quelli italiani e europei, al fine di concedere uno sguardo alla normativa d'oltre oceano rispetto agli argomenti trattati in questo capitolo, oltre che ad un ulteriore sguardo alla tecnologia dei Deepfake, si consiglia la lettura de: Y. Moncarol Wang "*Don't Believe Your Eyes: Fighting Deepfaked Nonconsensual Pornography with Tort Law*" (2023) University of Chicago Legal Forum University of Chicago Legal Forum; vol. 2022 articolo 16.

.3 La Cyber violenza in danno ai minori

Come si è visto lungo la trattazione fino a questo punto, negli ultimi anni, lo sviluppo di tecnologie basate sull'Intelligenza Artificiale ha rivoluzionato numerosi settori, dalla medicina alla finanza, dall'istruzione all'industria manifatturiera. Grazie alla capacità di apprendere dai dati e di automatizzare processi complessi, i sistemi di IA consentono di migliorare l'efficienza, ottimizzare risorse e personalizzare i servizi. Tuttavia, ogni innovazione tecnologica ha un rovescio della medaglia, e l'IA non fa eccezione. I criminali di tutto il mondo, infatti, stanno rapidamente sfruttando le nuove possibilità offerte dall'apprendimento automatico, dal data mining avanzato e dai modelli generativi per compiere reati sofisticati e, purtroppo, spesso difficili da identificare e contrastare.

Tra i soggetti più vulnerabili di fronte a questo scenario in evoluzione vi sono i minori. Bambini e adolescenti, sempre più connessi e presenti sulle piattaforme digitali, rappresentano un bersaglio facile per cybercriminali di vario tipo: pedofili che praticano grooming, hacker specializzati in furto di identità, truffatori che puntano a estorcere denaro o informazioni, e vere e proprie reti organizzate dedite alla produzione e alla distribuzione di materiale pedopornografico. Se le forme tradizionali di cyberbullismo o grooming online sono già da tempo motivo di preoccupazione per genitori, istituzioni scolastiche e forze dell'ordine, l'impiego di algoritmi di Intelligenza Artificiale apre un capitolo ancora più inquietante nella storia dei crimini informatici.[239]

Uno dei fattori che rende particolarmente insidiosa la combinazione fra tecnologia IA e criminalità è l'alta capacità di adattamento di questi sistemi. Un software in grado di apprendere può infatti migliorare costantemente le proprie strategie di adescamento, individuare in modo mirato le vulnerabilità di un minore (attraverso un'analisi estensiva dei dati condivisi sui social media) e persino falsificare la realtà in modo estremamente realistico. Questo livello di sofisticazione rende i minori ancor più esposti, poiché spesso non possiedono gli strumenti

[239] Cfr: Lappello del direttore della Postale Nunzia Ciardi ad una maggiore attenzione alluso della rete *"Sul web aumentano i reati contro i minori e si abbassa letà anagrafica delle vittime "* (10 Febbraio 2021) ministero dell'interno <www.interno.gov.it>

critici o l'esperienza necessaria per distinguere un profilo autentico da uno generato artificialmente, o per riconoscere un contenuto manipolato da un Deepfake. [240]

Uno dei fenomeni più connessi a questo impiego dell'Intelligenza Artificiale, è quello dei Deepfake. Quando si parla di minori, il problema assume contorni ancora più allarmanti rispetto a quanto visto fin ora. L'utilizzo dei Deepfake in campo pedopornografico consiste, ad esempio, nel prendere il volto di un bambino o di un adolescente e sovrapporlo a quello di un attore in un video pornografico, creando contenuti espliciti che "sembrano" coinvolgere effettivamente il minore. Nonostante la falsità dei contenuti, l'impatto psicologico e sociale sulla vittima può essere devastante. I criminali che realizzano e diffondono tale materiale possono ricattare la famiglia o il minore stesso, minacciando di pubblicare i video su forum clandestini o piattaforme di file sharing specializzate. In casi estremi, queste pratiche possono distruggere la reputazione della vittima e indurre conseguenze a lungo termine, come depressione, disturbi d'ansia e tendenze auto-lesionistiche.[241]

Le operazioni di polizia e gli studi condotti da organizzazioni come Europol e WeProtect Global Alliance mostrano come la disponibilità di software per creare Deepfake sia in costante aumento: esistono persino applicazioni che, con pochi clic, consentono di caricare una foto e generare un video. L'assenza di competenze tecniche avanzate non è più un ostacolo, e questo abbassa la soglia d'ingresso per i criminali, favorendo la crescita del fenomeno. Inoltre, la diffusione dei Deepfake non si limita al contesto pedopornografico: alcuni criminali li utilizzano per effettuare raggiri e truffe economiche, spacciandosi per un genitore (imitandone voce e aspetto) al fine di estorcere denaro o informazioni sensibili al figlio, o viceversa. [242]

Sul piano legale, diversi Paesi sono ancora alla ricerca di normative adeguate per affrontare il problema dei Deepfake. Attualmente, la maggior parte delle legislazioni punisce la diffusione e la creazione di contenuti pedopornografici[243], ma l'applicabilità delle norme ai Deepfake non è sempre chiara. Alcune legislazioni potrebbero non contemplare in modo

---

[240] Europol "*Internet Organised Crime Threat Assessment (IOCTA) 2024*" (2024) Ufficio delle pubblicazioni dellUnione europea , Lussemburgo; Pp: 23 - 26

[241] "*Global Threat Assessment 2023*" (2023) WeProtect Global Alliance <www.weprotect.org>; Pp. 20 - 32

[242] Ivi nota 240 Pp. 19 - 22; e Ivi nota 241 Pp. 34 - 41

[243] I contenuti pedopornografici e lo sfruttamento della pornografia minorile, nel nostro ordinamento, sono puniti agli att. 600 ter e 600 quater codice penale, e introdotti da prima con il DDL approvato 7 novembre 2003 recante: "*Disposizioni in materia di lotta contro lo sfruttamento sessuale dei bambini e la pedopornografia anche a mezzo internet*" e successivamente promulgato con la legge 6 febbraio 2006, n. 38 "*Disposizioni in materia di lotta contro lo sfruttamento sessuale dei bambini e la pedopornografia anche a mezzo Internet*."

esplicito i contenuti artificialmente generati, o non prevedere aggravanti specifiche per il caso in cui siano coinvolti minori. Ciò genera una pericolosa "zona grigia" nella quale i criminali sfruttano la possibilità di difendersi sostenendo la "non reale" natura dei contenuti prodotti.

Differentemente da quanto detto per le altre legislazioni, quella italiana ha previsto un articolo specifico per far fronte a queste evenienze, di fatti l'art. 600 quater 1[244] del Codice penale punisce chiunque detenga o produca materiale pedopornografico virtuale, permettendo una precisa copertura delle zone grige appena descritte.

Un secondo aspetto cruciale riguarda il cosiddetto grooming, ossia l'adescamento di minori via Internet. Il fenomeno, già conosciuto da tempo dalle forze dell'ordine, prevede l'instaurarsi di una relazione di fiducia tra l'aggressore (solitamente un adulto) e il minore, al fine di manipolarlo e ottenere prestazioni sessuali, fotografie o video compromettenti, oppure per condurlo a un incontro fisico.[245] L'uso dell'IA ha introdotto una variante ancora più subdola: l'impiego di chatbot e programmi in grado di simulare conversazioni umane in modo estremamente credibile.

Tali software analizzano i dati raccolti in rete sul minore – dai post sui social network ai commenti su forum e chat, fino alle preferenze e agli orari di connessione – e adattano il proprio stile conversazionale per risultare più attraenti e affidabili. Un chatbot potenziato da algoritmi di Natural Language Processing (NLP) può fingere di condividere gli stessi interessi della vittima, di frequentare la stessa scuola, o addirittura imitare il gergo giovanile per avvicinarsi emotivamente. Ciò è particolarmente insidioso poiché il minore non ha la percezione di interagire con un'entità artificiale: molte di queste conversazioni, per complessità e fluidità, appaiono del tutto normali, se non addirittura più "empatiche" rispetto a quelle con un coetaneo reale.[246]

Un ulteriore elemento di pericolo riguarda la "scalabilità" di questa pratica. Un singolo individuo malintenzionato, grazie all'IA, può avviare contemporaneamente decine o addirittura

[244] Articolo inserito nel nostro codice penale dalla legge del 6 febbraio 2006, n. 38; il fatto che questa legge sia stata promulgata ben 11 anni prima dello sviluppo del primo deepfake, non toglie valore alla copertura, anzi, permette di comprendere la migliore visione d'insieme del legislatore nell'approcciarsi alla pericolosità del mezzo "internet"; di fatti non viene mai esplicitato il metodo di realizzazione del materiale, permettendo al nostro ordinamento una copertura anche rispetto ad eventi e tecniche lesive che potrebbero sorgere in futuro.

[245] Ivi nota 240 P. 25

[246] V. Colomba e A. Souihel "*Child grooming: come riconoscerlo e come difendersi*" (14 febbraio 2023) Agenda digitale < www.agendadigitale.eu>

centinaia di interazioni con potenziali vittime, massimizzando le possibilità di ottenere informazioni personali, foto intime o inviti a incontri offline. Se anche solo una piccola percentuale dei minori adescati abbocca, il criminale ottiene un "ritorno" elevato in termini di materiale pedopornografico o di potenziale sfruttamento.[247]

La tecnologia di analisi dei dati, spesso designata con termini come data mining o big data analytics, consente ai criminali di profilare i minori a un livello di dettaglio impressionante. Le informazioni pubblicate sui profili social non si limitano alle foto e alle preferenze, ma comprendono la rete di contatti, i luoghi frequentati, le abitudini orarie, i gusti musicali o sportivi. Un algoritmo di IA ben progettato può integrare tutti questi dati e "consigliare" al criminale il miglior approccio psicologico da adottare. In tal senso, il grooming automatizzato diventa quasi una formula "scientifica", basata sulle statistiche di successo di determinate strategie di adescamento.

Le capacità di elaborazione e apprendimento dell'Intelligenza Artificiale trovano impiego anche in forme di estorsione e ricatto che coinvolgono i minori. Un esempio emblematico è l'uso di sistemi di sintesi vocale avanzati, in grado di imitare la voce di un membro della famiglia o di un amico. Immaginiamo un adolescente che riceve una chiamata su una piattaforma VoIP[248] con la voce che sembra, a tutti gli effetti, quella di un proprio coetaneo. Dall'altra parte, invece, potrebbe esserci un software di IA che riproduce frasi pre-registrate o generate al momento sulla base di un copione, il cui obiettivo è ottenere informazioni sensibili, foto compromettenti o addirittura denaro.[249]

L'ingegneria sociale, cioè l'insieme di tecniche volte a manipolare psicologicamente la vittima, viene così potenziata dall'IA fino a livelli prima impensabili. Non è raro che i criminali uniscano la simulazione vocale o la creazione di video falsificati (Deepfake) con la capacità di data mining per produrre "prove" incriminanti nei confronti della vittima. Il minore potrebbe ricevere un file video in cui, apparentemente, compare in comportamenti imbarazzanti o anche illegali. In realtà, si tratta di un contenuto generato artificialmente, ma l'impatto psicologico è

---

[247] E. Caffo per telefono azzurro "*dossier abuso 2024*" (novembre 2024) MAG Studio Milano; p. 28. Cfr E. Caffo per telefono azzurro "*la dignità dei bambini nel mondo digitale*" (5 maggio 2024) MAGcom srl; Pp. 32 - 40

[248] Per una più semplice comprensione: sono piattaforme VoIP tutti quei servizi che offrono la possibilità di effettuare telefonate attraverso protocolli Internet, a questo proposito, è facile trovare correlazione con applicazioni come Whatsapp, Instagram, Snapchat, Skype, Google Meet e molti altri.

[249] Cfr: R. Rogers "*Come non cascare nelle truffe telefoniche AI*" (10 aprile 2024) wired.it <www.wired.it>

devastante: la vittima, spaventata dalla possibile condivisione pubblica di quel video, può essere spinta a sottostare alle richieste del ricattatore.[250]

Queste richieste possono variare dal pagamento di somme di denaro all'invio di nuove foto intime, passando per la partecipazione a chat con contenuti sessuali o persino incontri di persona. L'estorsione può anche spingere il minore a reclutare altri coetanei, creando una catena di ricatti che si autoalimenta.[251] Le statistiche del National Center for Missing & Exploited Children (NCMEC) mostrano come, negli ultimi anni, siano aumentate le segnalazioni di sextortion, cioè di estorsioni a sfondo sessuale che coinvolgono minori.[252] L'uso dell'IA ha reso la realizzazione di materiali falsi molto più semplice, e la pressione psicologica che ne deriva ha avuto conseguenze drammatiche, in alcuni casi culminate in atti di autolesionismo o suicidio.

Le forze dell'ordine si trovano davanti a sfide enormi nel contrastare queste pratiche. Un criminale che opera da un Paese straniero può facilmente utilizzare reti virtual private network (VPN) e servizi di anonimizzazione, mentre l'IA genera innumerevoli trappole o "copie" di sé stessa, riducendo la possibilità di tracciare l'autore del reato. In molti casi, per la vittima e per la famiglia risulta difficile addirittura capire di essere caduti in un raggiro, poiché le prove sembrano inoppugnabili e l'imitazione della realtà può essere incredibilmente accurata.[253]

La violenza virtuale esercitata tramite l'Intelligenza Artificiale non si limita a conseguenze di tipo legale o reputazionale, ma si ripercuote in modo incisivo sull'equilibrio psicologico del minore. Nel caso di Deepfake pedopornografici, i bambini possono subire una sensazione di vergogna e colpa profonda, spesso acuita dall'incomprensione dell'ambiente circostante. È frequente che la vittima tema di non essere creduta o che si senta complice di quanto accaduto, senza avere gli strumenti per spiegare che i contenuti "non sono reali". Allo stesso tempo, la pubblicazione di immagini o video falsi in rete è un fenomeno difficilmente

---

[250] "*FBI Launches Sextortion Awareness Campaign in Schools*" (3 settembre 2019) FBI - Federal Investigation Bureau < www.fbi.gov >

[251] NETSmart "*I am a survivor of sextortion*" (2024) National center for missing & exploited children <www.missingkids.org >

[252] In particolare sono stati presi i dati riportati sulla pagina dedicata al CyberTipline report 2023, ove viene riferito che nel 2023 le segnalazioni alla Cyber Tipline sono state 36,2 milioni, ovvero un incremento del 12 % rispetto all'anno precedente; questo dato è sicuramente interessante perché pur riguardando solo gli stati uniti, di quelle milioni di segnalazioni solo 1,1 milioni (ovvero il 4,5% dei casi totali) di segnalazioni sono state riferite alle autorità federali, in quanto la maggior parte delle segnalazioni riguarda materiale sulle quali la forza pubblica statunitense non ha potere. Cfr "*CyberTipline 2023 Report*" <www.missingkids.org>

[253] Cfr idem nota 240

reversibile, poiché il materiale digitalizzato può circolare per anni attraverso siti di hosting, reti P2P e piattaforme di messaggistica privata.[254] [255]

Per i minori adescati tramite chatbot e strategie di grooming potenziate dall'IA, le conseguenze possono comprendere ansia, disturbi del sonno, crollo dell'autostima e un peggioramento del rendimento scolastico. Alcuni finiscono per allontanarsi dalla vita sociale, sviluppando fobie legate ai contesti che richiedono comunicazione online. Nei casi più estremi, l'isolamento e la disperazione possono sfociare in gesti autolesionistici o tentativi di suicidio, come testimoniano alcuni casi documentati da Telefono Azzurro nelle sue ricerche sulle emergenze dell'infanzia e dell'adolescenza.

A livello familiare, i genitori possono sentirsi impotenti e sopraffatti da una tecnologia che non conoscono a fondo o che cambia più velocemente di quanto loro possano capire. La difficoltà ad accedere a strumenti di segnalazione efficaci, unita alla complessità di coordinarsi con le forze dell'ordine (che devono spesso ricostruire reati transnazionali), genera frustrazione e un senso di fallimento nel proprio ruolo protettivo. Non meno rilevante è il danno sul piano sociale: la presenza di contenuti Deepfake e la facilità di manipolazione generano una sfiducia diffusa nella veridicità delle informazioni, minando ulteriormente la percezione di sicurezza nel mondo digitale.[256]

La lotta contro i crimini virtuali che sfruttano l'IA in danno dei minori richiede uno sforzo coordinato a livello globale, dal momento che i confini geografici hanno poca rilevanza negli spazi digitali. Organizzazioni come Europol, WeProtect Global Alliance e NCMEC sottolineano la necessità di un approccio integrato, che coinvolga legislatori, forze di polizia, piattaforme tecnologiche, scuole e famiglie.

Sul piano legislativo, è urgente colmare le lacune normative riguardo alla produzione e alla diffusione di contenuti artificialmente generati. Diverse nazioni si stanno interrogando su come includere nella definizione di "materiale pedopornografico" anche i file creati dall'IA che raffigurano in modo verosimile minori, indipendentemente dal fatto che questi ultimi esistano o meno nella realtà. Allo stesso tempo, occorre regolamentare l'uso di tecnologie di sintesi

---

[254] F. Vera-Gray, "*Key messages from research on the impacts of child sexual abuse*"(Marzo 2023)Child and Woman Abuse Studies Unit, London Metropolitan University <www.csacentre.org.uk>

[255] K. Chauviré-Geib, J. M Fegert "*Victims of Technology-Assisted Child Sexual Abuse: A Scoping Review*" (14 giugno 2023) Sage Journal, vol. 25(2) Pp.1335–1348.

[256] C. Fisher, A. Goldsmith, R. Hurcombe & C Soares "*The impacts of child sexual abuse: A rapid evidence assessment*" (luglio 2017) IICSA Research Team, Pp: 17 – 19. <www.iicsa.org.uk>

vocale e la creazione di chatbot in grado di impersonare terze persone. Tuttavia, trovare il giusto equilibrio tra tutela dei diritti e libertà di sviluppo tecnologico non è semplice: normative troppo restrittive rischiano di ostacolare l'innovazione, mentre regole troppo lasche lasciano mano libera ai criminali.

Dal punto di vista investigativo, le forze di polizia sono chiamate a formare personale specializzato in analisi forense digitale avanzata, capace di distinguere i Deepfake dai contenuti autentici e di ricostruire le catene di produzione e diffusione dei file. La collaborazione tra diversi Stati risulta essenziale, in quanto i server che ospitano materiale pedopornografico o Deepfake possono trovarsi in Paesi con legislazioni molto meno stringenti in materia di reati informatici. Europol, in particolare, si sta impegnando a creare task force internazionali che uniscano competenze di intelligence, analisti di dati e investigatori specializzati in crimini contro i minori.

Le piattaforme online, dal canto loro, hanno una responsabilità cruciale. Servizi di social networking, di messaggistica istantanea e di video sharing dovrebbero implementare sistemi di rilevamento automatico di contenuti illeciti o sospetti. Negli ultimi anni, si sono diffusi alcuni algoritmi di IA addestrati specificamente al riconoscimento di immagini pedopornografiche, capaci di confrontarle con database noti di materiale illegale. Tuttavia, contro i contenuti generati dal deep learning servono modelli costantemente aggiornati, in una sorta di corsa agli armamenti tra chi cerca di mascherare e chi cerca di smascherare la natura artificiale dei file.

Sul fronte educativo, è fondamentale avviare campagne di sensibilizzazione rivolte ai bambini e agli adolescenti, insegnando loro le buone pratiche in rete e i segnali di allarme di un possibile adescamento. Per quanto riguarda la scuola, sarebbe auspicabile introdurre percorsi formativi, fin dalla primaria, che spieghino cosa sono l'Intelligenza Artificiale, gli algoritmi e i pericoli legati alla manipolazione dei contenuti digitali. I genitori e gli insegnanti, dal canto loro, dovrebbero essere formati sulle dinamiche di grooming e sull'evoluzione degli strumenti di manipolazione online, in modo da poter intervenire tempestivamente se notano cambiamenti nel comportamento dei ragazzi.

La crescente presenza dell'Intelligenza Artificiale nel tessuto sociale ha inevitabilmente aperto scenari inquietanti per quanto concerne la criminalità informatica, in particolare quando a farne le spese sono i minori. I Deepfake, il grooming automatizzato, le estorsioni potenziate dall'ingegneria sociale e la normalizzazione di contenuti erotici che coinvolgono bambini o

adolescenti rappresentano pericoli reali e in rapida evoluzione. Se è vero che la "materia prima" di questo fenomeno è la tecnologia, è altrettanto vero che il suo combustibile principale è la mancanza di consapevolezza e di azioni coordinate tra i vari attori coinvolti.

I minori, in quanto anello debole della catena digitale, necessitano di protezione, educazione e supporto costante. Occorre rafforzare la collaborazione tra istituzioni, forze dell'ordine, piattaforme tecnologiche, scuole e famiglie, non solo per punire i colpevoli, ma anche per prevenire l'insorgere di condizioni che favoriscano il crimine. L'implementazione di leggi specifiche che puniscano la produzione e la diffusione di contenuti falsi a sfondo pedopornografico, l'aggiornamento della formazione per gli agenti di polizia e gli operatori sociali, la creazione di meccanismi di segnalazione e rimozione rapida di tali contenuti e la promozione di una cultura digitale responsabile sono solo alcuni degli interventi prioritari.

Le fonti analizzate, come il "Global Threat Assessment" della WeProtect Global Alliance, l'"Internet Organised Crime Threat Assessment (IOCTA)" di Europol, i report del National Center for Missing & Exploited Children (NCMEC) e le indagini di Telefono Azzurro, mostrano una tendenza all'aumento di casi di crimini online che coinvolgono minori. Sebbene questi dati possano apparire allarmanti, la conoscenza del fenomeno è anche la base per elaborare strategie di contrasto più mirate ed efficaci. La tecnologia IA, se ben orientata, può infatti diventare parte della soluzione, fornendo strumenti di monitoraggio e prevenzione più avanzati di quelli finora disponibili.

Rimane infine un'importante sfida culturale: occorre promuovere una responsabilizzazione digitale diffusa, coltivando nei giovani (e negli adulti) capacità di discernimento critico e consapevolezza dei rischi legati alla condivisione di dati personali e all'interazione con interlocutori sconosciuti. Solo attraverso un'azione sinergica su più fronti sarà possibile mettere un freno all'avanzata di questi crimini virtuali che utilizzano l'Intelligenza Artificiale in danno dei minori e garantire loro un ambiente online più sicuro e protetto.

10 Conclusioni

L'evoluzione tecnologica che si manifesta nell'ambito dell'intelligenza artificiale non si limita a trasformare i processi decisionali in contesti politici e sociali, ma investe anche il panorama delle pratiche commerciali.

Le riflessioni che sono emerse lungo la trattazione indicano come il panorama delle pratiche commerciali stia vivendo una trasformazione profonda, spinta dall'adozione di tecnologie sempre più potenti nell'ambito dell'intelligenza artificiale, del neuromarketing e del nudging. Questa evoluzione influenza radicalmente i processi decisionali dei consumatori, al punto da sollevare importanti questioni etiche e giuridiche.

L'analisi predittiva e la personalizzazione, con cui si ottengono raccomandazioni sempre più mirate, mostrano già in modo evidente come l'uso di algoritmi e di dati personali modifichi la relazione tra azienda e consumatore. Si fa strada una potenzialità persuasiva formidabile, che non si limita più a proporre un prodotto o un servizio sulla base di generiche preferenze, ma giunge a un vero e proprio coinvolgimento cognitivo ed emotivo dell'individuo, basato sull'osservazione e la comprensione di comportamenti e reazioni spesso inconsce.

La riflessione sul nudging evidenzia come questa pratica, in origine pensata per accompagnare le persone verso scelte che migliorino la loro vita senza imporre restrizioni, possa divenire uno strumento ambiguo se implementata in modo non trasparente. Il concetto di "spinta gentile" inizia a mostrare i propri limiti quando, grazie all'uso dell'IA, si trasforma in un sistema di influenze ripetute e sottili, capaci di indurre l'individuo a modificare il proprio percorso decisionale. Se da una parte le intenzioni originarie di Thaler e Sunstein miravano a intervenire in ambiti socialmente rilevanti, dall'altra si rischia di travalicare i confini etici non appena le "spinte" assecondano troppo gli interessi commerciali, ponendo in secondo piano la libertà effettiva di chi subisce l'influenza. La questione più delicata riguarda l'uso di queste tecniche senza la piena consapevolezza dei destinatari, che possono ritrovarsi condotti verso determinate azioni o scelte di acquisto senza averne percepito l'esistenza di una persuasione. Anche il neuromarketing si trova al centro di un dibattito analogo. Le sue origini teoriche poggiano sulla limitata razionalità degli esseri umani, come descritto da Herbert Simon, e da qui l'idea di cogliere le componenti emotive e inconsce che davvero determinano le decisioni. Le tecniche di misurazione come l'EEG, l'fMRI, l'eye tracking o la biometria facciale hanno

aperto la via a una comprensione incredibilmente approfondita delle risposte psicofisiche dei consumatori.

È evidente quanto ciò possa migliorare la qualità delle campagne di marketing, rendendole più efficaci e adatte alle reali esigenze del pubblico. Al contempo, la raccolta di dati così intimi e la possibilità di influenzare persino le emozioni sollevano interrogativi che vanno ben oltre il marketing tradizionale.

La consapevolezza del consumatore rispetto alla profilazione che subisce, la distinzione tra persuasione legittima e manipolazione occulta e la possibilità di esercitare un effettivo controllo sui propri dati assumono centralità in un contesto in cui le preferenze personali possono essere anticipate e plasmate. Quando queste tecniche si integrano con le potenzialità dell'intelligenza artificiale, il loro impatto si amplia ulteriormente. Le IA, attraverso l'analisi di enormi moli di dati e l'elaborazione di algoritmi predittivi, permettono una personalizzazione estrema che può sia migliorare l'esperienza del consumatore sia tramutarsi in uno sfruttamento delle sue vulnerabilità psicologiche. Il neuromarketing con l'IA opera più velocemente, su scala globale, e riesce a scomporre la complessità delle reazioni umane in una serie di pattern che possono essere utilizzati per persuadere. Se a questo si aggiungono le tecniche di manipolazione "sotterranee" come i messaggi subliminali, si comprende come venga messa in pericolo l'autonomia decisionale delle persone, spingendo molti osservatori a chiedere una disciplina giuridica più stringente e una supervisione indipendente dell'uso di questi strumenti. La dimensione etica resta un fattore cruciale.

La trasparenza delle tecnologie è spesso carente, e non sempre i consumatori hanno la possibilità di comprendere il modo in cui dati biomedici o comportamentali siano raccolti, analizzati o riutilizzati. Vi è il rischio di accentuare disuguaglianze già esistenti, poiché i soggetti più vulnerabili, o meno consapevoli degli strumenti digitali, possono diventare i primi bersagli di campagne aggressive. D'altronde, la stessa spinta a una continua ottimizzazione delle strategie promozionali potrebbe condurre le imprese a investire più risorse nel perfezionare la persuasione e meno nel migliorare effettivamente prodotti e servizi, con effetti discutibili sulla qualità generale del mercato.

Per tutte queste ragioni, emerge la necessità di un quadro normativo aggiornato e di linee guida etiche condivise, che possano garantire il rispetto della dignità e della libertà di scelta individuale. Il dibattito sull'uso del nudging, del neuromarketing e delle IA deve perciò

essere affrontato in modo integrato, coinvolgendo non solo le grandi aziende tecnologiche, ma anche i legislatori, gli esperti di etica, le associazioni dei consumatori e l'intera società civile. È infatti cruciale tracciare un confine preciso tra strategie di marketing legittime e pratiche scorrette, impedendo che l'innovazione scientifica e tecnologica si traduca in un futuro in cui la volontà delle persone venga manipolata senza alcun freno. Il percorso non è semplice, ma la consapevolezza crescente dei rischi e la diffusione di una cultura della trasparenza possono promuovere soluzioni in grado di salvaguardare tanto l'interesse economico, quanto l'autonomia e la dignità di ogni individuo.

Gli stessi interrogativi si collegano a un orizzonte normativo ed etico che, nel frattempo, si sta trasformando con grande rapidità, includendo un'attenzione specifica all'IA generativa. La panoramica proposta rivela un panorama normativo ed etico in continua evoluzione, in cui l'intelligenza artificiale, soprattutto quella generativa, è sempre più al centro dell'attenzione di legislatori, istituzioni internazionali e grandi aziende tecnologiche.

L'Unione Europea, con l'AI Act, il GDPR, le Linee Guida Etiche e il Piano Coordinato sull'Intelligenza Artificiale, sta costruendo un sistema di regole organico che punta a disciplinare lo sviluppo e l'impiego dell'IA in modo trasparente e sicuro, tutelando sia i diritti fondamentali sia la spinta innovativa. Pur avendo un carattere di assoluto divenire, queste disposizioni delineano un approccio olistico che ingloba anche l'IA generativa, considerata spesso ad alto rischio se impiegata in contesti delicati come la finanza, la selezione del personale o la sanità.

La novità del metodo europeo risiede nel cercare di armonizzare i princìpi etici con la normativa vincolante, tenendo conto delle raccomandazioni internazionali e puntando alla protezione della privacy, alla spiegabilità dei modelli e alla supervisione umana. Parallelamente, la dimensione internazionale del dibattito è ben evidenziata dalle iniziative di ONU, UNESCO e OCSE, che concentrano l'attenzione soprattutto sugli aspetti etici.

La Raccomandazione UNESCO, pur non essendo un testo vincolante, sottolinea la necessità di garantire trasparenza, responsabilità, inclusione e diversità nei dataset di addestramento.

L'idea di fondo è che le IA generative, capaci di creare contenuti autonomamente, comportino rischi specifici di manipolazione e falsificazione dell'informazione, rendendo

ancora più urgente un monitoraggio costante e l'adozione di linee guida etiche condivise. Le risoluzioni dell'ONU, in particolare quella sull'impatto delle tecnologie dell'informazione e della comunicazione sulla sicurezza internazionale, confermano come la comunità globale percepisca ormai l'IA non solo come un'opportunità di crescita, ma anche come un possibile strumento di destabilizzazione se utilizzata in modo malevolo.

Negli Stati Uniti, il quadro legislativo è più variegato, poiché si articola su livelli statali e federali. La California, con l'AB 2013, si è mossa in direzione di una maggiore trasparenza sui dati di addestramento, imponendo agli sviluppatori di dichiarare l'origine dei dataset, incluse eventuali informazioni protette da copyright. Il Colorado, con la SB 24-205, allarga il discorso all'ambito delle decisioni consequenziali, obbligando a valutazioni d'impatto e a audit specifici per prevenire discriminazioni algoritmiche. Altri stati, come Illinois o New York, hanno adottato leggi mirate alla prevenzione dei bias e alla garanzia di procedure di selezione del personale più trasparenti. Sul fronte federale, il National AI Initiative Act, insieme all'Algorithmic Accountability Act, fornisce le basi per una strategia nazionale che incoraggi lo sviluppo di tecnologie IA avanzate, tutelando però i consumatori e mettendo al centro la responsabilità e la supervisione umana. L'obiettivo è, si può supporre, mantenere la leadership statunitense in materia di innovazione, ma coniugarla con principi di correttezza e di sicurezza dei dati.

Un altro aspetto rilevante è l'impegno delle grandi aziende tecnologiche e dei social network nel fronteggiare fenomeni come i Deepfake e la diffusione di contenuti manipolati. Reddit, Facebook, Instagram, Twitter (ora X), YouTube e TikTok hanno sviluppato politiche specifiche per contrastare i media sintetici ingannevoli, adottando strumenti di rilevamento combinati con l'intervento umano e stringendo collaborazioni con fact-checker esterni.

Le piattaforme si sono rese conto che la fiducia degli utenti è un asset prezioso: per non minare tale fiducia, inseriscono avvisi sui contenuti sospetti, riducono la visibilità dei video manipolati e in certi casi procedono alla loro rimozione completa.

Questo testimonia che le aziende stanno riconoscendo la necessità di intervenire proattivamente, non soltanto per ragioni di responsabilità etica, ma anche per adempiere agli obblighi introdotti dalle nuove normative e per prevenire interventi ancora più stringenti da parte dei legislatori.

La convergenza tra l'azione dei governi e l'impegno delle organizzazioni internazionali e delle aziende private è un segnale di come la materia stia acquisendo una centralità inedita. Ciò implica un continuo aggiornamento delle disposizioni, dato che la tecnologia evolve molto rapidamente e le soluzioni odierne potrebbero non bastare domani. Sono quindi essenziali strategie di governance che promuovano la collaborazione tra pubblico e privato, così come la creazione di standard comuni e l'adozione di buone pratiche di intelligenza artificiale affidabile.

Si ravvisa, però, la necessità di un'educazione diffusa per sviluppatori, imprese e cittadini, affinché tutti gli attori siano consapevoli delle potenzialità e dei limiti dell'IA generativa.

Il quadro complessivo, dalla prospettiva dell'Unione Europea fino a quella statunitense, conferma una crescente attenzione a temi quali la trasparenza, la protezione dei dati, l'impatto etico, la discriminazione algoritmica e la responsabilità.

L'IA generativa, grazie alla sua capacità di produrre contenuti di grande impatto, costituisce una sfida ulteriore, poiché apre scenari che vanno ben oltre la semplice automazione di processi o la raccolta di dati.

Come si è visto lungo la trattazione dei casi (§5), è una tecnologia che tocca la sfera della creazione di informazione e che può plasmare, in misura rilevante, la percezione della realtà. Per questa ragione, appare inevitabile che la regolamentazione continui a rafforzarsi e che si lavori a strumenti sempre più raffinati per prevenire abusi e tutelare i diritti fondamentali. La collaborazione transnazionale, la formazione di standard tecnici condivisi e la costante ricerca di soluzioni innovative saranno indispensabili per assicurare che l'IA generativa porti benefici concreti all'umanità, evitando distorsioni sociali e culturali.

In parallelo, tale quadro giuridico, con le sue evoluzioni, incontra questioni concrete nell'ambito della tutela degli utenti e della salvaguardia dei principi democratici.

L'analisi dell'evoluzione normativa e della prassi applicativa in materia di intelligenza artificiale evidenzia come questa tecnologia, ormai penetrata in quasi ogni aspetto della vita sociale ed economica, sollevi questioni che coinvolgono contemporaneamente la sfera politica, quella commerciale e quella giuridico-autoriale.

La sua capacità di influenzare i processi decisionali e le percezioni degli individui ha indotto diversi Stati, tra cui l'Italia, a interrogarsi su quali meccanismi di tutela e regolamentazione fossero necessari per preservare i diritti fondamentali degli individui, la trasparenza del mercato e la correttezza dei processi democratici. Il fatto che l'intelligenza artificiale possa essere utilizzata tanto per la creazione di campagne propagandistiche e manipolative quanto per lo sviluppo di strumenti di analisi più affinati, sottolinea come questa tecnologia non sia di per sé né solo un'opportunità né solo un pericolo, ma piuttosto un mezzo in grado di potenziare finalità molto diverse tra loro.

Nell'ambito delle ingerenze politiche (§5.1), l'elemento cruciale è la necessità di garantire che l'IA, pur offrendo nuove forme di comunicazione e organizzazione, non sia impiegata per minare i valori democratici o per alterare il libero dibattito elettorale. In questo senso, la volontà di impedire l'uso di tecniche subliminali, la creazione di Deepfake ingannevoli e la diffusione massiva di contenuti falsi costituisce un punto centrale.

Gli strumenti normativi europei, a partire dal GDPR, fino alle disposizioni del Digital Services Act e del Digital Markets Act, evidenziano una tensione costante: da un lato, assicurare un ambiente digitale libero e innovativo; dall'altro, prevenire fenomeni di manipolazione su larga scala, specie laddove la raccolta e l'analisi dei dati degli utenti possono essere indirizzate a obiettivi non trasparenti.

L'Italia, in particolare, ha recepito il GDPR e adottato provvedimenti specifici attraverso il Garante per la protezione dei dati personali, con il fine di sensibilizzare sia le istituzioni sia i cittadini sulla rilevanza di questi temi. In parallelo, il dibattito su una legislazione ancora più mirata riguarda le disposizioni del Regolamento sull'Intelligenza Artificiale (AI Act), in cui si delineano livelli di rischio associati alle diverse applicazioni dell'IA, prevedendo obblighi severi per quelle potenzialmente lesive dei diritti fondamentali o della sicurezza pubblica. Nell'area del neuromarketing, si manifesta una questione altrettanto delicata. Qui la problematica centrale risiede nella capacità di analizzare i processi psicologici e neuronali dei consumatori a fini promozionali, con il rischio di agire su un livello quasi inconscio delle scelte di acquisto e di persuasione.

L'ordinamento europeo non ha ancora elaborato un corpus normativo rivolto esclusivamente al neuromarketing, ma le previsioni del GDPR, l'attenzione ai principi di

trasparenza e consenso informato e le clausole che vietano le manipolazioni subliminali delineano un quadro di tutela che può essere esteso anche a tale ambito.

Il fatto che il termine stesso "neuromarketing" non compaia nelle norme non significa che l'ordinamento sia disattento al fenomeno; semplicemente, l'approccio giuridico tende a definire i confini della liceità di queste tecniche in base ai principi generali di protezione dei dati personali e di rispetto della dignità individuale. Il cammino normativo, tuttavia, è solo agli inizi: mancano disposizioni più specifiche che regolino la sperimentazione neuro-cognitiva, la raccolta dei dati biometrici e la sofisticazione delle campagne pubblicitarie.

È verosimile che, a fronte di un utilizzo sempre più diffuso di strumenti scientifici in ambito commerciale, il legislatore nazionale ed europeo dovrà tornare sul tema, rafforzando gli obblighi di chiarezza e di tutela del consumatore.

Per quanto riguarda il copyright, la riflessione si fa altrettanto complessa. Se da un lato l'Unione Europea, con la Direttiva Copyright (2019/790) e con il successivo recepimento negli ordinamenti nazionali, ha cercato di dare risposte ai problemi posti dal mercato digitale, dall'altro l'emergere di sistemi di IA generativa ha aperto un nuovo fronte di incertezze. Il regime tradizionale del diritto d'autore, fondato sull'idea che l'opera sia espressione di una creatività umana e originale, fatica a identificare se e quando un contributo algoritmico possa considerarsi privo di creatività umana o, al contrario, frutto di una collaborazione uomo-macchina meritevole di protezione.

La controversia si amplia se consideriamo che molti sistemi di IA vengono addestrati su grandi quantità di materiale protetto, talvolta senza il consenso esplicito degli autori. L'eccezione del text and data mining, introdotta dalla Direttiva, permette di utilizzare contenuti protetti ai fini di ricerca, ma non è chiaro come si applichi a fini prettamente commerciali o in contesti in cui i dataset comprendono qualsiasi tipo di creazione artistica o letteraria preesistente. Lo stesso Regolamento sull'IA, pur focalizzandosi sui profili di sicurezza e responsabilità, potrebbe influire sulla circolazione e sull'uso delle opere, imponendo standard di trasparenza a chi sviluppa modelli di intelligenza artificiale.

A monte, rimane però la questione di chi debba rispondere in caso di violazione dei diritti: il programmatore che ha creato il modello? L'utente che genera contenuti? La piattaforma che li diffonde? In definitiva, il quadro che emerge è quello di un insieme

complesso di problemi interconnessi, la cui soluzione richiede un costante adeguamento normativo e una cooperazione tra autorità, imprese e società civile.

L'intelligenza artificiale potenzia possibilità straordinarie di sviluppo e di innovazione, ma al contempo obbliga i legislatori a ripensare categorie consolidate come l'autorialità, l'imputabilità delle condotte, la tutela dalla manipolazione subliminale e la stessa idea di "libertà di scelta" nella sfera politica e commerciale.

La normativa europea e italiana appare in progressiva evoluzione, segnando un percorso che, sebbene incerto, indica la volontà di preservare i fondamenti democratici e i diritti dei singoli, senza arrestare il progresso tecnologico. È proprio questo equilibrio - tra libertà e sicurezza, tra creatività e protezione, tra mercato e diritto - a rappresentare la sfida principale che le istituzioni dovranno affrontare negli anni a venire, con riferimento agli argomenti collegati all'intelligenza artificiale.

Tale sfida si estende anche alla salvaguardia dei diritti fondamentali e dei soggetti più vulnerabili, come emerge nel capitolo 9, in cui vengono discusse questioni collegate alla violazione dell'immagine, al revenge porn, alla cyberviolenza e alla protezione dei minori. Le riflessioni che emergono da questo ampio quadro normativo e giuridico, nonché dalle dinamiche psicologiche e sociali delineate, evidenziano come le nuove tecnologie abbiano trasformato in modo radicale il modo in cui i diritti fondamentali della persona possono essere lesi, soprattutto in contesti quali la violazione dell'immagine, il revenge porn, la cyberviolenza e la protezione dei minori.

Le tutele tradizionali, già presenti negli ordinamenti nazionali e internazionali, stanno evolvendo per rispondere alle sfide poste dall'intelligenza artificiale e dalle sue potenzialità manipolative. Dall'analisi delle disposizioni vigenti e delle iniziative in corso, emerge la necessità di aggiornare continuamente la cornice normativa, sia a livello europeo sia globale, al fine di colmare le lacune che consentono ai criminali di agire indisturbati, approfittando di ambiguità interpretative o di veri e propri vuoti legislativi.

Il capitolo appena citato sottolinea come, sul piano internazionale, non esista ancora un corpo normativo pienamente dedicato all'uso illecito dell'intelligenza artificiale, soprattutto per i temi concernenti la manipolazione delle immagini o per la creazione di contenuti pornografici non consensuali. Si osserva però una tendenza a utilizzare e adattare strumenti già esistenti,

come il GDPR per la protezione dei dati personali, o le leggi sulla diffamazione, che possono estendersi ai casi di contenuti manipolati con IA.

Lo stesso discorso vale per le norme sul diritto d'autore e sul diritto all'immagine, che benché nate in epoche precedenti, si stanno dimostrando ancora oggi un canale di tutela per chi subisce gravi compromissioni della propria reputazione o della propria identità visiva.

Tuttavia, la rapidità con cui le tecnologie di Deepfake si diffondono e l'estrema facilità di utilizzo di software in grado di generare contenuti altamente realistici rendono indispensabile un approccio più puntuale, basato sull'introduzione di obblighi di trasparenza e sull'imposizione di responsabilità a carico delle piattaforme che ospitano tali contenuti.

In questo senso, l'AI Act e il Digital Services Act mostrano un percorso di regolamentazione che tenta di contenere i fenomeni illeciti e di responsabilizzare i fornitori di servizi online, costringendoli a rimuovere tempestivamente i materiali palesemente lesivi dei diritti fondamentali.

Lo stesso quadro evolutivo si registra in ambito di cyberviolenza e revenge porn; l'Italia, con la legge sul "Codice Rosso", ha adottato misure specifiche per sanzionare chi diffonde contenuti sessualmente espliciti senza consenso, riconoscendo anche aggravanti nel caso in cui l'illecito sia commesso attraverso strumenti informatici. Sebbene questa legislazione risponda in modo abbastanza preciso al fenomeno del revenge porn, soprattutto in ambito di relazioni affettive e dinamiche di vendetta o ricatto, permangono ulteriori sfide, specialmente quando la rete diventa la sede di attacchi anonimi e su scala potenzialmente globale.

Come si diceva al capitolo 9, l'incrocio tra revenge porn e Deepfake, dove si manipolano immagini e video per creare materiale pornografico falso, implica che anche persone che non hanno mai prodotto contenuti intimi possano ritrovarsi esposte a forme di violenza sessuale digitale.

È dunque necessario integrare le fattispecie penali, estendendole alla creazione di materiale completamente falsificato ma comunque lesivo della dignità e dell'immagine di chi ne è vittima.

Un aspetto particolarmente drammatico che emerge è la vulnerabilità dei minori, i quali subiscono un rischio elevatissimo di adescamento e sfruttamento, complici la diffusione rapida

delle tecnologie di intelligenza artificiale e la scarsa consapevolezza che bambini e adolescenti hanno della dimensione digitale. La creazione di Deepfake pedopornografici, pur non coinvolgendo fisicamente il minore, genera gravissime conseguenze psicologiche e sociali, oltre a una pericolosa zona grigia in cui i criminali potrebbero trincerarsi dietro la natura "artificiale" del contenuto.

L'Italia, da parte sua, ha introdotto una tutela specifica contro la detenzione e la produzione di materiale pedopornografico virtuale (art. 600 quater 1 c.p.), colmando almeno in parte il rischio di impunibilità. Resta però il fatto che, a livello globale, le reti di criminali che sfruttano tali tecnologie riescono facilmente a oltrepassare i confini, frammentando le responsabilità giuridiche e rendendo più complesso l'intervento delle forze dell'ordine.

Tra i risvolti più preoccupanti vi è anche l'automazione del grooming, potenziato da software che simulano in modo credibile interazioni umane, analizzano i dati dei minori e sfruttano sofisticate tecniche di ingegneria sociale per stabilire rapporti di fiducia a fini di sfruttamento sessuale o ricatto.

La rapidità con cui questi sistemi evolvono e apprendono, così come la facilità di distribuire i materiali creati, costituisce una sfida costante per le autorità, che si trovano a dover combattere fenomeni in cui l'anonimato e la fluidità della rete offrono ai criminali moltissimi appigli. Le diverse organizzazioni internazionali e nazionali citate, come Europol, WeProtect Global Alliance e il National Center for Missing & Exploited Children, stanno lavorando per mettere a punto strategie di prevenzione e contrasto che comprendano non solo l'inasprimento delle pene, ma anche la creazione di task force transnazionali specializzate in analisi forense digitale.

Le piattaforme digitali, inoltre, sono sempre più sollecitate a migliorare i propri sistemi di moderazione e controllo, introducendo algoritmi in grado di riconoscere materiali pedopornografici o manipolati.

Resta comunque fondamentale un intervento a livello educativo: solo creando una cultura digitale improntata al rispetto e alla consapevolezza dei pericoli sarà possibile ridurre in modo significativo il bacino di possibili vittime.

A rafforzare l'attenzione sui profili di sicurezza delle persone, interviene un'ulteriore riflessione che prende in considerazione scenari correlati all'emergere di fenomeni come la

sextortion, la manipolazione digitale e il revenge porn, tutti strettamente legati all'avanzamento tecnologico.

Nell'era digitale, infatti, la manipolazione attraverso l'intelligenza artificiale ha contribuito all'ampliamento di rischi potenzialmente devastanti per la privacy e la sicurezza individuale, esponendo un numero crescente di persone a nuove forme di violazione e di ricatto. Guardando al futuro, è chiaro che sarà essenziale implementare regolamentazioni più rigide.

Le autorità governative e le organizzazioni internazionali sono sempre più consapevoli della necessità di legiferare in modo più stringente per regolamentare l'utilizzo etico delle tecnologie emergenti come l'intelligenza artificiale, nonché per punire severamente i trasgressori nel contesto del revenge porn. Queste misure legislative sono cruciali per creare un ambiente digitale più sicuro e responsabile.

Parallelamente alla legislazione, l'industria tecnologica è impegnata nello sviluppo di tecnologie di rilevamento all'avanguardia. Questi nuovi strumenti sono progettati per identificare, segnalare e talvolta rimuovere automaticamente contenuti falsi o manipolati. Sistemi sofisticati sono in fase di sviluppo per autenticare video e immagini, e per intercettare contenuti inappropriati come il revenge porn prima che possano fare danni significativi; ovvero per permettere una più agile segnalazione del fenomeno nel caso in cui i materiali non fossero stati rilevati tempestivamente; in questo contesto possiamo certamente nominare alcuni progeti, come l'osservatorio volto alla lotta contro il revenge porn e del materiale intimo senza che vi sia un consenso a fondamento della diffusione stessa (permesso negato[257]).

Oltre alla tecnologia e alla legislazione, l'educazione e la sensibilizzazione pubblica giocano un ruolo fondamentale. Le campagne di sensibilizzazione e gli eventi educativi sono essenziali per equipaggiare le persone con le conoscenze necessarie per riconoscere gli abusi digitali.

Comprendere come identificare e dove segnalare questi abusi è cruciale, così come è importante che tutti comprendano i loro diritti nell'ambito digitale. Una collaborazione interdisciplinare tra il settore tecnologico, i legislatori e le organizzazioni per i diritti umani è

[257]Permesso Negato è una realtà privata nata nel 2019, per offrire supporto tecnologico e orientamento legale alle vittime di diffusione non consensuale di materiale intimo e violenza online. Questo permette di coprire sia il vuoto per la segnalazione che per il vuoto educativo. <https://www.permessonegato.it/>

altrettanto vitale. Queste entità devono lavorare insieme per mitigare i rischi associati all'uso improprio delle tecnologie.

Le piattaforme di social media e i provider di servizi Internet, in particolare, hanno una grande responsabilità nel monitorare attivamente i loro ambienti per rimuovere i contenuti dannosi il più rapidamente possibile.

Per proteggere le vittime di questi crimini digitali, il supporto legale e psicologico è imprescindibile. Molte organizzazioni, sia governative che non profit, offrono servizi di consulenza legale e supporto psicologico, spesso gratuiti o a basso costo, per aiutare le vittime a navigare attraverso le sfide legali e a recuperare dal trauma psicologico subito. Inoltre, esistono hotline e servizi online che servono come primi punti di contatto per le vittime, offrendo assistenza immediata e indirizzandole verso ulteriori risorse di supporto. Anche le opzioni per la rimozione dei contenuti sono fondamentali.

Le normative in molti Paesi impongono alle piattaforme online di agire rapidamente per rimuovere i contenuti illegali, come quelli legati al revenge porn, poco dopo essere stati segnalati. Le piattaforme stesse spesso adottano politiche interne per facilitare una rapida rimozione di tali contenuti. Infine, la protezione della privacy attraverso tecnologie specifiche è un altro strumento cruciale per la difesa individuale. L'uso di VPN, la crittografia end-to-end nelle comunicazioni e gestori di password affidabili sono solo alcune delle misure che gli individui possono adottare per proteggere le proprie informazioni personali da accessi non autorizzati e usi impropri.

In definitiva, le conclusioni che si traggono evidenziano uno scenario complesso e in continua evoluzione, in cui la protezione dell'immagine, della reputazione e dei diritti dei minori richiede azioni mirate e su molti livelli.

I progressi normativi compiuti dall'Unione Europea, così come le riforme introdotte da singoli Paesi, indicano una crescente presa di coscienza delle problematiche legate all'abuso dell'IA e alla cyberviolenza, ma rimane un lungo cammino da percorrere. L'obiettivo è costruire un sistema che sappia cogliere le opportunità positive dell'innovazione digitale, garantendo al tempo stesso la sicurezza e la dignità delle persone, in particolare di quelle più esposte e fragili come i minori. L'unico modo per affrontare la complessità di questi crimini online è procedere con un approccio globale, che includa la cooperazione internazionale,

l’aggiornamento continuo delle leggi, l’educazione ai rischi e l’impiego di strumenti tecnologici avanzati per la prevenzione e l’individuazione tempestiva degli illeciti. Un simile sforzo corale potrà realmente arginare fenomeni come il deep porn e il revenge porn, tutelare il diritto all’immagine e alla privacy, e contrastare la minaccia crescente della violenza digitale in tutte le sue forme, specialmente quando sono i minori a trovarsi maggiormente esposti.

In conclusione, le prospettive future per combattere questi crimini digitali sono orientate verso un approccio integrato che include una legislazione più forte, avanzamenti tecnologici per il rilevamento e la prevenzione, una maggiore educazione pubblica e un sostegno robusto per le vittime, affrontando così efficacemente sia la prevenzione sia la mitigazione degli impatti di questi gravi problemi digitali. L’insieme di tutto quanto trattato fornisce dunque un quadro vasto e articolato, ma al tempo stesso un messaggio chiaro: il progresso dell’IA e di altre tecnologie non può prescindere da un impegno costante a livello normativo, formativo e sociale, per garantire la tutela della dignità umana, la sicurezza di chi naviga nel mondo digitale e il rispetto dei diritti di ciascun individuo, a cominciare dai soggetti più vulnerabili.

Riferimenti biblio-sitografici

Bibliografia

A. Barak, W. A. Fisher, S. Belfry, & D. R. Lashambe "*Sex, Guys, and Cyberspace: Effects of Internet Pornography and Individual Differences on Men's Attitudes Toward Women*" (1999) Journal of Psychology & Human Sexuality, vol. 11(1). DOI: 10.1300/J056v11n01_04

A. Ceschi, R. Sartori, "*Un approccio empirico per una tassonomia delle euristiche*" (1° gennaio 2021), Università di Verona

A. Cooper, E. Griffin-Shelley, D.L. Delmonico & R.M. Mathy "*Online Sexual Problems: Assessment and Predictive Variables.*" (2001). Sexual Addiction and Compulsivity, vol 8(3–4). DOI: 10.1080/107201601753459964

A. Cooper, N. Galbreath, & M.A. Becker, "*Sex on the Internet: Furthering Our Understanding of Men with Online Sexual Problems*." (2004). Psychology of Addictive Behaviors, vol. 18(3). DOI: 10.1037/0893-164X.18.3.223.

A. Gramsci, "*Quaderni del carcere*" (1971).

A. Maslow, "*Motivation and Personality*" (1954) Harper & Row, pubblicato digitalmente da Digital Library of India (31/12/2005) <www.archive.org>

A. Qobraty "*'Star Wars' fans love CGI Luke Skywalker, but deepfake implications are dangerous*" (3 marzo 2022) The daily Targum <www.dailytargum.com>.

A. Shaji George e A. S.Hovan George "*Deepfakes: The Evolution of Hyper realistic Media Manipulation*", (dicembre 2023), in Partners Universal Innovative Research Publication (PUIRP), Vol 1(2), DOI: 10.5281/zenodo.10148558

A. Smidts, A. T. H. Pruyn & C. B. M. Van Riel, "*The impact of consumer advertising on product awareness and perception*" (2001) Academy of Management Journal, Vol. 49(5) DOI: 10.2307/3069448

A. Turing, "*Computing Machinery and Intelligence*", (ottobre 1950), Mind, Vol. 49(236):

A.Tversky, D. Kahneman, "Judgment under uncertainty: heuristics and biases", in Prospect Theory: An Analysis of Decision Under Risk (1979), Econometrica 47, n. 2 DOI: 10.2307/1914185.

Aa.vv. "*Report of The Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns And Interference In The 2016 U.S. Election*" (10 novembre 2020) Vol. 1: Russian Efforts Against Election Infrastructure With Additional Views" Commitee sensitive – russia investigation only, 116th congress, 1st session.

Aa.vv., "*Report on The Investigation Into Russian Interference In The 2016 Presidential Election*", (marzo 2019), vol 2, U.S. Department of Justice, Washington D.C.

AAAI (American association of artificial intelligence) 1948

B. Oakley & O. Kenneth, "*Alvery: Britain's Strategic Computing Initiative*", (26 aprile 1990), MIT Press.

C. Cadwalladr, & E. Graham-Harrison, "*Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.*" (2018). The Guardian.

C. Fisher, A. Goldsmith, R. Hurcombe & C Soares "*The impacts of child sexual abuse: A rapid evidence assessment*" (luglio 2017) IICSA Research Team; ISBN 978-1-5286-3068-0 <www.iicsa.org.uk>

D. F. Aranha & J. van de Graaf, "*The Good, the Bad, and the Ugly: Two Decades of E-Voting in Brazil*" (Nov.-Dic. 2018) in IEEE Security & Privacy, vol. 16 (6), DOI: 10.1109/MSEC.2018.2875318.

D. Kahneman, "*Pensieri lenti e veloci*" (2012), Mondadori, Milano. ISBN: 9788804736127

E. Albrecht, E. Fournier-Tombs, & R. Brubaker, "*Disinformation and Peacebuilding in Sub-Saharan Africa: Security Implications of AI-Altered Information Environments*", (2024.) New York: United Nations University Centre for Policy Research.

E. Cheli, "*La realtà mediata*" (6° edizione 2002) FrancoAngeli srl, ISBN: 9788820475550

E. R. Murphy, J. Illes & P.B. Reiner "*Neuroethics of neuromarketing*." (2008) Journal of Consumer Behaviour vol. 7(4-5). DOI: 10.1002/cb.252

E. Sigel "P*redictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die di Eric Siegel*" (Febbraio 2016), John Wiley & Sons. ISBN: 978-1-119-14567-7

E. Sigel "*Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die di Eric Siegel*" (Febbraio 2016), John Wiley & Sons. ISBN: 978-1-119-14567-7

E. Tuccari "*Neuromarketing: Un'asistematica Disciplina … Oltre Il Consenso?*" (14 ottobre 2024) Persona E Mercato, vol. 2

Europol "*Internet Organised Crime Threat Assessment (IOCTA) 2024*" (2024) Ufficio delle pubblicazioni dell'Unione europea, Lussemburgo.

F fiore "*Bias cognitivi: come la nostra mente ci inganna*" (16 maggio 2018) Sigmund Freud University (Milano). < https://milano-sfu.it/ >

F. Di Tano "*I reati informatici e i fenomeni del cyberstalking, del cyberbullismo e del revenge porn*" in Th. Casadei, S. Pietropaoli, "*Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*" (2021) Wolters-Kluwer, Cedam giuridica, Milano.

F. G. Hoffman, "*Conflict in the 21st Century: The Rise of Hybrid Wars*" (2007) Potomac Institute for Policy Studies.

F. H. Hsu, "*Behind Deep Blue: Building the computer that defeated the world chess champion.*" (2002), Princeton University Press.

F. Machlup, "*The Production and Distribution of Knowledge in the United States*" (1962). Princeton University Press.

F. Pozzi "*Digital Nudge. L'architettura delle scelte nei contesti digitali*" (luglio 2022) L'edizioni. ISBN: 978-8855264631

G. Gerbner, & L. Gross, "*Living with Television: The Violence Profile*" (1976). Journal of Communication. Vol. 26.

G. Le Bon, "*La psicologia delle folle*" (1895).

"*Global Threat Assessment 2023*" (2023) WeProtect Global Alliance <www.weprotect.org>.

G. Sartor, F. Lagioia "*The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*" (2020). Brussels, European Parliament. DOI: 10.2861/293.

G. Zaltman "*How customers think: Essential insights into the mind of the market*." (2003). Harvard Business Press.

H. A Simon, "*Models of man; social and rational*" (1957) Wiley.

H. Zhenliang et al., "*AttGAN: Facial Attribute Editing by Only Changing What You Want*", (11 novembre 2019) IEEE transaction on image progression, vol. 28. DOI: 10.1109/TIP.2019.2916751.

I. Goodfellow et al., "*Generative Adversarial Networks*", (novembre 2020) ACM Digital Library, communication of the ACM, vol. 63 (11). DOI: 10.48550/arXiv.1406.2661.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, "*Generative Adversarial Nets. Advances in Neural Information Processing Systems*" (2014). DOI: 10.48550/arXiv.1406.2661.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville & Y. Bengio, "*Generative Adversarial Nets. Advances in Neural Information Processing Systems*" (2014). DOI: 10.48550/arXiv.1406.2661.

J. Lighthill, "*Artificial Intelligence: A General Survey*". (1973). Artificial Intelligence: a paper symposium. Science Research Council.

J. M. Twenge, "*iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy – and Completely Unprepared for Adulthood*" (2017), Simon and Schuster.

J. McCarthy, M.L. Minsky, N. Rochester & C.E. Shannon, "*A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*", (2006) AI Magazine, vol 4.

J. McGonigall "*La realtà in gioco: Perché i giochi ci rendono migliori e come possono cambiare il mondo*" (marzo 2011) Apogeo. ISBN: 978-8850330409

J. Piaget, "*Il giudizio morale nel fanciullo*" (1932).

J. S. Carroll, L. M. Padilla-Walker, L. J. Nelson, C. D. Olson, C. McNamara Barry & S. D. Madsen "*Generation XXX: Pornography acceptance and use among emerging adults*." (2008) Journal of adolescent research vol. 23(1). DOI: 10.1177/0743558407306348

K. Chauviré-Geib, J. M Fegert "*Victims of Technology-Assisted Child Sexual Abuse: A Scoping Review*" (14 giugno 2023) Sage Journal, vol. 25(2). Doi: 10.1177/15248380231178754

K. Kawon, K. Woo Gon e L. Minwoo "*Impact of dark patterns on consumers' perceived fairness and attitude: Moderating effects of types of dark patterns, social proof, and moral identity*" (ottobre 2023) Elesevier, vol 98 edizione online. DOI: 10.1016/j.tourman.2023.104763

K. Lewin, "*Field Theory in Social Science*" (1947).

K. R. Prajwal et al, "*Towards Automatic Face-to-Face Translation*" (October 2019), Proceedings of the 27th ACM International Conference on Multimedia. DOI: 10.1145/3343031.3351066

Kohavi, Ron; Xu, Ya; Tang, Diane "*Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing.*" (22 ottobre 2021) Cambridge University Press

L. Portinale "*Intelligenza Artificiale: storia, progressi e sviluppi tra speranze e timori*" (28 gennaio 2022) Media Laws, Focus: Blockchain e Intelligenza Artificiale vol. 3/2022.

L. Sze-Fung, "*Deepfakes with Chinese Characteristics: PRC Influence Operations in 2024*", (29 marzo 2024) jamestown.org, China Brief Vol. 24(7).

M. Brundage, S. Avin, J. clark et alt. "*The Malicious Use of Artificial Intelligence : Forecasting, Prevention, and Mitigation*" (febbraio 2018)

M. Di Paolo Emilio, "*Intelligenza artificiale, machine learning e deep learning: quali sono le differenze*", (19 dicembre 2022), Innovation Post.

M. Liu et al., "*STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing*", (2019), IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

M. McCombs &D.L. Shaw, "*The Agenda-Setting Function of Mass Media*" (1972). Public Opinion Quarterly. Vol. 36.

M. N. Schmitt, Nato "*Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*" (febbraio 2017) Cambridge University Press, ISBN: 9781316822524; DOI: 10.1017/9781316822524

M. Wilding, C Milmo et al. "*Met police computers access 'dangerous' facial recognition search engine*" (10 maggio 2024) liberty investigate <libertyinvestigates.org>.

McCarty et al., "*A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*" (31 agosto 1955), e riedito in AI Magazine, 2006, vol.4.

Minsky e Papert, "*Perceptrons: An Introduction to Computational Geometry*", (1969), Cambridge MA,

N. Ciardi "*Sul web aumentano i reati contro i minori e si abbassa l'età anagrafica delle vittime*" (10 Febbraio 2021) ministero dell'interno <www.interno.gov.it>

N. Lomba, C. Navarra & M. Fernandes "*Combating gender-based violence: Cyber violence*" (marzo 2021) European Parliamentary Research Service (EPRS). PE 662.621

N. Smuha et alt. "*orientamenti etici per un'IA affidabile*" (8 aprile 2019) Commissione europea

P. McCorduk, "*Machines Who Think: A personal inquiry into the history and prospects of aritficial intellignece*",2004, CRC Press,

P. Renvoisé & C. Morin "*Neuromarketing: Understanding the Buy Buttons in Your Customer's Brain.*" (2007) Thomas Nelson Inc.

P.W. Singer, E.T. Brooking, "*LikeWar: The weaponization of social media*." (2018) Eamon Dolan Book

R. B. Cialdini, "*Influence: Science and Practice*" (29 luglio 2008, quinta edizione). Allyn & Bacon

R. F. Baumeister, K. D. Vohs, & D. M. Tice, "*The Strength Model of Self-Control*" (2007).

R. H. Thaler e C. R. Sunstein "*Nudge: Improving Decisions About Health, Wealth, and Happiness*" (2008) Penguin Books (riedito 2009) ISBN: 978-0143115267

R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, & J.Lederberg. "*Dendral: A Case Study of the First Expert System for Scientific Hypothesis Formation. Artificial Intelligence*" 61, 2 (1993)

R. M. S. Wilson, J. Gaines & R. P. Hill "*Neuromarketing and consumer free will*." (2008) Journal of Consumer Affairs. Vol. 42(3)

R. Mirza, "*How AI deepfakes threaten the 2024 elections*" (16 febbraio 2024) The journalist's Resource.

R. Montinaro "*Riconoscimento Delle Emozioni E Marketing Personalizzato*", Persona e mercato, 2024, n. 3. ISSN: 2239-8570

R. Viale "*La razionalità limitata "embodied" alla base del cervello sociale ed economico.*" (2019) Il Mulino.

S. Aral & D. Roy, "*The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health - and How We Must Adapt*" (2020). Crown Publishing Group.

S. E. Asch, "*Effects of Group Pressure upon the Modification and Distortion of Judgments*" (1951).

S. J. Russell, & P. Norvig, "*Artificial intelligence: a modern approach.*" (2016), Pearson.

S. Milgram, "*Obedience to Authority: An Experimental View*" (1974).

Sharad Agarwal "*Neuromarketing for Dummies*" (2014), Journal of Consumer Marketing. Vol. 31(4). DOI: 10.1108/JCM-12-2013-0811

Stolle, Lien *"The need for protection of environmental defenders from digital intimidation: an analysis of Article 3(8) of the Aarhus Convention*" (Agosto 2023) Opolskie Studia Administracyjno-Prawne. Vol. 21 (1). DOI: 10.25167/osap.4980

V. Frosini "*Il giurista e le tecnologie dell'informazione*", Bulzoni editore, Roma, 1998,

V. Thambawita, J.L. Isaksen, S.A. Hicks, et al. "*DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine*." (9 novembre 2021) Nature. Vol. 11, art. n. 21896. DOI: 10.1038/s41598-021-01295-2

Y. Bengio, I. Goodfellow & A. Courville, "*Deep learning*" (18 novembre 2016) Cambridge, MIT press. Vol. 1.

Y. J. Zhu et al., "*Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks*", (2017), Proceedings of the IEEE International Conference on Computer Vision (ICCV),

Y. Mirsky & W. Lee, "*The creation and detection of deepfakes: A survey*" (2021). ACM computing surveys (CSUR), vol 54(1).

Y. Xu, P. Terhöst, M. Pedersen & K. Raja "*Analyzing Fairness in Deepfake Detection with Massively Annotated Databases.*" (marzo 2024). IEEE Transactions on Technology and Society. Vol. 5(1). DOI: 10.1109/TTS.2024.3365421

Young, Scott W. H "*Improving Library User Experience with A/B Testing: Principles and Process*" (Agosto 2014). Weave: Journal of Library User Experience. Vol. 1(1). DOI: 10.3998/weave.12535642.0001.101.

Sitografia

A. Devy "*How Russia's Troll Farm Is Changing Tactics Before The Fall Election*" (29 marzo 2020), New York Times (edizione digitale)

A. Funk, A. Shahbaz & K. Vesteinsson, "*The Repressive Power of Artificial Intelligence*" (2023) freedomhouse.org

A. Genco "*YouTube si prepara a contrastare i deepfake con nuovi strumenti di protezione per i creator*" (6 settembre 2024) HD Blog < www.hdblog.it >

A. K. Przybylski, K. Murayama, C. R., DeHaan, & V. Gladwell, "*Motivational, emotional, and behavioral correlates of fear of missing out*" (2013). Computers in Human Behavior. Vol 29. DOI: 10.1016/j.chb.2013.02.014.

A. Martin "*Chinese law enforcement linked to largest covert influence operation ever discovered*" (29 agosto 2023) The record <therecord.media>

A. Muglia "*Il re dei deepfake indiani: «Sono inondato dalle richieste dei politici»*" (30 aprile 2024) Corriere della sera (edizione digitale)

A. Rennie, J. Protheroe, "*Decoding Decisions: making sense of the messy middle*." (luglio 2020) Think with Google <https://www.thinkwithgoogle.com/>

A. Roosendaal "*Digital Personae and Profiles as Representations of Individuals*", in M. Bezzi et al. "*Privacy and Identity Management for Life*" (2010) Springer Science & Business Media. DOI: 10.1007/978-3-642-20317-6

Aa. vv., "*Intelligence and Security Committee of Parliament - Russia*", (21 luglio 2020), House of Commons

Aa.Vv. "*Deepfake-vademecum*" (28 dicembre 2020) Garante della Privacy.

Accademia della Crusca, sezione parole nuove < accademiadellacrusca.it.>

Adobe Communications Team "*Behavioral targeting - what it is, why it's important, and how to do it*" (15 novembre 2022) Adobe <business.adobe.com>

Autore sconosciuto "*Kiev, in videochiamata con un falso Klitschko: i sindaci di Madrid, Berlino e Vienna ingannati dal deepfake*" (25 giugno 2022) La Repubblica (edizione online)

B. Dasgupta "*BJP's deepfake videos trigger new worry over AI use in political campaigns*" (21 settembre 2020) Hindustan Times, New Delhi (edizione digitale)

C Newman "*Exclusive: Hundreds of British celebrities victims of deepfake porn*" (21 marzo 2024) Channel 4.

C. Carton Ferrer, B. Dolhansky, B. Pflaum, J. Bitton, J. Pan & J Lu, "*Deepfake Detection Challenge Results: An open initiative to advance AI*" (12 giugno 2020), Meta <ai.meta.com>

C. Gentilhomme " *«Une arme pour décrédibiliser un adversaire politique» : alerte contre les «deepfakes», ces vidéos truquées par intelligence artificielle*" (24 dicembre 2023), Le Figaro (edizione digitale)

C. Jeearchive, "*An Indian politician is using deepfake technology to win new voters*", (19 febbraio 2020), MIT technology review.

C. Nilesh "*We've Just Seen the First Use of Deepfakes in an Indian Election Campaign*" (18 febbraio 2020) Vice <www.vice.com>

"*Cosa intendiamo per dati personali?**" Garante della Privacy <garanteprivacy.it >

C. R. Sunstein "*The Ethics of Nudging*" (25 novembre 2021), Yale Jurnal on Reguation. Vol. 32(2), URI: http://hdl.handle.net/20.500.13051/8225,

C. Sernagiotto "*The Mandalorian, Lucasfilm assume lo YouTuber del video deepfake con Luke Skywalker*" (28 luglio 2021) Sky tg 24 <tg24.sky.it>

C.Watts, "*Russian US election interference targets support for Ukraine after slow start*" (17 aprile 2024) Microsoft on The Issues.

"*CyberTipline 2023 Report*" <www.missingkids.org>

"*Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*", (28 dicembre 20) Garante della privacy <www.garanteprivacy.it >

D.Harwell, "*Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*" (3 dicembre 2018) The Washington Post (ed online)

D.O'Sullivan "*Congress to investigate deepfakes as doctored Pelosi video causes stir*" (4 giugno 2019), CNN (edizione digitale)

E. Caffo per telefono azzurro "*dossier abuso 2024*" (novembre 2024) MAG Studio Milano;

E. Caffo per telefono azzurro "*la dignità dei bambini nel mondo digitale*" (5 maggio 2024) MAGcom srl;

E. Harrell "*Neuromarketing: what you need to know*" (23 gennaio 2019) Harvard business review <hbr.org>

E. Kamarck, "*Malevolent soft power, AI, and the threat to democracy*", (29 novembre 2018), brookings.edu

E. Pariser, "*The Filter Bubble: What the Internet Is Hiding from You*" (2012) UCLA journal of education and information studies. Vol. 8(2), DOI: 10.5070/D482011835.

F. Lanza "*La storia della regola 34 di Internet*." (6 novembre 2015) linkiesta.it

F. Michetti "*Deep Nostalgia non è un giochetto, ma la nuova arma della post-verità*" (19 marzo 2021) Agenda digitale <agendadigitale.eu>

F. Vera-Gray, "*Key messages from research on the impacts of child sexual abuse*" (Marzo 2023) Child and Woman Abuse Studies Unit, London Metropolitan University <www.csacentre.org.uk>

"*FBI Launches Sextortion Awareness Campaign in Schools*" (3 settembre 2019) FBI – Federal Investigation Bureau < www.fbi.gov >

G. Asta "*La Russia blocca l'accesso a RaiNews.it e altri 80 media dell'Unione europea*" (25 giugno 2024) RaiNews.it.

G. D'Amico "*Behavioural Marketing e implicazioni privacy: di cosa parliamo e il ruolo delle parti in gioco*" (12 ottobre 2020) Cyber Security 360 < cybersecurity360.it/legal >

G. Giacobini "*Storia dell'app che genera(va) false foto di nudo femminile*" (28 giugno 2019) Wired < www.wired.it/>

G. Sartor & F. Lagioia "*The impact of the General Data Protection regulation (GDPR) on artificial intelligence*", (giugno 2020) EPRS European Parlamentary reserch service. DOI: 10.2861/293

G. Sartor & F. Lagioia "*The impact of the General Data Protection regulation (GDPR) on artificial intelligence*", (giugno 2020) EPRS European Parlamentary reserch service. DOI: 10.2861/293

G. Scorza, componente del Garante per la protezione dei dati personali "*Contrasto alla pornografia non consensuale: istruzioni per l'uso"* (5 dicembre 2022) HuffPost.

Gleason (2016) Imdb.com <https://www.imdb.com/title/tt4632316/>

"*HoloLens 2 x Education*" Microsoft Education. < www.microsoft.com/en-us >

https://deepnostalgia.ai/

https://malariamustdie.com/

Infodata, "*Il funzionamento delle reti neurali convoluzionali spiegate bene in un video*", (11 novembre 2023) il sole 24 ore, edizione digitale < www.infodata.ilsole24ore.com>

"*Informazioni sul fact-checking su Facebook, Instagram e Threads*" centro assistenza per le aziende di meta <www.facebook.com/business/help/ >

J. Peters "*Reddit bans impersonation on its platform / But impersonation that's satire will still be allowed*" (9 gennaio 2020) The Verge <www.theverge.com>

K. Lyons "*New AI 'Deep Nostalgia' brings old photos, including very old ones, to life*" (28 febbraio 2021) The verge <www.theverge.com >

"*La Commissione pubblica orientamenti nell'ambito della legge sui servizi digitali per l'attenuazione dei rischi sistemici online per le elezioni*" (26 marzo 2024)

L. Keolin "*Reddit bans deepfake porn videos*" (8 febbraio 2018) BBC <bbc.com>

L. Saulle "*Che cos'è il Digital Nudging: la persuasione digitale*" (10 dicembre 2020) PXR Italy <pxritaly.com/it >

L. Tremolada "*Dai deepfake di Taylor Swift alle fake news con ChatGpt: come difendersi dalla disinformazione elettorale?*" (21 agosto 2024) Il Sole 24 Ore (edizione digitale)

M Maratona e R. Savella "*AI generativa, la Cina mette i paletti: luci e ombre delle nuove regole*" (10 febbraio 2023) Agenda Digirale <agendadigitale.eu>

M. Bickert "*Our Approach to Labeling AI-Generated Content and Manipulated Media*" (5 aprile 2024) Meta <about.fb.com/news>

M. Bicket "*Enforcing Against Manipulated Media*" (6 gennaio 2020) Meta <about.fb.com>

M. Cazzaniga, "*Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfake, Media Laws*", (14 giugno 2023) Medialaws <Medialaws.eu>

M. Jesse, D. Jannach "*Digital nudging with recommender systems: Survey and future directions*" (13 gennaio 2021) Computer in Human Behavior, report 3, <sciencedirect.com> Doi: 10.1016/j.chbr.2020.100052

M. Martorana & R. Savella "*IA e riconoscimento delle emozioni: rischi e possibili vantaggi*" (21 settembre 2023) Agenda Digitale <www.agendadigitale.eu >

M. Santarelli, "*La Russia Cambia La Strategia Di Disinformazione Politica Negli Usa*", (23 marzo 2021), Agenda Digitale

M. Serafini "*L'hacker che rubava le vite via webcam*" (15 marzo 2012) il Corriere.it, sez. Esteri

N. Arora et alt. "*The value of getting personalization right - or wrong - is multiplying*" (12 Novembre 2021) McKinsey & company <mckinsey.com>

N. Badshah "*Nearly 4,000 celebrities found to be victims of deepfake pornography*" (21 marzo 2024) The guardian (ed digitale).

N. Degli Innocenti "*Gran Bretagna: influenze russe nel voto per la Brexit. Governo Uk sotto accusa*", (19 gennaio 2023) Il Sole 24 Ore, edizione online.

<https://oecd.ai/en/ai-principles >

NETSmart "*I am a survivor of sextortion*" (2024) National center for missing & exploited children <www.missingkids.org >

<https://www.permessonegato.it/>

"*PimEyes: The Digital Detective Redefining Cold Case Investigation*" pimeyes.com

Rappresentanza in Italia <italy.representation.ec.europa.eu>

Redazione Ansa "*Dopo deepfake Zelensky, anche video finto sindaco Kiev*" (28 giugno 2022) Ansa.

R. Natsume, T. Yatagawa & S. Morishima, "*RSGAN: Face Swapping and Editing using Face and Hair Representation in Latent Spaces*" (18 aprile 2018) Arxiv. DOI: 10.1145/3230744.3230818.

R. Riccardi "*Neuromarketing e AI: Strategie per la Mente del Consumatore*" (30 maggio 2023) Università del Marketing < www.universitadelmarketing.it>

R. Rijtano "*Da Nancy Pelosi alla Gioconda mai viste prima: l'era dei deepfake è appena cominciata*" (29 maggio 2019), La Repubblica (edizione digitale)

R. Rogers "*Come non cascare nelle truffe telefoniche AI*" (10 aprile 2024) wired.it <www.wired.it>

R. Towels, "*Alphafold is the most important achivement in AI ever*" (3 ottobre 2021), Frobes (digitale)

"*Russia, censura dell'informazione e persecuzione delle proteste contro la guerra. Crescono opposizione alla guerra e repressione del dissenso*" (28 febbraio 2022) amnesty.it

S. Cole, "*We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*", (24 gennaio 2018) Vice.

S. Ghaffary "*Twitter is finally fighting back against deepfakes and other deceptive media*" (4 febbraio 2020) Vox <Vox.com>

S. Mlot "*TikTok Bans Deepfakes, Expands Fact-Checking to Fight Election Misinformation*" (6 agosto 2020) PC MAG <pcmag.com>

Staff di About amazon "*Amazon introduce nuovi strumenti basati sull'IA generativa per i partner di vendita europei*" (20 giugno 2024) about amazon < www.aboutamazon.it>

"*Taking medical instruction remote and to mixed reality at Case Western*" Microsoft Customer Stories <www.customers.microsoft.com/en-us >

T. Chivers "*Internet rules and laws: the top 10, from Godwin to Poe*" (23 ottobre 2009) The Telegraph (ed digitale)

T. Leopold "*meet the rules of internet*" (15 febbraio 2013) CNN business (ed digitale).

V. Berra "*Zelensky chiede alle truppe di deporre le armi, ma è un falso: il video deepfake rimosso da Facebook e YouTube*" (17 marzo 22) Open.online

V. Colomba e A. Souihel "*Child grooming: come riconoscerlo e come difendersi*" (14 febbraio 2023) Agenda digitale < www.agendadigitale.eu>

V. Giacometti, "*Il deepfake: il sottile confine fra fantasia, gioco e reato*" (13 dicembre 2023) Forensicnews <forensicnews.it>.

V. Mancini & P. Dell'Osso "*Focus Ucraina – La censura del Cremlino sulla guerra*" (6 aprile 2022) irpa.eu

www.gamification.it/

www.myheritage.it

Y. Choi et al., "*StarGAN v2: Diverse Image Synthesis for Multiple Domains*", (26 aprile 2020) arxiv, DOI: 10.48550/arXiv.1912.01865

Y. Mirsky, & W. Lee, "*The Creation and Detection of Deepfakes: A Survey*." (2021). ACM Computing Surveys (CSUR), 54(1), DOI: 10.1145/3425780.

Y. Moncarol Wang "*Don't Believe Your Eyes: Fighting Deepfaked Nonconsensual Pornography with Tort Law*" (2023) University of Chicago Legal Forum University of Chicago Legal Forum; volume 2022 articolo 16

Y. Ruvalcaba & A. Eaton "*Nonconsensual Pornography Among U.S. Adults: A Sexual Scripts Framework on Victimization, Perpetration, and Health Correlates for Women and Men. Psychology of Violence*" (febbraio 2019) DOI: 10.1037/vio0000233

Z. Yang "*I deepfake dei cari defunti sono un business cinese in piena espansione*" (7 maggio 2024) MIT Technology Review <technologyreview.it>

Riferimenti normativi e giurisprudenziali

Aa. Vv., "Ethics guidelines for trustworthy ai" (8 aprile 2019) Commissine europea

aa.vv., "Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence", (2023), UNESCO, DOI: 10.54678/YTSA7796

California ab 2013/24

California Consumer Privacy Act (CCPA)

California Privacy Rights Act (CPRA)

Cass. civ., Sez. I, Ordinanza, 19/01/2023, n. 1674

Cass. pen. Sez. V, Sent., (ud. 05/03/2024) 27/06/2024, n. 25516

Cass. pen. Sez. V, Sent., (ud. 22/02/2023) 07/04/2023, n. 14927

Cass. pen., Sez. V, Sent., (data ud. 01/03/2024) 20/06/2024, n. 24379

Code of Virginia

Convenzione di Istambul (11 maggio 2011) ratificata in Italia con la legge n. 77/2013

D.lgs 196/2003

D.lgs 2 agosto 2007, n. 145

D.lgs 25 marzo 2024, n. 50, "Disposizioni integrative e correttive del decreto legislativo 8 novembre 2021, n. 208",

D.lgs 8 novembre 2021, n. 177, recante "Attuazione della direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale".

D.lgs 9 aprile 2003, n. 70 "Attuazione della direttiva 2000/31/CE relativa a taluni aspetti giuridici dei servizi della società dell'informazione nel mercato interno, con particolare riferimento al commercio elettronico"

Direttiva (UE) 2018/1808

Direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale (c.d. Direttiva Copyright)

Direttiva 2005/29/CE

Direttiva Del Parlamento Europeo E Del Consiglio COM/2022/105 final

European Parliamentary Research Service < www.europarl.europa.eu >

GU n.90 del 17-04-2024

L. 6 febbraio 2006, n. 38 "Disposizioni in materia di lotta contro lo sfruttamento sessuale dei bambini e la pedopornografia anche a mezzo Internet."

Legge 22 aprile 1941, n. 633 (Legge sul diritto d'autore).

Legge 31 dicembre 1996, n. 675, Codice Della Privacy

Online Safety Act 2023, UK Government Bill, Originated in the House of Commons, Sessions 2021-22, 2022-23

Parere del Comitato economico e sociale europeo 2017/C 288/01, 31 maggio 2017.

Regolamento (UE) 2016/679

Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio

Regolamento (UE) 2022/1925 del parlamento europeo e del consiglio del 14 settembre 2022

Regolamento (UE) 2022/2065 del parlamento europeo e del consiglio del 19 ottobre 2022

Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024Regolamento 2022/2065

S.3572 - Algorithmic Accountability Act of 2022, 117th congress, USA

Texas Penal code