UNIVERSITÁ DEGLI STUDI DI PADOVA

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata

Corso di laurea triennale in Comunicazione

L'ARMA DELL'INGANNO: ANALISI DEGLI EFFETTI DEL DEEPFAKE NEL MONDO DIGITALE

Relatore:

Bruno Mastroianni

Laureanda:
Martina Lupo

Numero di matricola:
1224153

Anno accademico

2021/2022

Sommario

Abstract

I nuovi sviluppi digitali stanno creando la possibilità di rendere sempre più indistinguibili tra di loro i contenuti falsi da quelli reali. La nuova tecnologia che sta contribuendo ad alimentare questo problema è il deepfake, che genera video sfruttando l'Intelligenza Artificiale (AI) a partire da algoritmi di *deep learning* e reti neurali, generando così una nuova forma di manipolazione dei contenuti video. L'uso dannoso di questa tecnologia ha creato più pericoli che opportunità soprattutto per i cittadini. Questi video, infatti, sono spesso usati per danneggiare la reputazione delle celebrità e guidare l'opinione pubblica, minacciando gravemente la stabilità e la fiducia sociale. Sebbene il deepfake possa essere utilizzato per scopi positivi, come la realizzazione di film e la realtà virtuale, è ancora ampiamente applicato per usi dannosi e scopi negativi. Partendo dalla considerazione di alcuni studi che hanno approfondito questo campo, cercheremo di percorrere ed esplorare gli effetti che questa nuova tecnologia genera nel mondo digitale di oggi.

Introduzione

Questa tesi si pone come obbiettivo quello di affrontare il problema sempre più diffuso dei deepfake per esplorare le sue conseguenze ed effetti sulla società. Nel primo capitolo si partirà dal tema della disinformazione che è strettamente connesso all'uso dei deepfake. Successivamente verranno descritte in modo approfondito le tecniche utilizzate per la creazione dei video deepfake, dal *face swapping* ai sintetizzatori audio, applicate attraverso delle reti neurali denominate GAN. In seguito si cercherà di capire da chi sono prodotti e quali sono i motivi che stanno dietro la loro creazione.

Nel secondo capitolo verranno inizialmente descritte le conseguenze dannose che il deepfake porta, dalla disinformazione al *deepnude*, per poi passare a una descrizione degli aspetti utili veicolati da questo fenomeno, che vanno dal campo commerciale a quello didattico. Verranno presentati, inoltre, alcuni esempi che mostrano le possibilità costruttive e distruttive di questa tecnologia. L'ultimo paragrafo descriverà poi un particolare effetto del deepfake che viene impiegato nelle immagini di persone care decedute per poterle animare. Si esploreranno le problematiche etiche legate a questa tecnica denominata Deep nostalgia.

Il terzo capitolo riporta i risultati di un questionario sottoposto a un campione di 160 persone per comprendere la difficoltà dei soggetti coinvolti nel riconoscere i video veri da quelli falsi. Ai partecipanti sono stati somministrati sette video, due reali e cinque fake, nei quali dovevano selezionare l'opzione corretta. Le risposte hanno permesso di avere un quadro articolato delle difficoltà riscontrate dai rispondenti e sulla gravità del non poter più credere non solo a ciò che si legge, ma anche a ciò che si vede con i propri occhi.

L'ultimo capitolo, infine, cercherà di indagare le misure di risposta ai danni dei deepfake messi in campo dalla legislazione e dalle piattaforme digitali. Si tenterà, inoltre,

di capire se siano sufficientemente efficienti o se c'è ancora strada da fare prima di arrivare a non riconoscere più il vero dal falso.

La scelta dell'argomento risale a sei mesi fa, in cui durante una lezione di teoria e pratica dell'argomentazione digitale ho sentito il professor Bruno Mastroianni, nonché relatore di questa tesi, nominare il termine deepfake. Non conoscendone il significato ed essendone incuriosita sono andata a ricercare la definizione e le sue funzionalità, rimanendone colpita. Ho deciso quindi di unire quelli che fin dall'inizio del mio percorso universitario sono stati gli argomenti che mi hanno coinvolto più appassionatamente (come l'argomentazione, lo sviluppo tecnologico, i mezzi di informazione e le fake news) con la curiosità e il trasporto che questo fenomeno ha suscitato in me.

Questa tesi, infine, oltre a descrivere le potenzialità e le funzionalità del deepfake, cercherà di esporre un pensiero personale a partire dall'applicazione del pensiero critico di diversi autori che hanno studiato e giudicato il fenomeno in profondità, in modo tale da comprendere come questo si applichi ai giorni nostri.

CAPITOLO 1
INTRODUZIONE AL DEEPFAKE

1.1 Fake News e "Behavioral Economics"

"Mentre una bugia sta facendo il giro del mondo, la verità si sta ancora allacciando le scarpe" affermava Mark Twain (Cosentino,2017). Questo aforisma può descrivere perfettamente un fenomeno che è sempre più in crescita: le fake news. Nonostante sia datato, ben noto ancor prima dell'avvento di Internet, con il tempo questo termine è diventato sempre più presente nell'ambito pubblico, politico, sociale e dei media, soprattutto a causa della crescita delle più recenti tecnologie dell'informazione e della comunicazione, dell'utilizzo massiccio dei social network e degli eventi che, recentemente, hanno pervaso la nostra quotidianità. L'effetto principale delle fake news e della disinformazione in generale è quello di intercettare un malcontento sociale pubblico riguardo a tematiche di attualità in modo tale da far leva su di esso e causare indignazione all'interno della società.

Il principale pericolo è che mirando a questo tipo di scopo, rischiano di diventare un vero e proprio problema per la salute sociale e pubblica, in quanto hanno l'obiettivo di influenzare e orientare il pensiero della gente comune per alimentare teorie complottiste o contro il sistema. La causa di tutto ciò, per giunta, non è solo la disinformazione, ovvero la deliberata creazione e condivisione di notizie false, ma anche la misinformazione, ovvero la creazione e divulgazione involontaria di esse.

Il "*World Economic Forum*", a questo proposito, ha introdotto nel "*Global Risk report*" del 2013 il problema di un rischio globale legato alla misinformazione digitale, posizionandolo al centro dei pericoli legati alle nuove tecnologie (Quattrociocchi W., Vicini A. 2016). Infatti, come affermano Quattrociocchi e Vicini, viviamo in una realtà

in cui non vi è più solo la figura dell'*opinion leader* classico, del giornalista, dell'esperto, perché tutti noi possiamo diventare possibili emittenti e abbiamo la possibilità di creare, di condividere contenuti e di divulgare la nostra visione sulla realtà "*worldwide*" senza nessun tipo di controllo. Inoltre affermano che in questo scenario l'importanza della qualità delle informazioni, su cui viene prodotta la visione di un determinato fenomeno, rischia di risultare fortemente compromessa.

Per comprendere i processi psicologici e comportamentali che stanno dietro la diffusione delle fake news, vorrei collegarli ad alcuni concetti della *Behavorial Economics* un "settore interdisciplinare dell'economia e della psicologia cognitiva, che studia il comportamento decisionale dal punto di vista economico con metodo sperimentale"[1]. Gli studi della *Behavioral Economics* mostrano che le persone compiono le loro scelte spesso in modo irrazionale.

Spesso si ipotizza che gli attori razionali prendano decisioni deliberate, ma la ricerca psicologica suggerisce che non è sempre così. Infatti, particolarmente interessante in questo campo è stato il lavoro svolto dallo psicologo israeliano Daniel Kahneman, vincitore del Premio Nobel per l'economia nel 2002 e autore del celebre saggio "*Thinking, Fast and Slow*" del 2011, tradotto da Laura Serra nel 2012 in "Pensieri lenti e veloci" (Kahneman, 2012). Kahneman iniziò a lavorare per le Forze Armate Israeliane, dove si occupava di valutare l'idoneità dei candidati. In questo periodo, però, notò di aver commesso degli errori di giudizio nella loro selezione e per questo motivo si interessò ai meccanismi che stanno alla base del comportamento decisionale umano. Nel 1968 Kahneman iniziò a lavorare assieme ad Amos Tversky, anche lui psicologo israeliano, per capire come si sviluppano i giudizi e i processi decisionali. Kahneman presuppone che, metaforicamente, la mente umana sia dotata di due sistemi: il Sistema 1, che produce un pensiero veloce, istintivo, intuitivo e funziona continuamente, fornendoci involontariamente delle conclusioni senza che noi ci pensiamo in modo consapevole, e il Sistema 2 che rende, invece, un pensiero più impegnativo, analitico, cosciente e che richiede più tempo per arrivare a una conclusione (Kahneman, 2012).

Applicando questi criteri di Kahneman alla disinformazione potremmo ipotizzare che quando un individuo si trova di fronte a una notizia falsa, che è creata per attirare

[1] https://www.treccani.it/enciclopedia/behavioral-economics/#:~:text=Branca%20interdisciplinare%20dell'economia%20e,scelta%20economica%20con%20metodo%20sperimentale.

intenzionalmente l'attenzione in modo istintivo, l'individuo reagisce attivando solo il Sistema 1, che è quello utilizzato principalmente per prendere decisioni in modo veloce, senza un grande sforzo cognitivo. Invece, quando ci si trova di fronte a una fake news, per non cadere nell'imbroglio, bisognerebbe attivare il Sistema 2, che richiede uno sforzo mentale più impegnativo e che comporterebbe, quindi, un'analisi più laboriosa in modo tale da esaminare più accuratamente le fonti, la verificabilità di ciò che viene affermato e l'affidabilità di un autore.

Quindi, a meno che non si sia motivati a fare uno sforzo cognitivo per impiegare il Sistema 2, si accetta semplicemente la conclusione del Sistema 1 con poca riflessione.

Questo significa che seppur il Sistema 1 sia fondamentale per prendere decisioni in modo veloce e intuitivo nella nostra quotidianità, queste ci influenzano anche di fronte a questioni che richiedono un maggiore sforzo cognitivo e razionale. Per questo motivo, il risultato imminente non può essere che la produzione di quelli che Kahneman chiama "bias cognitivi", che l'autore definisce come "preconcetti che ricorrono in maniera prevedibile in particolari circostanze" (Kahneman, 2012). In altre parole, i bias sono distorsioni delle valutazioni delle persone su fatti ed eventi, che ci portano a ricreare un nostro scenario soggettivo che non corrisponde esattamente alla realtà, portandoci così a distorcere ciò che ci circonda. In pratica, il Sistema 1 porta a prendere decisioni in modo veloce, ma nella maggior parte dei casi fondate su pregiudizi e percezioni che non hanno la garanzia di non essere errate o distorte.

1.2 Fake News e video: l'origine del deepfake

Nel mondo digitale in cui oggi viviamo le fake news e i pericoli ad esse associati si sono moltiplicati a causa dell'avanzamento tecnologico. Infatti, ad oggi, è sempre più difficile distinguere le fonti affidabili da quelle false e le notizie inattendibili si diffondono molto velocemente attraverso i social, mediante i quali possono avere un impatto su milioni di utenti.

Nel 2016 l'*Oxford Dictionary* ha selezionato il termine *post-truth*, ovvero post-verità, come parola dell'anno, definendola come "relativa circostanza in cui i fatti oggettivi sono meno influenti nel plasmare l'opinione pubblica rispetto agli appelli alle emozioni e alle convinzioni personali" (Oxford dictionary, 2016). L'aggettivo è stato

coniato per definire un nuovo tipo di politica, ovvero la "politica della post-verità". Infatti, come afferma Gabriele Cosentino sembrerebbe che il 2016 sia stato l'anno che ha portato all' "inizio di una nuova fase politica e la rottura con equilibri precedenti, sia per l'imprevedibilità di eventi che hanno sorpreso anche i più attenti osservatori, sia per la loro portata epocale" (Cosentino, 2017). L'autore afferma che durante questo periodo di cambiamenti ha preso vita un nuovo rapporto tra politici e media e la causa di tutto ciò sembrerebbe la perdita di legittimità delle fonti ufficiali, dei media tradizionali e di una nuova comunicazione basata sui social media, che sono sempre più sfruttati in ambito politico.

Se fino a qualche anno fa erano soprattutto le immagini, le foto e gli elementi testuali a prevalere nei social come Instagram o Facebook, che nascono inizialmente come social network dedicati alla fotografia, oggi con l'avvento di un social come Tik Tok[2], i video brevi stanno prendendo il sopravvento. Tik Tok è un social network nato nel 2016 dalla fusione con il precedente Musical.ly, che permette ai suoi utenti di registrare e condividere video creativi di durata variabile sulla piattaforma, permettendo anche di modificare la velocità di riproduzione, di aggiungere filtri, audio, doppiaggi e alti effetti particolari che consentono di editare il video. Questa rapida diffusione dei video brevi, però, ha anche permesso l'emergere di una nuova tecnica che sfrutta l'intelligenza artificiale per sovrapporre il volto di una persona ad un'altra ripresa in un video: il deepfake. Nel 2020 il Garante della protezione dei dati personali ha definito i deepfake come:

> "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" (GPDP, 2020).

La tecnologia deepfake può generare, ad esempio, un video umoristico, politico o addirittura pornografico senza il consenso della persona presa in causa, e il fattore più sorprendente, quanto inquietante, è che la tecnologia che viene impiegata è talmente sofisticata da rendere questi video indistinguibili rispetto ai contenuti autentici.

[2] https://www.tiktok.com

Il termine deepfake secondo l'enciclopedia Treccani deriva "dall'espressione inglese deepfake, che incrocia la locuzione *deep learning* ('insieme di tecniche che permettono all'Intelligenza artificiale di imparare a riconoscere le forme') con *fake* ('falso, notizia falsa')" (Treccani, 2018).

Il primo tentativo di scambio di un volto risale al 1865, quando la testa del presidente degli Stati Uniti Abraham Lincoln è stata trasposta nel corpo del politico sudista John Calhoun. Dopo l'assassinio di Lincoln, la richiesta di litografie che lo ritraevano era talmente tanta che iniziarono ad apparire incisioni della sua testa su altri corpi (Güera D., Delp E. J., 2018). Con il tempo però, e con i recenti progressi è stato cambiato radicalmente il campo di manipolazione di immagini e video. Come afferma Filippo Torrini nel blog di divulgazione "*UniverseIT*"[3], i video deepfake nascono come tecnica di ausilio agli effetti speciali cinematografici. Infatti, spesso il cinema di Hollywood ha utilizzato questa tecnologia per sostituire un volto reale o immaginario su altri attori (basti pensare a film campioni di incassi come *Avatar*), o addirittura per riportarli in vita, com'è accaduto per esempio con la pellicola di *Star Wars Story* nel 2016.

L'azienda ha dovuto usufruire della la CGI (*computer generated imagery*) e filmati d'archivio modificati per ricreare le fattezze di Peter Cushing, interprete del personaggio di Tarkin in "Una nuova speranza", sul corpo dell'attore Guy Henry (Scarselli L., 2020). Infatti l'attore Peter Cushing scomparve nel 1994 e, in previsione dell'uscita nel 2016 di "*Rogue One: A Star Wars Story"*, ambientato prima del film originale della famosa saga di Guerre Stellari, l'azienda decise di recuperare numerose ore di filmati di Cushing per usarli come modello e per permettere così all'attore Guy Henry di interpretare il personaggio di Tarkin sul set attraverso la sostituzione del suo volto con una maschera digitale creata appositamente per l'attore (Scarselli L., 2020).

Il termine "deepfake" è stato coniato però solo nel 2017, a causa della diffusione di video pornografici fake che vedevano come protagonisti personaggi celebri la cui identità era stata rubata senza il loro consenso, pubblicati da un utente chiamato "*deepfakes*" su Reddit (Güera D. e Delp E., 2018). Il problema, però, è che nonostante la circolazione di questi video sia stata bloccata e sia stata proibita la diffusione di porno fake, la divulgazione dei video da quel momento ha iniziato a diventare inarrestabile anche attraverso altre applicazioni come, per esempio, *FaceApp*, che genera

[3] https://universeit.blog/deepfake/

automaticamente trasformazioni altamente realistiche dei volti nelle fotografie, e *FakeApp*, un'applicazione desktop in grado di generare i video deepfake.

1.3 Dal face swapping all'audio deepfake

Per capire come funzionano attualmente i video e le tecniche deepfake, è importante innanzitutto introdurre due concetti fondamentali: il *deep learning*, che abbiamo già menzionato in precedenza, e le reti neurali. I deepfake, infatti, si basano su reti neurali introdotte nel 2014 da un lavoro intitolato "*Generative adversarial networks*"[4] dell'informatico e ricercatore statunitense Ian Goodfellow (2020) e altri ricercatori dell'Università di Montreal, denominate GAN (*Generative Adversial Network*), che l'autore definisce come un tipo di algoritmo di intelligenza artificiale progettato per risolvere il problema del modello generativo. L'obiettivo del modello generativo, secondo l'autore, è quello di studiare un insieme di esempi di addestramento e imparare la probabilità di distribuzione che li ha generati. Le GAN sono poi in grado di generare altri esempi a partire dalla distribuzione di probabilità stimata.

Goodfellow afferma, inoltre, che i modelli generativi basati sul *deep learning* sono comuni, ma le GAN sono tra i modelli generativi di maggior successo, soprattutto in termini di capacità di generare immagini realistiche ad alta risoluzione. Il processo prevede l'inserimento dei filmati di due persone in un algoritmo di *deep learning* che è in grado di scambiare i volti attraverso la tecnica del *face swapping*. In altre parole, i deepfake utilizzano la tecnologia di mappatura facciale e l'intelligenza artificiale che sono in grado di scambiare in un video il volto di una persona con quello di un'altra in modo estremamente realistico. Per spiegare il *face swapping*, è necessario introdurre quattro tipi di manipolazioni, che sono stati riportati nel 2022 nel blog di divulgazione "*UniverseIt*" dall'autore Filippo Torrini[5]: la sintesi dell'intero volto, il cambio di identità, la manipolazione di un attributo e il cambio di espressione.

Nel primo caso, la sintesi dell'intero volto dà origine a volti umani perfettamente identici a quelli di persone reali, ma che in realtà sono del tutto generati dal computer attraverso la rete generativa contradditoria (GAN), che in termini tecnici è

[4] "Reti generative contradditorie"
[5] https://universeit.blog/deepfake/

sostanzialmente un'architettura composta da due reti neurali in competizione l'una con l'altra - dove la prima è definita "generatore" ed è, infatti, quella che genera nuovi dati, che passano poi alla seconda rete, chiamata "discriminatore", che distingue i contenuti creati artificialmente da quelli reali. Il compito del generatore è quindi quello di trarre in inganno il discriminatore e, sfruttando le risposte da esso fornite, è programmato per migliorarsi in modo da creare dati sempre più verosimili (Iozzia G., 2022).

Il cambio di identità, invece, si ha quando in un video si sostituisce il volto di una persona con quello di un'altra, senza permettere di poter riconoscere la manipolazione, come accaduto con i video pornografici fake divulgati su Reddit dov'è stata coinvolta l'identità di alcune celebrità, come abbiamo visto.

La manipolazione di un attributo, dall'altro lato, consente di cambiare alcuni aspetti e caratteristiche del volto umano (colore della pelle, dei capelli, genere ed età), come succede utilizzando *FaceApp*, che è un'applicazione che permette di elaborare le foto e di appore dei filtri ai volti modificandoli (per esempio invecchiando il viso di una persona).

Infine, il cambio di espressione consente, ad esempio, di rendere un viso triste sorridente o un viso alterato rasserenato. Questo tipo di *face swapping* può avere conseguenze estremamente dannose dal momento in cui diventa possibile creare da zero i movimenti labiali che riguardano un discorso e trasporli sul volto di un'altra persona.

Per di più, l'innovazione più grande di queste tecnologie è che non richiedono una grande abilità da parte dell'utente per creare questo tipo di contenuti. Infatti, non è necessario modificare manualmente le luci, i colori, i pixel e fare operazioni complesse per rendere l'immagine il più verosimile possibile come si farebbe invece per modificare un'immagine su Photoshop. Questo perché la rete neurale è in grado di apprendere in modo automatico come agire a partire dai dati disponibili. Per questo i video deepfake possono avere delle conseguenze estremamente pericolose: hanno fatto sì che chiunque potesse creare questo tipo di manipolazioni che prima richiedevano un certo livello di competenza e, inoltre, il problema più grande è che a oggi chiunque abbia un computer può creare video falsi che sono sempre più indistinguibili da quelli autentici.

Per quanto riguarda l'audio, invece, la creazione di un audio falso che imita la voce del soggetto coinvolto oggi è reso possibile in maniera rapida ed economica anche grazie ad aziende come *Lyrebird*, che ha sviluppato un'applicazione chiamata con lo

stesso nome dell'azienda in grado di imitare in modo estremamente accurato la voce di qualsiasi persona. La tecnologia della *Lyrebird* e quella sviluppata all'Università di Washington (UW), "permette quindi di sincronizzare il movimento labiale con un audio e di elaborare un video in cui il soggetto "parla" con un audio rielaborato e mai veramente pronunciato" (Arruzzoli F., 2019).

L'audio sintetizzato viene prodotto attraverso i cosiddetti *synthetizer*, dispositivi elettronici che sono in grado di generare autonomamente segnali audio, molto utilizzati nell'ambito della musica pop. Diverse aziende hanno realizzato dei sintetizzatori audio in grado di riprodurre una voce dopo averla sentita. Per esempio, nel documentario sulla vita di Anthony Bourdain "*Roadrunner: A Film About Anthony Bourdain*"[6], gastronomo e personaggio televisivo statunitense morto suicida nel 2018, è stato creato un suo clone vocale da inserire nel documentario. Il regista Morgan Ville ha affermato:

> "C'erano tre battute che avrei voluto sentirgli pronunciare con la sua voce, ma non esistevano registrazioni, quindi mi sono messo in contatto con una società di intelligenza artificiale, e gli ho fornito ore e ore di girato perché me la producessero" (Zanon Giusto M., 2021).

Basta registrare delle frasi standard e poi gli algoritmi riproducono un modello molto accurato della nostra voce, il che può essere utilizzato per farci dire qualsiasi cosa. Questo tipo di sistema è chiamato *Voice cloning*, un ramo del *Digital cloning*, che è una tecnologia emergente che permette di simulare artificialmente la voce di una persona e di manipolare foto e video (Grandis N., 2022).

Gli esordi di questo metodo nascono nel 1970 a partire dallo studioso giapponese di robotica Masahiro Mori che pubblicò sulla rivista "*Energy*" un celebre articolo dal titolo "*Uncanny Valley*" (Mori M., 2012), uno studio sul rapporto tra verosimiglianza antropomorfica e accettazione emozionale che le persone provano verso i robot. Mori riportò su un grafico cartesiano la nostra inclinazione ad accettare volentieri istruzioni da un robot molto simile a una persona, mentre siamo meno inclini quando esse provengono da un robot verso il quale percepiamo meno familiarità.

[6] "Roadrunner: un film su Anthony Bourdain", https://www.youtube.com/watch?v=ihEEjwRlghQ

L’ipotesi di fondo si basa su un concetto semplice: tanto più un robot assume fattezze umane, tanto più alto sarà il grado di affinità emotiva che un individuo prova nei suoi confronti. Quindi è importante considerare due fattori fondamentali: il livello di somiglianza che l’oggetto ha rispetto all’essere umano e la percezione di familiarità che lo stesso evoca per l’individuo (Cherchi F., 2022).

Consideriamo per esempio i robot industriali: nonostante svolgano movimenti ed azioni simili a quelle umane, come per esempio un braccio che si piega o una pinza che afferra un oggetto, essi hanno un basso grado di somiglianza con gli esseri umani e quindi registreranno uno scarso livello di affinità nei partecipanti.

È diverso però, come sostiene Mori, se consideriamo, per esempio, un robot giocattolo: si tratta di un oggetto che non solo imita le nostre azioni funzionali, ma riproduce alcune fattezze degli esseri umani come le gambe, la faccia, le braccia, le mani (Mori M., 2012). In questo caso il livello di affinità cresce, ed infatti sono dei tipi di automi che ritroviamo spesso nel mercato dedicato ai bambini e allo svago. Finora, quindi, se dovessimo immaginare un asse cartesiano che presenta una curva, saremmo nella parte crescente e le cose rispetterebbero un andamento lineare. Tanto più aumenta la somiglianza del robot a un essere umano, tanto più cresce il grado di affinità e familiarità percepita da degli ipotetici partecipanti nei suoi confronti.

Mori, però, a questo punto evidenzia un problema: secondo i suoi studi, all’aumentare del grado di somiglianza umana del robot esiste un punto oltre il quale il grado di affinità percepita presenta una considerevole riduzione. Questa riduzione può essere causata da una reazione emotiva negativa nei confronti di un oggetto inanimato che, nel suo tentativo di sembrare umano, fallisce.

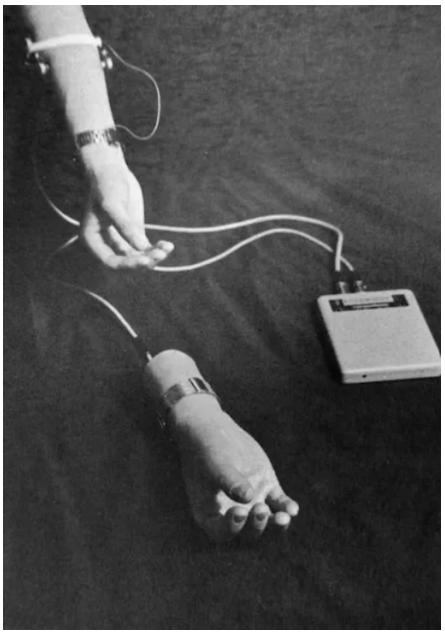

Figura 1 Braccio umano che controlla mano mioelettrica, chiamata mano di Vienna

Immaginiamo una protesi robotica a forma di mano, caratterizzata da una pelle sintetica che imiti bene l'epidermide e unghie di plastica: essa riprodurrà in modo fedele sia le azioni funzionali sia le fattezze umane, solo che adesso il grado di somiglianza umana cresce maggiormente rispetto ai giocattoli robot (Cherchi F., 2022). In questo caso, una volta che gli ipotetici partecipanti si rendono conto che ciò che a prima vista sembrava reale in realtà è artificiale, sperimenteranno delle emozioni negative, inquietanti e, riportando le parole del professor Mori, "*we lose our sense of affinity, and the hand becomes uncanny*"[7] (Mori M., 2012). La sensazione *uncanny* (perturbante) di stringere una mano meccanica come se fosse vera, e la freddezza ad essa associata, fa cambiare radicalmente la percezione di affinità: adesso la curva decresce, fino a toccare il fondo di quello Mori definisce la *Uncanny Valley*, ritrovandoci improvvisamente nel punto più basso della curva.

Mori prende in considerazione un altro oggetto, ovvero una bambola *Bunraku* (una bambola tradizionale usata negli spettacoli teatrali giapponesi). La bambola, in questo caso, riproduce azioni umane funzionali (si muove e agisce come un umano), assomiglia concretamente a una persona (capelli, occhi, espressioni facciali), e presenta una maggiore somiglianza con l'essere umano rispetto alla mano meccanica menzionata nel paragrafo precedente (infatti adesso l'oggetto viene presentato nella sua interezza), anche se in termini di dimensioni e di realismo appare effettivamente diversa da una persona. Nonostante ciò, il livello di familiarità percepito in questo caso cresce, fino al punto tale da uscire fuori dalla profondità della *Uncanny Valley* (Mori M., 2012).

In altre parole, assistere a uno spettacolo teatrale di bambole giapponesi non suscita nello spettatore quelle emozioni negative che percepisce invece nella mano meccanica; l'oggetto in questione acquisisce un grado di somiglianza umana tale per cui il suo "tentativo" di replicare le fattezze di un essere umano è soddisfacente. Torniamo, quindi, sulla curva ascendente: abbiamo lasciato la *Uncanny Valley* e ritroviamo

[7] "perdiamo il nostro senso di affinità e la mano diventa misteriosa"

un'associazione positiva tra somiglianza umana e familiarità percepita. L'ultimo step, infatti, vede un oggetto robotico con fattezze umane indistinguibili ed estremamente verosimili che si comporta, si muove, ed appare esattamente come un essere umano "sano". Tale robot registrerebbe un punteggio alto per entrambe le variabili in quanto sarebbe estremamente simile all' essere umano e avrebbe alti livelli di familiarità percepita (Cherchi F., 2022).

Il fenomeno della mano meccanica, che troviamo nel punto più basso della *Uncanny Valley*, è utile a descrivere come la modalità di interazione tra uomo e robot sia molto complessa: il problema deriva dal fatto che quando un oggetto tenta di sembrare umano, ma fallisce, evoca un profondo senso di non familiarità, di emozioni negative, che possono alterare il rapporto tra oggetto e individuo. Per questo motivo Mori rivolge un messaggio ai designer: creare robot con fattezze umane è una disciplina che deve tenere ben presente la *Uncanny Valley* (Mori M., 2012).

La sfida è quella di creare un livello stabile ed equilibrato, dove un oggetto robotico, sebbene simile all'umano, non evochi emozioni negative nella persona che ne usufruisce. Una possibile soluzione è quella di creare oggetti aventi una fattezza volontariamente non umana. Cherchi a questo proposito propone di considerare per esempio dei semplici occhiali da vista: se questi avessero fattezze umane, per esempio un design simile all'occhio umano all'interno delle lenti, si ritroverebbero nella *Uncanny Valley* (Cherchi F., 2022). Questo non accade perché gli occhiali semplicemente evidenziano degli aspetti estetici, senza diventare umani. Lo stesso vale nell'esempio della mano meccanica: tentare di ricreare un oggetto troppo umano che, riproduce le unghie, l'epidermide e la colorazione della pelle, porta al fallimento di quel tentativo e il rischio di ritrovarsi nella *Uncanny Valley*. Ciò non accade, invece, per le mani di legno delle statue buddiste che riporta Mori nei suoi studi perché nonostante siano dotate di articolazioni e unghie, non determinano alcuna emozione negativa, proprio per il fatto che rinunciano consapevolmente a ricreare le fattezze umane in quanto fatte di legno (Mori M., 2012)

Secondo la spiegazione di Mori, la sensazione di negatività che proviamo di fronte a una mano meccanica sembrerebbe una sorta di istinto di protezione verso una fonte di pericolo: la stessa sensazione, infatti, viene riscontrata anche per i cadaveri (oggetti non mobili), in quanto entrambi vengono associati ad una sensazione di freddezza, mancanza

di vita organica e morte. Infatti, se nella profondità della *uncanny valley* troviamo questi due oggetti, dall'altra parte, troviamo un oggetto che è stato in grado di assumere alla perfezione le fattezze di una "persona sana", dotata di vita, che non ha bisogno dell'attivazione di alcun meccanismo di auto-preservazione (Mori M., 2012).

Gli studi di Masahiro Mori possono essere molto interessanti se applicati al fenomeno del deepfake. La domanda da porsi è: questa tecnologia emergente permette effettivamente che la versione reale e la sua copia digitale siano assolutamente indistinguibili? E se non lo fossero, quale sarebbe la reazione di un'eventuale osservatore nel vedere una copia umana simile, ma non perfettamente e dettagliatamente realistica?

Se proprio vogliamo trovare una tecnica che possa in qualche modo farci capire che il video che stiamo osservando non è reale, è quello di ascoltare le nostre percezioni. Può accadere che, trovandoci di fronte a un video, proviamo delle sensazioni di disagio derivanti dal fatto che il video ha lo scopo di farci credere che sia reale, quando in realtà è ingannevole perché generato attraverso l'intelligenza artificiale. I deepfake sono, infatti, video iperrealistici, e questo, se riprendiamo i risultati di Mori, può portarci a percepire delle sensazioni negative esattamente come l'esempio della mano meccanica. Questo può accadere quando vediamo quelle pubblicità in televisione che fanno sembrare il cibo molto più appetitoso di quello che è in realtà, o fanno apparire un oggetto che in realtà non è presente nel set, infatti l'industria pubblicitaria sta utilizzando sempre più spesso tecniche che sfruttano l'intelligenza artificiale. In altre parole, non siamo allarmati quando sappiamo che qualcosa è una finzione ovvia o una rappresentazione irreale e, allo stesso modo, non lo siamo neanche quando siamo sicuri che ciò che stiamo vedendo è reale. Ma c'è un punto intermedio in cui un essere umano si sente a disagio, ovvero quando percepiamo che qualcosa sta cercando di ingannarci. Ad esempio, quando ci rendiamo conto che un video, essendo stato creato attraverso l'intelligenza artificiale, è estremamente nitido, perfetto in ogni minimo dettaglio, che presenta una figura talmente impeccabile da non sembrare "umana", "reale", esattamente come l'esempio della mano meccanica, che evoca immediatamente in noi un senso di freddezza ed estraneità.

1.4 Chi li produce e perché?

Esistono almeno quattro tipi di produttori di video deepfake: comunità di appassionati di deepfake, attori politici come governi stranieri e vari attivisti, altri individui ostili come truffatori e, infine, autori legittimi, come le società televisive (Westerlund M., 2019)

Dopo l'introduzione di video deepfake pornografici di celebrità condivisi su Reddit da parte dell'utente "*deepfakes*" nel 2017, non ci è voluto molto prima che un'altra comunità di hobbisti deepfake raggiungesse 90.000 membri. Westerlund afferma che, in generale, molti utenti tendono a vedere i video falsi realizzati attraverso l'intelligenza artificiale come una nuova forma di umorismo online piuttosto che come un modo per ingannare o minacciare le persone (Westerlund M., 2019). Questi deepfake sono pensati per essere divertenti o satirici e la loro condivisione può aiutare a guadagnare seguaci sui social media. Alcuni hobbisti, invece, potrebbero essere alla ricerca di vantaggi più concreti, come aumentare la consapevolezza sul potenziale della tecnologia deepfake oppure pubblicare video falsi per mostrare di essere esperti nel campo, con lo scopo di ottenere un lavoro correlato a questo ambito, ad esempio attraverso la produzione di video musicali o programmi televisivi.

Mentre i deepfake possono intrattenere e divertire gli utenti online, in questo campo sono coinvolti anche individui che procedono secondo scopi intenzionalmente più dannosi. Vari attori sociali, come agitatori politici, attivisti informatici e terroristi possono utilizzare il deepfake a scopo disinformativo per manipolare e influenzare l'opinione pubblica e per ridurre la fiducia nelle istituzioni di un determinato paese. I deepfake, infatti, stanno diventando sempre di più armi di disinformazione volte a interferire con le elezioni e generare disordini civili. Potremmo prevedere un aumento sempre maggiore di disinformazione supportata dagli stessi Stati che utilizzano l'intelligenza artificiale per diffondere video falsi politici che hanno lo scopo di influenzare gli utenti dei social media. I deepfake sono anche sempre più utilizzati dai truffatori sul campo economico, al fine di manipolare il mercato e le azioni o di commettere altri reati finanziari. Alcuni, infatti, hanno già utilizzato audio falsi generati attraverso l'intelligenza artificiale per impersonare un dirigente al telefono chiedendo un trasferimento di denaro e, per di più, i

materiali necessari per produrre questo tipo imitazioni da parte di falsi dirigenti sono spesso disponibili su Youtube (Westerlund M., 2019).

Per concludere, in questo capitolo ci siamo soffermati sulle peculiarità più importanti che caratterizzano il deepfake a partire da alcuni studi che hanno approfondito questa tecnologia emergente. Abbiamo esplorato le sue origini, le sue caratteristiche tecniche e le sue implicazioni fino ad arrivare ai motivi per cui questa innovazione viene complessivamente utilizzata. La domanda da porsi ora è: le ragioni che spingono alla creazione di video deepfake con scopi funzionali e positivi bastano per giustificare l'esistenza di una tecnologia che viene sfruttata a scopo dannoso e deleterio?

CAPITOLO 2
DEEPFAKE: PERICOLO DI UNA FALSA REALTÁ O UTILE FINZIONE?

2.1 Disinformazione e “infodemia”

È il 2019, siamo a un evento del Centro per il progresso americano e la Speaker della Camera Nancy Pelosi viene ripresa dalle telecamere mentre pronuncia delle parole incomprensibili, confuse e balbettanti, facendo pensare a tutti di essere ubriaca (Skytg24, 2019). Il video fa il giro del mondo, raggiungendo più di due milioni di visualizzazioni, oltre 45 mila condivisioni e più di 23 mila commenti in cui gli utenti giudicano la Speaker a causa delle sue condizioni. Tutto questo non sapendo che il video era stato rallentato del 75% e il suo tono di voce aumentato, per rendere più credibile la sua condizione poco dignitosa (Skytg24, 2019).

Politici e opinion leader sono spesso colpiti dai video deepfake, soprattutto nei periodi delle elezioni in modo tale da influenzare l’opinione pubblica. Il problema più grande arriva nel momento in cui questi video sono entrati nel mercato della disinformazione politica.

Si è constatato inoltre che questo fenomeno, in piena emergenza sanitaria a causa della pandemia per Covid-19, abbia allarmato anche gli esperti dell'Organizzazione Mondiale della Sanità (OMS), che hanno ventilato la diffusione di una “infodemia”. Nel 2020 il *Pan American Health Organization* (PAHO) ha dichiarato che l’infodemia consiste in una “sovrabbondanza di informazioni, alcune accurate e altre no, che rende difficile per le persone trovare fonti attendibili e affidabili quando ne hanno bisogno” (PAHO, 2020).

L'infodemia si riferisce a un grande aumento del volume di informazioni associate a un argomento la cui crescita può avvenire in maniera esponenziale in un breve periodo di tempo a causa di un evento specifico, come l'attuale pandemia in corso. In questa situazione, appaiono sulla scena disinformazione e voci di corridoio e questo fenomeno si amplifica soprattutto attraverso i social network, diffondendosi sempre più velocemente, come un virus. In altre parole, stiamo parlando di una vera e propria "epidemia informativa", ossia una diffusione di notizie false o, appunto, di deepfake che contribuiscono a creare psicosi nella popolazione (PAHO, 2020).

Secondo il PAHO, è fondamentale interrompere questo pericoloso circolo vizioso: la disinformazione si espande allo stesso ritmo della produzione e distribuzione di contenuti. La stessa infodemia, inoltre, non solo accelera la disinformazione, ma anche la misinformazione, di cui abbiamo già parlato nel primo capitolo. Anche il Garante per la protezione dei dati personali inserisce il problema della disinformazione e delle fake news nella scheda informativa: i deepfake, infatti, possono coinvolgere politici o *opinion leader*, con lo scopo di confondere e influenzare l'opinione pubblica (GPDP, 2020). Per esempio, possono essere utilizzati per mostrare un politico in contesti e posizioni sconvenienti, o che pronuncia discorsi e parole che non ha mai detto, in modo tale da allontanare gli elettori che simpatizzano per lui. In questo senso, quindi, il deepfake può anche sottrarre alle persone l'autodeterminazione informativa e la libertà decisionale (GPDP, 2020).

Il problema più grande è quindi che la disinformazione, a causa di questa tecnologia emergente, sta diffondendo contenuti sempre più realistici, arrivando fino al punto in cui non saremo più in grado di distinguere il vero dal falso.

2.2 Dal furto di identità al cyberbullismo

L'aumento incessante della diffusione dei video deepfake ha inevitabilmente portato anche all'aumento dei rischi ad esso connessi. In questo paragrafo cercheremo di ripercorrere i pericoli principali legati agli effetti di questa tecnologia, che possono risultare molto lesivi soprattutto dal punto di vista della privacy. A questo proposito il Garante per la protezione dei dati personali nel 2020 ha messo a disposizione una scheda informativa per sensibilizzare le persone riguardo gli usi dannosi di questi video fake che

sono sempre più frequenti e sempre più facili da riprodurre anche attraverso il semplice utilizzo di uno smartphone. Uno dei problemi più ricorrenti e pericolosi legati al deepfake è il furto d'identità: spesso nei video gli individui che vengono riprodotti a loro insaputa e senza il loro consenso vengono privati non solo del controllo sulla diffusione della loro immagine, ma anche sulla divulgazione delle loro reali opinioni e idee (GPDP, 2020). Non solo, nella scheda informativa viene riportato anche il problema che questi soggetti potrebbero essere rappresentati in luoghi e contesti in cui non sono mai stati, o con persone che in realtà non hanno mai incontrato, potendoli mostrare, così, in situazioni compromettenti (GPDP, 2020).

Come afferma il Garante:

> "un deepfake può ricostruire contesti e situazioni mai effettivamente avvenuti e, se ciò non è voluto dai diretti interessati, può rappresentare una grave minaccia per la riservatezza e la dignità delle persone" (GPDP, 2020).

L'intelligenza artificiale, inoltre, sta offrendo sempre più possibilità per condurre attacchi informatici, che sono sempre più difficili da distinguere allo sguardo anche dei più esperti (Brighi C., Chiara P.G., 2021). In tutto ciò, sono le persone a dover essere protette dalla manipolazione che questa nuova tecnologia sta diffondendo, in quanto online

> "comportamenti lesivi quali molestie, odio, bullismo e stalking hanno - un forte potenziale lesivo della reputazione in virtù della persistenza delle informazioni e della loro potenziale diffusione virale. Il furto d'identità, agevolato spesso da lacune di sicurezza, può diventare strumento, oltre che per commettere truffe e frodi, per reati di diffamazione e per diffondere contenuti riservati a scopo di vendetta (*revenge porn*) o di estorsione e ricatto (*sexstortion*) e atti persecutori verso individui vulnerabili" (Brighi C., Chiara P.G., 2021).

Il problema è che l'origine e la causa di questi attacchi non dipende solo da chi li sferra o dalla sicurezza informatica delle organizzazioni ma anche dai soggetti stessi che vengono attaccati, in quanto spesso non vi è una sufficiente attenzione nella difesa dei propri dati personali. Questo porta spesso di conseguenza al rischio del furto della propria identità, di un danno alla propria reputazione o dell'esposizione crescente al *cyber crime* (Brighi C., Chiara P.G., 2021).

Il *cyber crime*, infatti, è un'altra conseguenza del deepfake che viene riportata nella scheda informativa del Garante. Le attività informatiche illecite che fanno parte del *cyber crime* sono principalmente tre: lo "*spoofing*", il "*phishing*" e il "*ransomware*" (GPDP, 2020).

Lo *spoofing* viene definito come "il furto di informazioni che avviene attraverso la falsificazione di identità di persone o dispositivi, in modo da ingannare altre persone o dispositivi e ottenere la trasmissione di dati" (GPDP, 2020).

Il *phishing* è "truffa informatica effettuata inviando un'e-mail con il logo contraffatto di un istituto di credito o di una società di commercio elettronico, in cui si invita il destinatario a fornire dati riservati, motivando tale richiesta con ragioni di ordine tecnico" (Oxford languages, 2022).

Il *ransomware*, infine, viene definito come "un tipo di software progettato per bloccare l'accesso a un sistema informatico fino al pagamento di una somma di denaro" (Oxford languages, 2022).

Attraverso queste attività vengono modificati il volto e la voce di un soggetto in modo tale da ingannare i sistemi di sicurezza basati su dati biometrici facciali e vocali (GPDP, 2020). Infatti, oggi molti sistemi digitali come assistenti vocali e alcuni sistemi bancari o sanitari, si servono di dati biometrici vocali e facciali come sistema di autenticazione per l'accesso e, in questo caso, video e audio deepfake potrebbero essere utilizzati per ingannare tali sistemi (GPDP, 2020). Ad esempio, gli individui che si dedicano a queste attività illecite possono creare video o audio-messaggi deepfake che vengono inviati alla gente comune per indurla a entrare su link o aprire allegati a dei messaggi che espongono computer, smartphone o altri dispositivi a intrusioni rischiose, oppure per convincerli a concedere, ingannandoli, informazioni e dati sensibili, come il numero della carta di credito. Quindi, anche se attualmente il livello delle tecnologie di sicurezza è abbastanza elevato e la qualità dei deepfake è relativamente ancora imprecisa, è importante prestare comunque sempre attenzione (GPDP, 2020).

Vi è poi un'altra attività dalle conseguenze estremamente dannose che rappresenta uno degli effetti molto gravi del deepfake: il *cyberbullismo*. Anche questo fenomeno rientra nella scheda informativa del Garante della protezione dei dati personali, dove viene affermato infatti che

> "un deepfake può essere realizzato per denigrare, irridere e screditare le persone coinvolte, o addirittura per ricattarle, chiedendo soldi o altro in cambio della mancata diffusione del video oppure per la sua cancellazione se è già stato diffuso" (GPDP, 2020).

Federico Tonioni (2014) approfondisce accuratamente questo fenomeno, affermando che si tratta purtroppo di un termine ormai attuale, in quanto rappresenta per la stragrande maggioranza degli adolescenti una minaccia molto concreta, quasi come l'alcol e la droga. Questo anche a causa di una società in cui la vita privata diventa sempre più pubblica nella dimensione digitale dei social network, in particolare di quella dei più giovani.

Anche se il bullismo tra adolescenti, per quanto possa essere lesivo, sia sempre esistito, il *cyberbullismo* è molto più crudele, tanto da poter diventare un "fenomeno incontrollabile" (Tonioni F., 2014). Questo perché il *cyberbullismo* permette ai bulli del web di rimanere protetti dall'anonimato della rete e di diventare disinibiti a causa della mancanza di contatto fisico con la vittima, a differenza delle relazioni dal vivo in cui istintivamente limitiamo l'espressione dei nostri pensieri e i comportamenti più aggressivi. Inoltre tutto ciò spesso si svolge sotto gli occhi dell'indifferenza degli altri utenti della rete, aggravata ancor di più dalla mancanza di controllo da parte degli adulti che spesso non sono a conoscenza delle dinamiche in cui si imbattono quelli che Tonioni soprannomina "nativi digitali".

Secondo l'autore il problema più grande di questo fenomeno è probabilmente la visibilità che è offerta dalla rete in quanto l'atto di bullismo richiede tre elementi: il carnefice, la vittima e degli spettatori. Quest'ultimi in particolare possono sia consapevolmente che inconsapevolmente aggravare l'atto di bullismo rimanendo indifferenti o condividendo il video deepfake della vittima. La divulgazione di un video che commette un furto dell'identità di un soggetto, soprattutto se giovane, può portare a delle conseguenze estremamente gravi, suscitando nella vittima dolore, incapacità di socializzazione e reazioni che potrebbero portare ad esiti estremamente drammatici. Per questo motivo è importante rivolgersi a delle istituzioni di tutela in modo tale da essere messi nelle condizioni di trovare delle soluzioni efficaci in caso di furto della propria identità in un video deepfake.

2.3 Deepnude: quando il deepfake ti spoglia

Vi è un'altra conseguenza molto grave legata al deepfake, anch'essa menzionata nella scheda informativa del Garante: il *deepnude*. Questo fenomeno potrebbe probabilmente essere posizionato nel podio degli effetti lesivi del deepfake, in quanto non solo genera un furto d'identità, ma anche una lesione della dignità e del pudore della vittima coinvolta in quanto in questa particolare tipologia di deepfake

> "persone ignare possono essere rappresentate nude, in pose discinte, situazioni compromettenti (ad esempio, a letto con presunti amanti) o addirittura in contesti pornografici" (GPDP, 2020).

Questo fenomeno permette infatti di prelevare il volto di un soggetto e posizionarlo attraverso particolari software sul corpo di una persona nuda o impegnata in atti e pose esplicitamente sessuali o pornografiche (GPDP, 2020). In tutto ciò, il Garante afferma che attraverso questa tecnologia è possibile anche "spogliare" dei corpi ricostruendo l'aspetto che quel soggetto ha sotto i vestiti in modo estremamente realistico.

Questo fenomeno ha coinvolto e coinvolge tutt'ora in modo particolare personaggi famosi, spesso con lo scopo di ricattarli o minacciarli. Questo tema è stato approfondito anche in un servizio[8] de "Le Iene" di Matteo Viviani del 2019, in cui spiega che il *deepnude* che spoglia i corpi delle persone è creato attraverso la GAN (rete generativa contradditoria) di cui abbiamo parlato nel primo capitolo.

L'intervistato Davide Cozzolino, ricercatore universitario, spiega che attraverso questa rete il software crea un'immagine coerente a quella che gli viene data. In pratica, questa strategia di addestramento prevede due ragionamenti: "uno che ha l'obiettivo di generare un'immagine sempre più realistica e un altro che ha come obiettivo di capire se un'immagine è realistica o meno".

Matteo Viviani cerca dunque di spiegare in parole semplici che il programma è stato precedentemente addestrato inserendo migliaia di nudi femminili. In questo modo, la rete generativa ha imparato a riprodurne uno partendo da zero, per cui se viene dato al programma un'immagine di partenza, lui si occupa di studiarla, di capire le proporzioni

[8] https://www.iene.mediaset.it/2019/news/deepnude-foto-donne-nude_578538.shtml

del corpo, la sua inclinazione, le ombre, le luci, le parti che sono coperte dai vestiti e le parti nude.

In seguito la rete generativa inizia a riprodurre da zero, pixel per pixel, una nuova immagine identica a quella che è stata caricata. Nel momento in cui la rete arriva ad analizzare le parti del corpo coperte dai vestiti, a esse non vengono sovrapposte delle parti del corpo nude di qualcun altro, ma vengono ricostruite da zero a partire dalle migliaia di esempi attraverso i quali il programma è stato addestrato. Quindi la rete è in grado di generare una parte del corpo nuda che non ha mai visto prima.

A questo punto, se il risultato è realistico, il programma lo propone; se invece, ritiene che non lo sia, la rete prova a creare un'immagine ancora più credibile, fino a che quella foto non sia più distinguibile dall'originale. Viviani ha anche deciso di mettere alla prova questi software sulle foto di due donne famose: Chiara Ferragni, imprenditrice e blogger italiana, e Veronica Ruggeri, inviata delle Iene. Caricando delle foto in costume delle due donne su un creatore di deepfake ha scoperto che non solo il programma ti dà la possibilità di scegliere quanti fake della stessa foto vuoi creare, ma che ti permette anche di selezionare la misura delle parti intime della vittima (misura dei seni, del capezzolo, dell'areola del capezzolo, della vagina e addirittura della quantità dei peli pubici da inserire nell'immagine di nudo).

A questo punto Viviani decide di mostrare le foto di nudo fake delle due donne ai rispettivi compagni, Federico Leonardo Lucia, in arte Fedez, rapper italiano, e Niccolò Devitiis, inviato delle Iene. Fedez ha confermato di avere un certo dubbio sulla veridicità della foto, ma dimostrando comunque una certa preoccupazione. In seguito all'invio delle foto fake alla moglie, comunque, scopre che in realtà quelle foto non sono altro che un fotomontaggio, ma conferma l'estrema credibilità di esse. Niccolò Devitiis, invece, si angoscia immediatamente alla visione di quelle immagini così credibili, infastidendosi e allarmandosi per la presunta divulgazione delle stesse.

Dopo aver smascherato il fatto e aver avvisato i compagni delle donne dell'inganno, Viviani intervista Fabiana, un'influencer e modella curvy di 24 anni, che ha subito la divulgazione di 36 foto deepfake di lei completamente nuda in un forum frequentato da milioni di utenti, in cui le discussioni si concentrano spesso sul tema del *deepnude*. In queste discussioni, migliaia di ragazzi e adulti chiedono di "spogliare" delle foto di donne e adolescenti su commissione. Come afferma Fabiana, ad oggi risulta

impossibile distinguere il reale dal falso, e dovrebbero essere previste delle conseguenze serie per questi atti. A questo proposito, infatti, questo fenomeno ha portato lo Stato della Virginia ad emanare la prima legge federale che rende i deepfake un reato pari a quello del *revenge porn*. Anche quest'ultimo viene menzionato nella scheda informativa del Garante, categorizzandolo come

> "la condivisione online - a scopo di ricatto, denigrazione o vendetta, da parte di ex partner, amanti o spasimanti respinti - di foto e video a contenuto sessuale o addirittura pornografico, che, nel caso del *deepnude*, sono ovviamente falsi" (GPDP, 2020).

Il Garante afferma che i video deepfake possono essere utilizzati per alimentare la pratica del *sexting*, che si tratta secondo la stessa fonte dello "scambio e diffusione di immagini di nudo, che a volte coinvolge anche soggetti minori", della pornografia illegale e della pedopornografica (GPDP, 2020), come ha dimostrato il servizio di Matteo Viviani.

2.4 Casi in cui il deepfake ha creato problemi: da Obama a Zelensky

"*We're entering an era in which our enemies can make anyone say anything at any point in time*"[9] , afferma l'ex presidente degli Stati Uniti d'America Barack Obama in un video su Youtube di BuzzFeed[10], con il fine di mettere in guardia dalla manipolazione dell'informazione presente nel mondo digitale in cui viviamo oggi.

Il problema è che in realtà Obama non ha né mai pronunciato quelle parole né mai registrato quel video, che è stato invece prodotto dal comico statunitense Jordan Peel, per dimostrare le conseguenze dannose che può provocare la divulgazione di un video deepfake. È facile immaginare la portata esplosiva di un video del genere che viene postato in rete e che subisce, inevitabilmente, un'inarrestabile diffusione: diventa una forma per influenzare e confondere l'opinione pubblica e per incrementare sempre più la sfiducia nelle istituzioni e nelle fonti di informazione. Infatti, uno degli aspetti negativi del deepfake che abbiamo già menzionato è la divulgazione della disinformazione, che

[9] "Stiamo entrando in un'era in cui i nostri nemici possono far dire qualsiasi cosa a chiunque in qualsiasi momento"
[10] https://www.youtube.com/watch?v=cQ54GDm1eL0

coinvolge spesso politici o *opinion leader* che possono essere mostrati in contesti e posizioni sconvenienti o pronunciando discorsi e parole che non hanno mai detto (GPDP, 2020).

È ciò che è accaduto nel video deepfake educativo di Obama divulgato su Youtube da BuzzFeed, nel quale non solo gli è stato fatto esprimere parole che non ha mai detto, ma gli è stato fatto anche pronunciare un pesante insulto nei confronti dell'ex Presidente degli Stati Uniti Donald Trump. Alcuni dei commenti sotto il video mostrano l'indignazione delle persone nei riguardi di questo fenomeno, per esempio un utente afferma: "come ho sempre detto, non fidarti di nessuno o di niente, soprattutto se è politico o ci sono soldi coinvolti, non importa la fonte, specialmente Internet"; un altro dichiara ironicamente: "Jordan Peel, per favore, potrebbe scrivere un altro thriller/film horror su questo? Sarebbe così interessante!"; in un altro commento un utente sostiene "Questo è BuzzFeed che ti dice di non fidarti di BuzzFeed", a riprova del fatto che questi tipi di video alimentano una forte sfiducia nei confronti delle fonti che divulgano notizie sui social media.

A questo proposito nel 2020 è stato condotto uno studio da parte di Vaccari Cristian e Andrew Chadwick: sulla base del loro studio, i due ricercatori hanno rilevato che le persone identificano correttamente i fake sono nel 50% dei casi (Vaccari C., Chadwick A., 2020). I due studiosi hanno intervistato online un campione di 2.005 persone rappresentative del Regno Unito da un ampio panel gestito da "*Opinium Research*". Hanno innanzitutto testato la loro fiducia nei confronti delle notizie che trovano sui social media; in seguito, per analizzare gli effetti della visione di video deepfake e i livelli di fiducia nei confronti delle notizie sui social media, hanno diviso in modo casuale il campione in tre gruppi, ognuno dei quali è stato sottoposto a tre parti differenti del video di BuzzFeed del 2018 in cui Jordan Peel prende le parti di Obama.

La prima parte dura 4 secondi e mostra Obama che apostrofa pesantemente Trump, rappresentando i tipici video "*troll*" che vengono condivisi più probabilmente in rete; il secondo gruppo, invece, è stato sottoposto alla seconda parte che dura 26 secondi, in cui viene mostrata la completa dichiarazione falsa di Obama, senza la rivelazione di Jordan Peel. Il terzo gruppo, infine, ha visto l'intero video di BuzzFeed che inizia con il video fake di Obama e termina con la rivelazione della manipolazione da parte del comico

statunitense che avvisa gli spettatori di prestare attenzione a questi video (Vaccari C., Chadwick A., 2020).

A questo punto i due ricercatori hanno chiesto ai partecipanti se pensassero che Obama avesse realmente pronunciato l'insulto a Trump e quanto si fidassero delle notizie sui social media. In seguito all'analisi delle risposte, i due ricercatori hanno scoperto che chi era stato sottoposto ai primi due video, quindi quelli senza la rivelazione di Peel, non avevano maggiori possibilità di essere fuorviati dal video fake rispetto a chi aveva osservato l'intero video educativo. Quindi, nonostante non gli fosse stata mostrata la verità, non erano più propensi ad essere convinti che Obama avesse effettivamente insultato Trump rispetto a coloro che invece avevano ascoltato le parole di Peel.

Tuttavia, i primi due gruppi avevano molti più dubbi sulla veridicità del video, infatti il 36% ha affermato di non poter dire se il video fosse reale o falso. Nel gruppo che ha visto completamente il video, invece, solo il 25, 7% dei partecipanti ha avuto dubbi sulla sua veridicità (Vaccari C., Chadwick A., 2020).

Successivamente Vaccari e Chadwick si sono concentrati sulla percezione dell'affidabilità dei partecipanti nei confronti delle notizie sui social media prima e dopo aver visto il video di Obama. I due ricercatori attraverso queste interviste hanno scoperto che le persone incerte sulla realtà o falsità del video avevano anche meno fiducia nei confronti delle notizie sui social media rispetto alle persone che non erano incerte in seguito alla visione del video.

Questo è importante perché il calo della fiducia potrebbe essere una risposta agli scandali di disinformazione online che hanno pervaso gli ultimi anni. Questo significa che la più grande minaccia della democrazia causata dai deepfake potrebbe non essere diretta, bensì indiretta.

I due ricercatori, infatti, affermano che il problema più grande non è che i deepfake potrebbero far credere alle persone qualcosa di falso, ma che potrebbero contribuire ad alimentare lo scetticismo e la sfiducia nei confronti delle fonti che divulgano le notizie, e questo porterebbe inevitabilmente alla nascita di una spirale interminabile, in quanto i contenuti condivisi sui social media generano incertezza; l'incertezza genera sfiducia; la sfiducia genera indifferenza e l'indifferenza rende le persone meno attente alla qualità dei contenuti che condividono sui social media.

Ma il caso di Obama non è l'unico. Ci sono stati dei casi politici più recenti legati allo scoppio della guerra in Ucraina, che hanno visto la diffusione di un video[11] deepfake in cui il Presidente ucraino Volodymyr Zelensky ordina alle forze armate del suo Paese di deporre le armi di fronte ai militari russi. Il video in questione sarebbe stato creato da degli hacker russi che lo hanno condiviso su un sito di news ufficiale ucraino a seguito di un attacco informatico, per poi diffonderlo sui social russi come "*Vk*" e successivamente anche su Twitter, Facebook e Youtube (Carboni K., 2022).

Il Presidente ucraino ha ovviamente smentito il video attraverso i social affermando:

> "rispetto all'ultima infantile provocazione e al consiglio di deporre le armi voglio consigliare alle truppe della Federazione russa di deporre le armi e di tornare a casa. Noi siamo a casa e difendiamo l'Ucraina". [12]

Anche le piattaforme come Twitter e Youtube hanno rimosso il video, ad eccezione dei social russi in cui circola ancora liberamente.

Questo caso, tuttavia, è interessante per un motivo: nonostante il video fake non fosse stato realizzato molto bene, l'accento non fosse preciso, il labiale non corrispondesse perfettamente e la sua testa fosse visibilmente più grande rispetto al corpo, questo non ha escluso la divulgazione e la preoccupazione di molti utenti, che hanno creduto al video nonostante le diverse imprecisioni. Infatti, un utente su Twitter afferma "Il video ha una qualità mediocre, ma è pur sempre un'arma"; un altro sostiene "Vogliono far credere che questo video sia falso, se tutto twitter è ai piedi di Zelensky, come può ancora circolare un video del genere?... Questo video è vero. È un ragazzo pericoloso e malato"; infine un altro utente dichiara "Come sappiamo che Zelensky è anche un vero leader? Potrebbe essere un CGI deepfake? In futuro, i deepfake ricopriranno effettivamente una carica politica e saranno i volti dei media e della politica controllati dagli oligarchi. Un po' come il mago di Oz", dimostrando la confusione che il video ha portato tra la folla.

Questo dimostra che questi video ora come ora possono avere ricadute estremamente pesanti soprattutto per le persone più fragili e più adulte che non sono

[11] https://www.youtube.com/watch?v=X17yrEV5sl4
[12] https://www.wired.it/article/guerra-russia-ucraina-video-deepfake-zelensky-resa/

abituate o non prestano abbastanza attenzione nel riconoscere un video falso. Ma in rete circolano già video di una qualità estremamente realistica, che rende ormai impossibile riconoscere il vero dal falso, colpendo così anche le fasce più giovani che sono sempre a contatto con questi strumenti e saprebbero riconoscerlo con più facilità: cercherò di trattare questo aspetto soprattutto nel terzo capitolo, in cui ho somministrato dei video falsi a un campione casuale di persone per osservare la loro capacità nel smascherare l'inganno.

2.5 Il deepfake e i suoi vantaggi: esiste un inganno positivo?

"Non cercare vantaggi dal male; i vantaggi del male sono equivalenti al disastro" affermava il poeta greco Esiodo (2016). Effettivamente i pericoli legati al deepfake sono probabilmente troppo gravi per giustificare alcuni aspetti che potrebbero rivelarsi utili, tuttavia non sono da escludere. La tecnologia deepfake presenta impieghi positivi in diversi settori e Westerlund (2019) ne elenca alcuni, tra cui il cinema, i media educativi, le comunicazioni digitali, giochi e intrattenimento, social media e assistenza sanitaria, scienze dei materiali e vari settori commerciali, come la moda e l'e-commerce: il primo impiego utile va sicuramente all'industria cinematografica, che trae vantaggio dal deepfake in diversi modi.

Il primo è sicuramente attraverso la CGI (*computer generated imagery)*, che si tratta di un'area delle pratiche di visualizzazione digitale che, dopo la sua comparsa alla fine degli anni '60, è arrivata rapidamente a ricoprire un ruolo privilegiato nella produzione cinematografica, che interessa in particolare l'animazione, gli effetti speciali e il *blockbuster* ad alto budget (Rehak B., 2011).

In queste aree, l'*imaging* digitale è costantemente spinto verso suoi limiti da uno stato dell'arte sempre in evoluzione. La CGI è una tecnologia che si identifica fortemente nel genere fantasy con spettacolari effetti speciali, come dimostrazione dell'abilità di Hollywood nel realizzare visioni fantastiche. Ma essa gioca un ruolo più significativo nei cosiddetti "effetti invisibili", che iniziano con il ritocco impercettibile della "verità" filmata e si propagano verso l'esterno verso ciò che alcuni hanno avvertito come la destabilizzazione del mezzo del cinema, presumendo una futura sostituzione

dell'industria a tutti i livelli, dalla produzione alla distribuzione, con la sua versione digitale (Rehak B., 2011).

L'attenzione alla CGI negli studi sui film e sui media si è in gran parte incentrata sugli effetti speciali, ovvero la creazione di illusioni sullo schermo con mezzi ingannevoli, ma non solo: un altro settore in cui viene impiegata è nei cortometraggi animati e nei lungometraggi, esemplificati dalla produzione della Pixar da Toy Story (1995) in poi (Rehak B., 2011). Westerlund, però, ricorda che questa tecnologia può essere anche utilizzata per creare voci digitali per attori che hanno perso la loro voce a causa di una malattia o di un problema momentaneo (Westerlund M., 2019). Il cinema è inoltre in grado di ricreare scene classiche nei film e creare nuovi film con attori morti da tempo (come si è visto nel primo capitolo per Peter Cushing in "*Rogue One: A Star Wars Story"* del 2017), utilizzare effetti speciali, un *editing* facciale avanzato in post-produzione e migliorare i video amatoriali con una qualità decisamente professionale.

La tecnologia deepfake consente anche il doppiaggio vocale automatico e realistico per film in qualsiasi lingua, consentendo così a un pubblico eterogeneo di godersi meglio film e media educativi. A questo proposito, Westerlund ricorda che una campagna globale di sensibilizzazione sulla malaria del 2019 con David Beckham[13], dove la star del calcio parla in nove lingue diverse per supportare l'associazione "*Malaria No More Uk*", ha abbattuto le barriere linguistiche attraverso una pubblicità educativa che ha utilizzato la tecnologia di alterazione visiva e vocale per farlo apparire multilingue.

La tecnologia deepfake è infatti in grado di rompere la barriera linguistica traducendo il parlato e alterando contemporaneamente i movimenti del viso e della bocca per migliorare il contatto visivo e far sembrare che tutti parlino la stessa lingua adattando i movimenti del labiale a ciascuna di esse. Questa tecnica può avere un grande impatto soprattutto sui mercati esteri dal punto di vista del marketing nei casi in cui la barriera linguistica può costituire un aspetto problematico.

Un altro impiego utile del deepfake va sicuramente anche ai giochi multiplayer e ai mondi di chat virtuali, che presentano maggiore telepresenza e assistenti intelligenti dalla voce più umana e naturale, il che aiuta a sviluppare migliori relazioni umane e interazioni online (Westerlund M., 2019). Lo stesso vale per i media educativi e la didattica, in quanto è possibile inserire all'interno di un video ambientato nei giorni nostri

[13] https://tg24.sky.it/mondo/2019/04/09/david-beckham-campagna-contro-malaria-video

un personaggio storico, un'artista o una celebrità della musica e della letteratura ormai deceduti, che permettano agli studenti di studiare e di osservare con maggior coinvolgimento alcuni aspetti didattici (Torrini F., 2022). Quindi potremmo osservare Van Gogh che spiega "La notte stellata" o Napoleone Bonaparte che racconta in un video didattico le sue gesta agli studenti, rendendoli così ancora più partecipi.

Per quanto riguarda le aziende, invece, esse sono interessate al potenziale della tecnologia deepfake per la sua possibile applicabilità al marchio, in quanto può trasformare in modo considerevole l'e-commerce e la pubblicità (Westerlund M., 2019).

Ad esempio, i marchi di moda possono assumere top model che non sono realmente top model per poi modificare la loro tonalità della pelle, l'altezza o il peso. Inoltre, i deepfake consentono di sviluppare dei contenuti personali che trasformano i consumatori stessi in modelli, attraverso la creazione dei cosiddetti "camerini virtuali".

Di questo nuovo progetto, sottomesso a dicembre 2018 al bando MISE "Fabbrica Intelligente, Agrifood e Scienze della vita" ed approvato dal Ministero dello Sviluppo Economico, ne ha parlato ampiamente Jolanda Coppola. L'autrice nell'articolo descrive *Try it on*, un "nuovo sistema di prova d'abiti virtuale (*virtual fitting*)", che rivoluziona il modo di scegliere e di provare i capi d'abbigliamento sia nei negozi fisici, provando i vestiti nel cosiddetto "*smart mirror*", sia online, attraverso il proprio smartphone (Coppola J., 2020).

La novità principale sta nelle tecniche utilizzate: *try it on* è stato addestrato a riconoscere le caratteristiche uniche della fisicità, modo di muoversi e postura degli utenti per permettergli di "sovrapporre, in *real time*, all'immagine reale della persona la visione degli abiti simulati, che si muoveranno seguendone il corpo e i movimenti" (Coppola J., 2020). Quindi, a differenza degli altri camerini virtuali già presenti nel mercato, *try it on* permette non solo di vedere sovrapposto un vestito alla nostra immagine, ma anche provare la sensazione di indossarlo veramente, e questo grazie all'utilizzo delle tecniche di *deep learning* e reti neurali attraverso i quali il sistema viene addestrato per riconoscere l'aspetto dei vari utenti (Coppola J., 2020).

Figura 2 "Smart mirror"

Ma il deepfake può rivelarsi particolarmente utile anche in un altro settore, ovvero nel caso in cui l'identità di un soggetto vuole e necessita di essere tenuta nascosta. Sono quei video in cui, grazie alle tecniche di deepfake, vengono nascoste per privacy le identità di alcuni individui, come quegli attivisti che sono promotori di determinate posizioni e dichiarazioni che non vengono accettate all'interno di determinati scenari socio-politici (Torrini F., 2022).

Un caso esemplare è quello del documentario "*Welcome to Chechnya*"[14], prodotto dall'emittente televisiva HBO, sulla persecuzione degli individui LGBTQ (acronimo italiano di: Lesbica, Gay, Bisessuale e Transgender) nella repubblica russa. Si tratta del primo video dove il deepfake viene utilizzato per tutelare l'identità di determinati soggetti a rischio. A questo proposito il direttore del documentario David France, ha voluto descrivere e raccontare in un'intervista su *Witness*[15] del 2020 tutta l'esperienza trascorsa, dai protocolli di sicurezza impiegati alle tecniche utilizzate.

France racconta di aver deciso di intraprendere questo progetto nel 2017, quando per la prima volta ha sentito parlare del disastro che stava colpendo la popolazione in Cecenia a causa delle persecuzioni nei confronti della comunità LGBTQ, sentendo così la necessità di esporre e condividere con il resto del mondo ciò che stava accadendo del sud della Russia (France D., 2020). La sfida più grande, afferma il direttore, è stata quella di aver a che fare per la prima volta con un mezzo visivo che richiedeva l'anonimato

[14] https://www.youtube.com/watch?v=_2KMm49B6pE, trailer "Welcome to Chechnya"(2020), HBO
[15] https://www.youtube.com/watch?v=2du6dVL3Nuc, "Identity protection with deepfakes: 'Welcome to Chechnya' director David France" (2020)

dell'identità degli attori coinvolti, per cui ha dovuto introdurre dei protocolli di sicurezza nelle riprese che proteggessero gli stessi, cercando però comunque di narrare la loro storia nel modo più intimo possibile.

La prima idea, è stata quella di usare la tecnica della "*rotoscoping animation*", che permette di "tagliare" un attore fuori dalla scena, per poi applicare una serie di filtri che lo trasformano in una versione animata. Tuttavia, la versione da cartone animato non permetteva di nascondere perfettamente l'identità degli individui coinvolti, in quanto non manipolava i corpi e i visi degli attori ma semplicemente riproduceva una figura animata simile al soggetto coinvolto. Inoltre, molti di loro non acconsentirono a questa pratica in quanto si sentivano insicuri di fronte a un eventuale rivelazione della loro identità.

Per evitare dunque che il filmato finisse nelle mani sbagliate, France ha innanzitutto trasferito il filmato dopo aver girato su unità crittografate, impedendo così a utenti non autorizzati di accedervi, nonostante in Russia non fosse legale. Il passo successivo è stato quello di reclutare un volontario e sottoporlo a una sessione di acquisizione dei dati facciali da diverse angolazione e con diverse telecamere, per usarla poi in una scena ad alta risoluzione del documentario, accorgendosi così che erano in grado di ottenere non solo una qualità superiore rispetto alla tecnica del *rotoscoping*, ma anche un livello di espressione estremamente realistica.

A quel punto, France ha deciso di appoggiarsi allo sviluppatore di software e specialista di VFX (effetti visivi) Ryan Laney, che è rimasto per un lungo periodo in una stanza di montaggio segreta completamente offline durante il suo lavoro, ma che gli avrebbe poi permesso di impiegare l'intelligenza artificiale e la tecnica del *deep learning* per mascherare i volti di 23 attori consentendo comunque di conservare un attaccamento emotivo, grazie al mantenimento dei soggetti di un aspetto umano.

France ha dunque assunto alcuni soggetti non correlati, per lo più attivisti *queer* di New York, che prestassero il loro volto ai soggetti russi che sono stati filmati. Inoltre, France afferma che la sua preoccupazione principale legata a questa tecnica, era quella di gettare gli spettatori nella "*uncanny valley*" di cui abbiamo ampiamente parlato nel primo capitolo, ovvero la paura che un eccessivo avvicinamento a un aspetto perfettamente umano dei soggetti sottoposti al *deep learning* avrebbe rischiato in qualche modo di alienare il pubblico dalla storia.

Per evitare questo inconveniente, France ha deciso quindi di rivolgersi alla dottoressa Thalia Wheatley, esperta della *uncanny valley*, per intraprendere una ricerca su come il pubblico avrebbe reagito di fronte a i doppi facciali. La dottoressa Wheatley ha dunque organizzato uno studio nel suo laboratorio a Dartmouth coinvolgendo 109 studenti universitari per testare la loro risposta emotiva al documentario, mostrando loro quattro versioni diverse di una stessa clip: la prima mostrava una clip del documentario con i volti reali degli attori, senza mascherarli attraverso la tecnologia deepfake; nella seconda, invece, veniva mostrata la clip con il *face swapping* completo dei soggetti, che è la tecnica finale che poi France ha deciso di utilizzare per il film; la terza clip mostrava un *face swapping* parziale, ovvero che modificava il viso degli attori, mantenendo però gli occhi reali dei soggetti; infine, l'ultima clip mostrava una versione animata del solo soggetto che necessitava di rimanere mascherato.

Sorprendentemente, la risposta emotiva più alta è stata registrata con la seconda clip, ovvero quella che prevedeva una completa sostituzione facciale degli attori coinvolti, motivo per cui poi France ha deciso di metterla in atto nel suo documentario. Assieme alla dottoressa Wheatley, France è arrivato alla conclusione che probabilmente il risultato sia dovuto a una questione di attrazione nei confronti della bellezza del viso mascherato misurata dagli studenti che erano coinvolti in questo studio, e anche al fatto che quegli stessi volti sono riusciti a superare brillantemente la cosiddetta valle misteriosa, a detta della dottoressa Wheatley.

Questo lungo e complesso lavoro di "*Welcome to Chechnya*" ha quindi permesso di dimostrare che è possibile usare in modo conveniente ed etico questa tecnologia che sfrutta l'intelligenza artificiale e l'apprendimento automatico, permettendo ai registi di documentari di raccontare delle storie che altrimenti non sarebbero mai state esposte al pubblico, e di conseguenza anche di consentire alle persone che soffrono a causa di questi abusi di reclamare la propria voce, aprendo così la strada a quello che France chiama il "cinema di giustizia sociale".

2.6 Deep nostalgia: il futuro riporta in vita il passato

Un altro settore molto ampio in cui gli aspetti vantaggiosi del deepfake si applicano in maniera particolarmente funzionale è quello sociale e medico: come afferma

Westerlund, questa tecnologia può per esempio aiutare le persone ammalate di Alzheimer a interagire con un viso più giovane per favorire un ricordo; oppure, può ricreare digitalmente l'arto di una persona che è stata amputata per aiutarlo a superare il trauma; inoltre, può consentire alle persone transgender di ricreare una migliore versione digitale del genere con cui preferiscono vedersi (Westerlund M., 2019).

Ma c'è un campo ancora più ampio e interessante nella quale il deepfake opera che ha preso il nome di "Deep nostalgia". Questo argomento è stato ampiamente approfondito da Noelia Barbero che elabora il suo pensiero a partire dalla teoria che Horst Bredekamp sviluppa nel suo libro "Immagini che ci guardano. Teoria dell'atto iconico", dove si interroga sul perché l'idea stessa di immagine, il suo fascino e la sua potenza siano temi sempre attuali (Bredekamp H., 2015). La sua idea di fondo, si basa sul fatto che le immagini non siano prodotti privi di capacità che possiedono, come nel caso delle opere d'arte, dei puri fattori e valori estetici, ma agiscono su di noi e sui nostri sentimenti e influenzano il nostro mondo psichico e simbolico.

È come se le immagini non fossero semplici immagini, ma avessero la capacità di provocare qualcosa, di influenzare il pensiero e il modo di agire di chi le osserva (Bredekamp H., 2015). L'autore nel suo libro propone inoltre tre tipi di atti di immagine: l'atto schematico dell'immagine, l'atto di sostituzione dell'immagine e l'atto intrinseco dell'immagine. Quello che ci interessa maggiormente e che potrebbe essere collegato alla Deep nostalgia è l'atto di sostituzione dell'immagine. Bredekamp definisce questo atto come la possibilità che nasce nello scambio reciproco tra corpo e immagine. L'autore nel suo libro propone l'esempio della *vera-icon*, che è un nome che viene dato a certe reliquie religiose. Per, esempio, la reliquia autentica di Cristo è stata utilizzata per ricreare attraverso particolari tecnologie delle immagini iconiche che avessero l'aspetto di Cristo, per questo motivo queste immagini possono essere percepite come se quella foto rappresentasse Cristo stesso.

A questo proposito Noelia Barbero (2021) sostiene che la Deep Nostalgia possa essere considerata come un atto di sostituzione dell'immagine. Quest'ultima si tratta di un servizio creato dal social network specializzato in genealogia "*My Heritage"* che, come viene affermato nel sito stesso, permette di animare i volti nelle vecchie foto di famiglia e nelle foto storiche e creare sequenze video realistiche di alta qualità.

Questo è stato possibile grazie all'utilizzo della tecnologia di D-ID, una società israeliana specializzata nell'animazione di immagini tramite *deep learning*. Il servizio sfrutta inoltre "*MyHeritage Photo Enhancer*", che secondo il sito, permette di mettere a fuoco e rendere più nitide le vecchie foto storiche che spesso tendono a essere rovinate e sfocate.

Per osservare il *modus operandi* di questo servizio, ho deciso di caricare una vecchia foto in bianco e nero di mio nonno da giovane: mi accorgo immediatamente che la risoluzione della foto è molto più alta, quasi come se fosse stata scattata ai giorni nostri; poi, da una semplice fotografia immobile, il viso di mio nonno prende subito vitalità, trasformandosi in un video di pochi secondi. Sbatte le palpebre, guarda a sinistra, le sbatte di nuovo e accenna un sorriso, a differenza della foto in cui assume un aspetto serioso; infine abbassa il capo e rivolge lo sguardo verso di me, come se mi stesse guardando veramente. Assieme al suo sguardo muove anche il capo, e tutto in modo estremamente realistico L'impatto emotivo è sicuramente molto forte, dato il grado di realismo che genera il servizio. Tutti questi movimenti sono stati registrati precedentemente da un gruppo di persone, per addestrare il programma a riconoscerli e a riprodurli attraverso la tecnica del *deep learning*.

Figura 3 Foto reale

Figura 4 Frammento di video modificato

Figura 5 Frammento di video modificato

In Deep Nostalgia di *My Heritage*, basta caricare una foto affinché questa diventi un video di pochi secondi che riproduce i movimenti facciali del soggetto coinvolto. Non

è possibile, inoltre, selezionare i tipi di movimenti e gesti da riprodurre nel viso, ma è il software stesso a generarli in maniera casuale sulla base dei dati che ha a disposizione.

Tuttavia, la raccomandazione e l'idea di fondo del sito di *My Heritage* è quello di sfruttare il servizio per animare le foto di parenti deceduti. Il nome Deep Nostalgia deriva proprio da questo: suscitare in chi usufruisce di questo servizio un sentimento nostalgico nel vedere animare il viso di una persona cara ormai deceduta, sfidando virtualmente la morte e riportando in vita il passato.

La domanda da porsi ora è: che impatto emotivo può avere un servizio del genere se la foto che viene caricata è di una persona deceduta per cui qualcuno sta soffrendo un lutto? Le risposte sono sicuramente contrastanti. Secondo la FAQ di *My Heritage*, alla domanda "questa tecnologia è affascinante, ma anche un po' inquietante, non ritenete?", il sito risponde:

> "Alcune persone adorano la funzionalità Deep Nostalgia e la considerano magica, mentre altri la trovano inquietante e non la apprezzano. In effetti, i risultati possono essere discutibili ed è difficile rimanere indifferenti a questa tecnologia. Ti invitiamo a creare video utilizzando questa funzionalità e condividerli sui social media per vedere cosa ne pensano i tuoi amici e familiari. Questa funzionalità è pensata per un uso nostalgico, cioè per riportare in vita i propri amati antenati. I nostri video di modello non includono parlato per prevenire abusi, come la creazione di video "deepfake" di persone viventi. Utilizza questa funzione sulle tue foto storiche e non su foto di persone in vita senza il loro permesso".

Westerlund, per esempio, sostiene che il deepfake può aiutare le persone ad affrontare un lutto riportando virtualmente "in vita" una persona cara deceduta (Westerlund M., 2019). Andrea Pinotti, invece, sostiene che l'idea che il digitale possa essere in grado di creare delle immagini che presentano una forma di immortalità disincarnata è pura illusione (Pinotti A., 2016). Ad ogni modo, questo tipo di servizio permette di esplorare ampiamente quali siano le potenzialità dell'Intelligenza Artificiale e dell'impatto emotivo che essa può avere sulle persone.

Questa nuova possibilità è anche spopolata in social come Tik Tok, in cui attraverso un filtro è possibile rianimare la foto di una persona cara deceduta esattamente come nel sito di *My Heritage*. Questo ha fatto sì che si creasse un trend mondiale per cui moltissime persone hanno deciso di rianimare i loro vecchi parenti o amici scomparsi e

filmare la loro reazione emotiva. Possiamo dunque vedere un video[16] con 4 milioni di visualizzazioni in cui un signore anziano vede una vecchia foto di sua moglie rianimarsi e, incredulo, si commuove, sostenendo "lei è viva, non ci posso credere". Nei commenti, inoltre, tutti si uniscono alla sua commozione, raccontando le proprie storie e chiedendo informazioni sulla modalità di creazione del video.

Persino su Netflix, in una puntata della seconda stagione della famosa serie "*Black Mirror*", che attraverso delle storie racconta i disagi che colpiscono il mondo moderno, si è parlato indirettamente della Deep Nostalgia. In questo caso però, la puntata non presenta gli aspetti poetici di questo servizio, ma quelli inquietanti e cupi e in una forma decisamente più evoluta che, chissà, magari un giorno potremmo raggiungere in un futuro.

La protagonista è Martha, che perde il marito a causa di un incidente stradale. La sua morte non gli dà pace e un giorno una delle donne che era presente al funerale del defunto marito, la iscrive a sua insaputa a un nuovo servizio che le permette di rientrare in contatto con il marito attraverso una *chat bot*. Dopo essersi intrattenuta a messaggiare con il finto marito virtuale, però, la donna non si accontenta e decide di iniziare a parlare al telefono con questa voce virtuale che simula le risposte che il marito avrebbe dato in base ai dati dei video che la donna aveva caricato e che sono stati assorbiti dal software per addestrare la voce.

Martha sembra essere contenta e appagata dalle conversazioni che intrattiene con la voce virtuale, tuttavia il *bot* gli consiglia di elevare ancor di più la loro relazione attraverso un ulteriore servizio sperimentale che permette di ricevere a casa un clone sintetico estremamente realistico del marito, che avrebbe avuto la funzionalità di parlare, di muoversi e di comportarsi come lui in base alle informazioni concesse dalla donna.

Quando il corpo del finto marito si attiva attraverso gli elettrodi, però, se in un primo momento a Martha non sembrava vero riavere suo marito di fronte a sé, in un secondo momento inizia a sentire un forte distacco e una sensazione di freddezza verso quel corpo così perfetto e non reale, cadendo in una sorta di *uncanny valley*. Mantenere una relazione con il suo finto marito defunto non gli permetteva di superare il lutto e di

16 https://www.tiktok.com/@storytimewithpapajake/video/6951857071373552901?is_copy_url=1&is_from_webapp=v1&q=deep%20nostalgia&t=1666344809982

trovare un altro compagno; inoltre, la faceva allontanare dalle persone a lei vicine, a cui era costretta a raccontare una menzogna per non far scoprire ciò che stava accadendo.

L'esempio riportato dall'episodio di *Black Mirror* è sicuramente a un livello più evoluto di tecnologia e di realismo rispetto alla Deep Nostalgia di *My Heritage*, anche se in realtà le *chat bot* per parlare virtualmente con una persona defunta sono già state programmate da Eugenia Kuyda (Cosimo S., 2016). La 29enne, infatti, in seguito alla perdita di un amico, ha deciso di creare una *chat bot* che sulla base dei messaggi che si inviavano su Telegram riproducesse delle risposte simili a quelle che lui stesso avrebbe dato. Le conseguenze a queste pratiche sono sicuramente molteplici e molto discusse: l'impatto psicologico che può avere sulle persone che soffrono un lutto è sicuramente discutibile, anche se questa pratica potrebbe essere utilizzata anche in altri settori, come quello cinematografico, per esempio. Tuttavia è necessario regolamentarne l'utilizzo, dato che è ampiamente disponibile e accessibile a qualsiasi persona e a basso budget.

In questo capitolo abbiamo esplorato principalmente gli aspetti negativi e positivi legati al deepfake e le conseguenze che questi possono portare. Siamo partiti dagli aspetti lesivi, che possono provocare un danno alla privacy e alla reputazione delle persone, per poi approfondire gli aspetti positivi, che possono aiutare a migliorare diversi settori che vanno da quello commerciale a quello psicologico. Sulla base delle informazioni riscontrate, risulta sensato sostenere che gli aspetti negativi e dannosi di questa tecnologia siano decisamente più gravi rispetto al beneficio che essa può portare, anche a causa della disinformazione e ignoranza diffusa in materia.

Tuttavia, la limitazione degli usi dannosi del deepfake attraverso dei provvedimenti e delle legislazioni potrebbe non escludere un utilizzo positivo dello stesso, prevedendo anche un miglioramento tecnologico che possa essere utile soprattutto in campo cinematografico e psicologico.

CAPITOLO 3
SONDAGGIO

3.1 Impostazione del sondaggio

Per osservare gli effetti del deepfake in modo diretto ho deciso di somministrare un questionario ad amici e conoscenti contenente 18 domande chiuse, di cui due relative ai dati anagrafici. Nella descrizione del questionario ho chiesto la collaborazione dei partecipanti, avvisandoli del tempo che avrebbero impiegato per rispondere a tutte le domande (5 o 10 minuti al massimo) e assicurandoli sull'anonimato delle risposte.

Per aiutarli nella compilazione ho inoltre fornito una breve esplicazione del termine deepfake, definendolo come "una nuova tecnologia emergente che permette di creare video fake attraverso l'intelligenza artificiale, manipolando il volto e la voce di una persona". Nelle prime 7 domande ho posto dei quesiti semplici e generali, per capire quanto i partecipanti fossero a conoscenza dell'argomento trattato. Nelle seguenti 7 domande ho deciso di somministrare ai partecipanti 7 frammenti di video derivanti da Youtube che durano dagli 8 ai 25 secondi, chiedendogli di indovinare quali tra quelli secondo loro fossero reali e quali falsi e manipolati.

In tutti i video sono presenti personaggi per lo più conosciuti, da politici come Barack Obama ad attori famosi come Morgan Freeman e Tom Cruise. Tra quei video, solo due sono reali: il numero 1, che ha come protagonista Elon Musk impegnato in un'intervista diretta da Jonathan Nolan riguardante il business e l'intelligenza artificiale, e il numero 6, che vede il comico Bill Hader cimentarsi in un'intervista condotta da Jimmy Fallon.

Tutti gli altri video sono stati modificati attraverso dei programmi deepfake che hanno permesso di manipolare il volto e la voce dei soggetti coinvolti, a partire dal

numero 2, che mostra un'intervista in cui viene sovrapposto il volto di Arnold Schwarzenegger a Bill Hader ; il numero 3 che vede come protagonista una falsa copia di Tom Cruise che lo imita alla perfezione; il numero 4 in cui Bill Hader impersona sempre il volto e la voce di Tom Cruise; il numero 5 in cui in un video educativo Jordan Peel prende le vesti di Barack Obama (lo stesso che si è visto nel secondo capitolo); infine, il numero 7, in cui Bob de Jong assume le fattezze di Morgan Freeman per mostrare agli spettatori la potenzialità di questa tecnologia.

Ho selezionato accuratamente tra i video falsi quelli che potessero al meglio confondere i partecipanti in modo tale da ricevere dei risultati veritieri sulla capacità tecnologica del deepfake di ingannare i rispondenti. Ho infatti posto alla fine due ultime domande sul livello di difficoltà riscontrato da quest'ultimi nel riconoscere la manipolazione presente dietro ai video. Dopo aver concluso l'elaborazione del questionario l'ho condiviso ad amici e conoscenti che l'hanno a loro volta condiviso con altre persone, fino al raggiungimento di 160 risposte.

3.2 Considerazioni generali sulle risposte del sondaggio

Tra le 160 risposte ricevute, innanzitutto, non tutti hanno risposto a tutte le domande: deduco quindi che alcuni si siano trovati in difficoltà di fronte ad alcuni quesiti.

I rispondenti sono stati per il 56,3% donne e per il 43,8 % uomini. L'età molto variabile mi ha permesso poi di avere un quadro ancora più ampio della tendenza diffusa: essa varia, infatti, dai 12 anni di età ai 70. Tuttavia, la maggior parte dei rispondenti appartiene alla fascia d'età 21/22 anni, quindi una fascia giovanile che mi ha permesso di verificare il grado di capacità di coloro che nella maggior parte dei casi hanno a che fare più probabilmente con i video deepfake, essendo più a contatto con il mondo dei social.

Il primo passo, dopo aver fornito una breve definizione del deepfake e una descrizione del questionario è stato quello di chiedere quanti partecipanti fossero a conoscenza del termine prima della somministrazione del questionario. Le risposte sono state molto equilibrate: il 51,9% era a conoscenza del termine mentre il 47, 5% no. Quasi metà dei rispondenti, quindi, non aveva mai sentito nominare questo termine e di conseguenza probabilmente non è neanche a conoscenza degli aspetti più tecnici che lo riguardano, rendendo così più possibile una non riconoscibilità di un video deepfake.

Inoltre un partecipante afferma: “Non sapevo avesse questo nome, avevo sentito parlare del fenomeno”.

Figura 6 Domanda 1

Alla domanda “Ne avevi mai sentito parlare senza sapere di cosa si trattasse?”, invece, il 61,3% afferma di no e il 35,6% risponde positivamente. Il fatto che molti non conoscano più approfonditamente il concetto deepfake fa capire quanta poca consapevolezza vi è riguardo a questo argomento, infatti un anonimo afferma: “Sapevo che esistesse ma non sapevo si chiamasse deepfake”. Solo 4 anonimi, tuttavia, affermano di essere a conoscenza dell’esistenza di questo termine e di sapere di cosa si tratta.

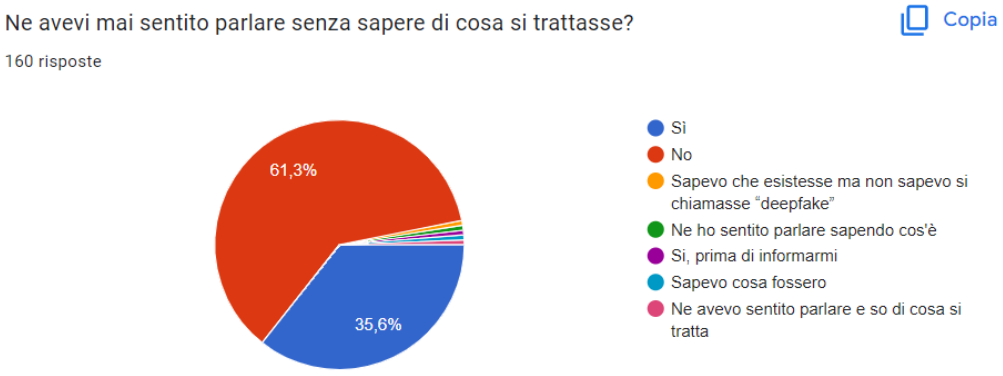

Figura 7 Domanda 2

Nonostante la poca conoscenza e consapevolezza del termine, alla domanda “Pensi di esserti mai imbattuto in un video deepfake?” il 66,3% risponde sì e solo il 26, 9% risponde no. È probabile, comunque, che anche questa piccola fetta di persone si sia trovato di fronte a un video falso almeno una volta nella vita, e il fatto che rispondano di

no fa dedurre che probabilmente non abbiano riconosciuto la manipolazione di un eventuale video deepfake in cui si sono imbattuti, o che forse non sono molto a contatto con i social media dato che ci sono moltissimi video deepfake che sono diventati virali in pochissimo tempo, da quello di Tom Cruise che gioca a golf a quello della Regina Elisabetta che balla durante il suo discorso natalizio. Secondo uno studio di *Sensity* (Ajder H., Patrini G., Cavalli F., Cullen L., 2019) il numero totale di video deepfake online è in rapido aumento, rappresentando quasi il 100% di aumento in base alla loro precedente misurazione (7.964) effettuata a dicembre 2018. Il numero totale di deepfake online infatti nel giro di sei mesi sarebbe ammontato a 14.678.

Inoltre 4 partecipanti, usufruendo dell'opzione "Altro", rispondono con sincerità "Non lo so", un altro risponde "Forse", altri due "Probabile" o "Potrebbe essere" e infine altri due affermano "Non sapendo di cosa si tratti, non saprei" o "Non so cos'è". Si percepisce quindi una certa confusione nei partecipanti che hanno risposto negativamente sulla possibile visione di un video deepfake.

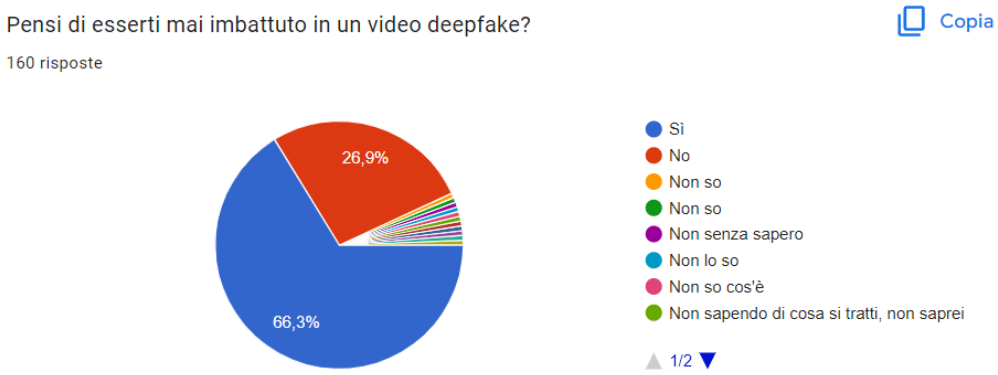

Figura 8 Domanda 3

La quarta domanda riguarda i social in cui le persone pensano sia maggiormente diffusa la divulgazione di questi tipi di video: ho deciso di porre questa domanda per capire principalmente se i partecipanti fossero a conoscenza della piattaforma "Reddit", social da cui i video deepfake hanno preso il nome e in cui tutt'ora continuano a circolare in maniera molto diffusa. Secondo la mia previsione, molti non sono neanche a conoscenza dell'esistenza di questo social. Le risposte, hanno confermato le mie ipotesi, infatti solo il 13, 1 % sostiene che Reddit sia il social in cui i video deepfake circolano

maggiormente. Questo significa che molti non sono a conoscenza della piattaforma e dei sui utilizzi.

Secondo lo studio di *Sensity*, uno dei social in cui maggiormente circolano video deepfake (prevalentemente a sfondo pornografico) è proprio Reddit. Il numero totale di visualizzazioni dei soli video deepfake pornografici nei primi 4 siti dedicati alla pornografia deepfake (tra cui Reddit) sono 134.364.438.

La maggior parte dei partecipanti, invece, sostiene che il social dove sono presenti maggiormente questi video è Tik Tok, che è stato selezionato dal 65%. Segue Instagram che si aggiudica il secondo posto con il 62, 5% e Facebook con il 60,6%. Infine vi è Twitter con il 16,3%, Youtube con il 2,5% e Telegram con l'1,3%. Interessante anche la scarsa selezione di Youtube: infatti dallo studio di *Sensity* è stato riscontrato che la pornografia deepfake prende di mira e danneggia esclusivamente le donne. Tuttavia, il deepfake non pornografico è molto diffuso su YouTube, che coinvolge per la maggior parte soggetti maschi, come si può evincere dai video che ho somministrato ai partecipanti e che sono stati selezionati esclusivamente da quest'ultimo (Ajder H., Patrini G., Cavalli F., Cullen L., 2019).

In realtà anche su Facebook girano diversi video deepfake, tanto che nel 2020 è stato condotto uno studio da alcuni ricercatori che hanno selezionato 120 video dal dataset di Facebook per capire se 20 soggetti riuscissero a riconoscere se fossero veri o falsi (Korshunov P., Marcel S., 2020). Questo social ha uno dei database più grandi e più recenti e presenta molte varianti differenti di deepfake, che vanno da quelli più palesi a quelli che sembrano molto realistici.

Anche social come Tik Tok, Twitter, Instagram e Telegram stanno comunque diventando il fulcro della diffusione dei deepfake. I dati presenti in questi social sono sicuramente moltissimi, basti pensare che Tik Tok conta oltre 1 miliardo di iscrizioni: più dati sono disponibili per alimentare gli algoritmi dell'intelligenza artificiale, più l'intelligenza artificiale diventa capace di manipolare o abusare di dati personali e

biometrici. Questo, di conseguenza, porta al pericolo che i dati vengano utilizzati deliberatamente per scopi criminali o comunque dannosi.

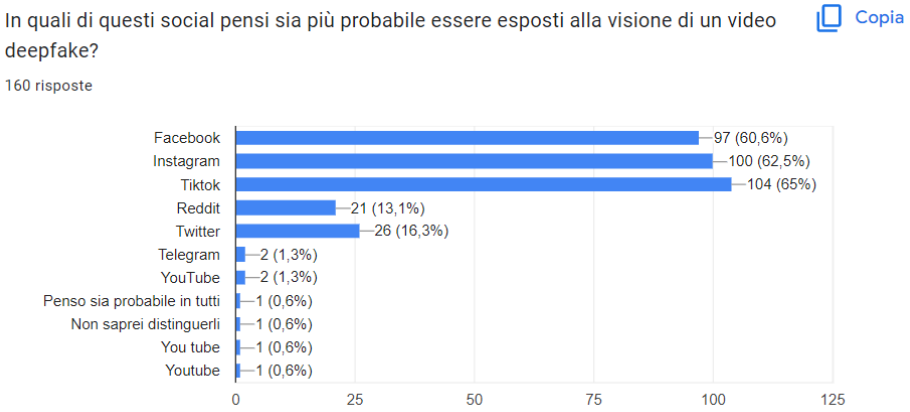

Figura 9 Domanda 4

La quinta domanda “Su una scala da "per niente dannoso" a "molto dannoso", quanto pensi possa esserlo un video deepfake?” è stata posta per rilevare la percezione della gravità avvertita dai partecipanti nei confronti di eventuali conseguenze dannose legate al deepfake. A questa domanda il 45,6% ha risposto “molto dannoso” e il 47,5% “abbastanza dannoso”. Nessuno tra i partecipanti sostiene che i video deepfake siano “per niente dannosi”.

La tendenza dimostra quindi che i rispondenti sono piuttosto consapevoli sul livello di lesione che può provocare la divulgazione di un video falso. Tuttavia vi è un 6,9% che sostiene che sia “poco dannoso”: questa piccola fetta di partecipanti probabilmente non ha messo in conto che il 96% del totale dei video deepfake online sono video pornografici falsi creati senza il consenso dei soggetti coinvolti; o che i primi quattro siti web dedicati alla pornografia deepfake ha ricevuto più di 134 milioni di visualizzazioni su video destinati a centinaia di celebrità femminili in tutto il mondo; o che l’aumento dei deepfake sta destabilizzando sempre di più i processi politici, mettendo a rischio l’integrità delle democrazie in tutto il mondo; o che questo fenomeno sta influenzando sempre di più il panorama della sicurezza informatica, migliorando le minacce informatiche tradizionali e abilitando nuovi vettori di attacco (Ajder H., Patrini G., Cavalli F., Cullen L., 2019).

Il 2019, infatti, ha visto diverse segnalazioni di casi in cui audio sintetici e immagini di persone sintetiche inesistenti hanno utilizzato il "*social engineering*" contro imprese e governi.

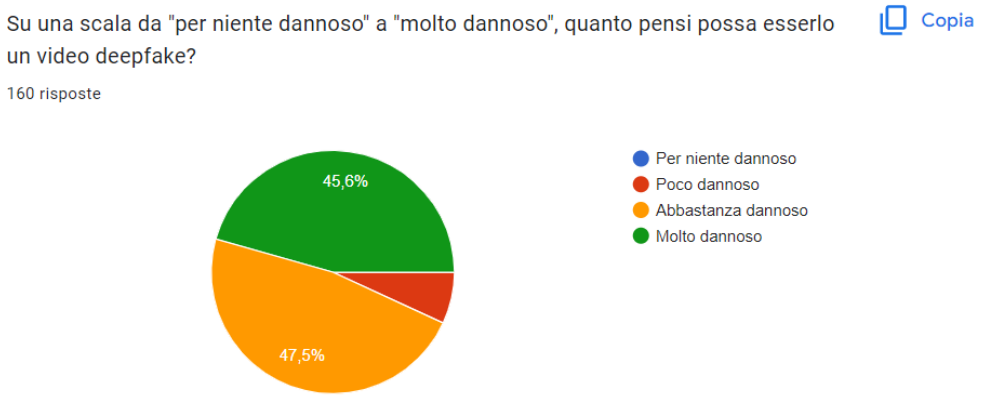

Figura 10 Domanda 5

Al contrario, la sesta domanda riguardava gli impieghi utili del deepfake. Alla domanda "Ritieni che i video deepfake possano essere utili in qualche campo?" il 51,6% ha risposto affermativamente, mentre il 48,4% ha risposto negativamente: quest'ultima fetta abbondante porta a dedurre che probabilmente non abbiano pensato o che non siano a conoscenza del fatto che la tecnologia deepfake viene utilizzata in modo etico in campo cinematografico, commerciale, didattico, tutelare, sociale e addirittura medico. Quindi questa risposta rappresenta un ulteriore dimostrazione della poca consapevolezza presente riguardo a questo fenomeno.

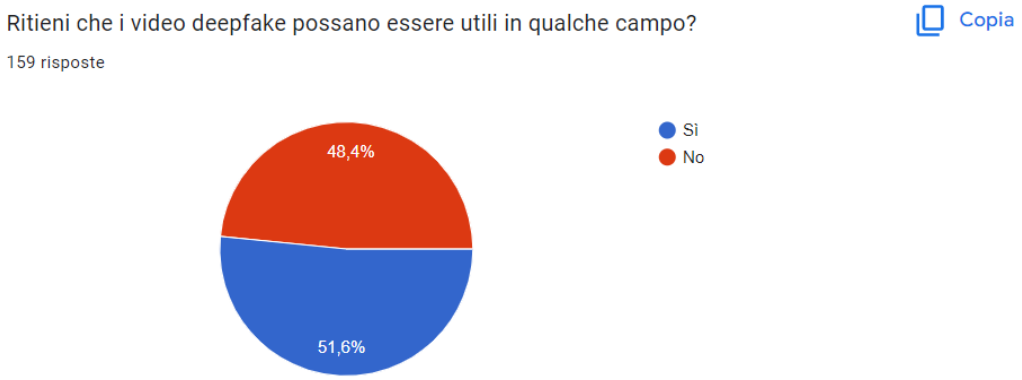

Figura 11 Domanda 6

Nella settima domanda chiedo ai partecipanti se si ritengono in grado di riconoscere un video deepfake. Il 34,4% risponde di sì, il 48,8% risponde consapevolmente di no. Altri anonimi affermano "Forse", "Non sono del tutto sicura", "Non credo di essere sempre in grado di riconoscere", "A volte", "Non sempre", "Abbastanza", "A volte sì a volte no" "Non lo so", non essendo sicuri di poterli riconoscere sempre con facilità.

Altri rispondono "Dipende", "Dipende ma tendenzialmente sì", "Forse, non ne sono sicura, dipende dalla qualità del video" "Dipende dalla tematica di riferimento", "Dipende da quanto bene è fatto", "Dipende dal contesto, ma spesso è davvero difficile" dimostrando che spesso dipende da alcuni fattori se è possibile riconoscere un video manipolato.

Un partecipante afferma anche "Nel dubbio verifico con una fonte sicura", azione che la maggior parte delle persone non fanno. Infine un anonimo risponde "Alcuni sono troppo accurati per essere riconosciuti", a riprova del fatto che molti non avendo gli strumenti necessari fanno fatica a riscontare l'inganno che sta dietro a un video deepfake.

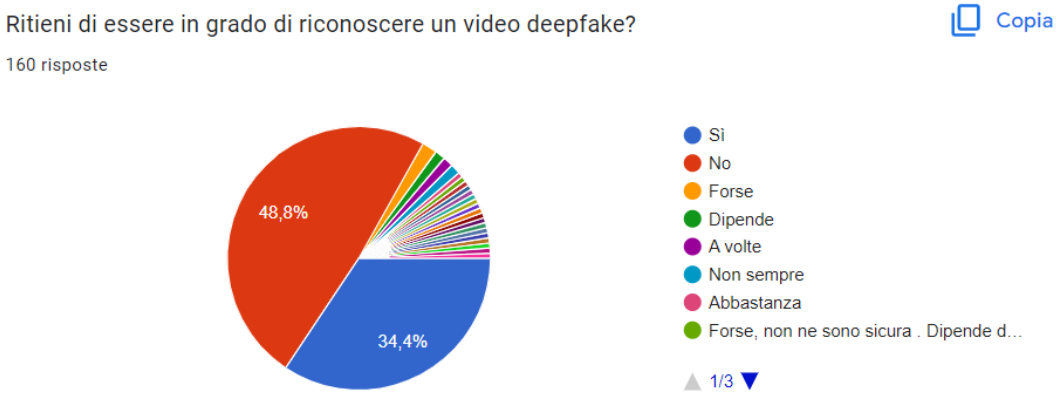

Figura 12 Domanda 7

Dall'ottava domanda si passa finalmente alla fase pratica: da questo momento somministrerò ai partecipanti cinque video deepfake e due video reali per capire quanti di loro riescono a distinguerli. Questa fase è fondamentale per comprendere quanto questo fenomeno sia evoluto e capace di confondere le persone, fino a non permetterci più di distinguere il vero dal falso. L'aspettativa infatti, era quella che buona parte di loro non riuscissero a riconoscere, tra tutti, i video deepfake e i risultati sono stati sorprendenti.

Il primo video è reale e ha come protagonista Elon Musk che risponde a un'intervista diretta da Jonathan Nolan riguardante il business e l'intelligenza artificiale. La percentuale che ha indovinato è solo il 53,1%. Il 46, 9 % afferma, invece, che si tratta di un video fake. Il fatto che quasi metà dei partecipanti abbia selezionato "fake" su un video che è reale, ci fa capire quanta confusione provoca questo fenomeno.

Il video non presenta alcuna anomalia, se non per il fatto che la qualità è leggermente scarsa, motivo per cui non mi aspettavo che una parte così alta dei partecipanti selezionasse "fake".

Figura 13 Video 1

Il secondo video è falso e riguarda un'intervista che è stata fatta al comico Bill Hader a cui è stato sovrapposto il volto di Arnold Schwarzenegger. Effettivamente il video appare molto realistico, grazie anche alla capacità di Bill Hader di imitare in modo molto similare la voce di Schwarzenegger. Il viso è perfettamente sovrapposto a quello del comico, inoltre il labiale è perfettamente sincronizzato. Possiamo dunque affermare che a meno che non ci si accorga che il volto sovrapposto è quello del celebre attore, sia molto complesso riconoscere l'inganno.

I risultati parlano chiaro: ben il 76,3% dichiara che il video è reale, mentre solo il 23,8% scopre l'inganno. Il dislivello è sorprendente, anche perché la qualità del video è più elevata rispetto al primo, quindi dovrebbe essere più facile scorgere eventuali anomalie. Tuttavia, la maggioranza dei rispondenti non è riuscita a riconoscere la manipolazione, cadendo così nella trappola.

Figura 14 Video 2

Anche il terzo video è falso e rappresenta una versione molto realistica di Tom Cruise che racconta di un incontro con Michail Gorbačëv, ex Presidente dell'Unione delle Repubbliche Socialiste Sovietiche. Anche in questo caso, solo il 54,7% ha riconosciuto l'inganno, mentre il 45,3% crede che il video sia reale. Anche questo video è di una qualità decisamente superiore rispetto al primo e non presenta indizi che potrebbero far pensare che sia falso, infatti una buona fetta dei partecipanti non è riuscito a scorgere che fosse un deepfake.

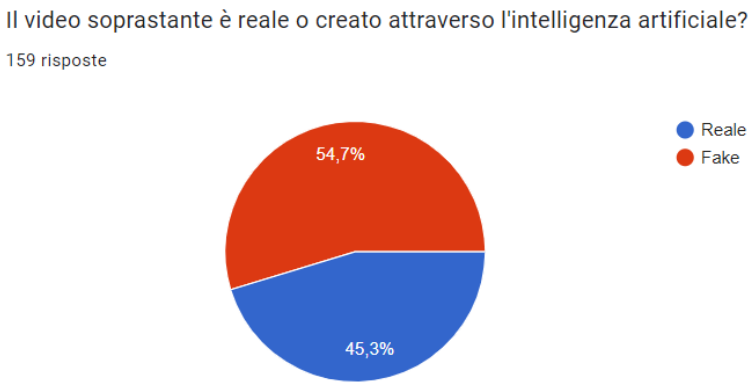

Figura 15 Video 3

Il quarto video ha sempre come protagonista Tom Cruise che, come possiamo notare, è stato coinvolto in diversi video deepfake. Anche questo video è falso, ma questa volta è Bill Hader a interpretare il suo volto e la sua voce in maniera estremamente realistica. La qualità e le luci del video in questo caso appaiono leggermente più scarse, ma anche in questo caso non vi sono irregolarità che possano far pensare che sia fake.

Infatti, ben il 62% ritiene che il video sia reale e solo il 38% indovina che si tratta di un deepfake.

Figura 16 Video 4

Passando al quinto video, possiamo vedere l'ex Presidente degli Stati Uniti Barack Obama che afferma che "stiamo entrando in un'epoca in cui i nostri nemici possono far sembrare che qualcuno stia dicendo qualcosa in qualsiasi momento, anche se non direbbero mai quelle cose". Il video è falso come si è visto nel secondo capitolo e la qualità è leggermente scarsa: tuttavia, questo non ha fermato il 27,8% ad affermare che fosse reale. Al contrario, il 72,2% seleziona l'opzione fake. In questo caso è stato quindi riscontrato abbastanza facilmente l'inganno, probabilmente per la qualità del video e per i movimenti leggermente meccanici del labiale e del corpo del falso Presidente.

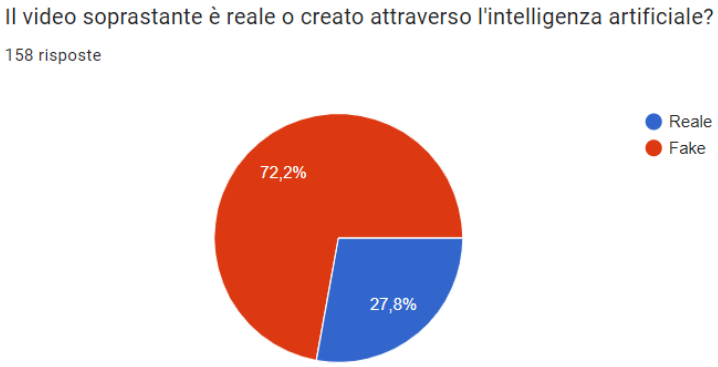

Figura 17 Video 5

Il sesto video, invece, è uno dei due reali e vede Bill Hader cimentarsi in un'intervista condotta da Jimmy Fallon. Nel video intero, il volto di Bill Hader veniva

scambiato con quello di Al Pacino: tuttavia ho selezionato una parte del video in cui il suo volto non veniva manipolato ma rimaneva quello reale del comico. Il 74,8% dei rispondenti in questo caso afferma che il video è reale, a differenza del 25,2% che ritiene che invece sia falso. Come nel video 1, quindi, nonostante il video sia reale e non vi sia alcuna modica del volto o della voce del soggetto coinvolto, vi è comunque una fetta di rispondenti che ritiene che sia falso, a riprova del fatto che l'eccessiva presenza di video deepfake che circolano nel web non ci rende più in grado di percepire cosa effettivamente sia vero e cosa sia fake.

Figura 18 Video 6

L'ultimo video, il numero 7, è falso e rappresenta un video educativo di Morgan Freeman che chiede agli spettatori quale sia la loro percezione della realtà. Nel video intero si mostra l'attore che afferma di non essere Morgan Freeman e di non essere nemmeno un essere umano, per poi scoprire alla fine che si tratta di un video deepfake educativo creato da Bob de Jong per mostrare le potenzialità di questa tecnologia.

Ho deciso di tagliare solo una parte del video per non renderlo così esplicito. Il 62,7% dei rispondenti in questo caso indovina la falsità che si cela dietro al video, ma rimane comunque un 37,3% che crede sia reale. Quel 62,7% ha probabilmente prestato attenzione alla lentezza delle parole pronunciate da Morgan Freeman. Tuttavia più di qualcuno in privato e nella domanda successiva ha confessato che è stato lo sfondo nero alle spalle dell'attore a fargli pensare che fosse un video fake. Questo mi fa dedurre che probabilmente se lo sfondo fosse stato un set o un altro ambiente più realistico, molti avrebbero risposto il contrario in quanto il video risulta estremamente realistico e similare

all'attore, sia dal punto di vista del volto, che è stato sovrapposto a un uomo bianco, sia dal punto di vista della voce, che è molto simile a quella di Freeman.

Figura 19 Video 7

Nella domanda successiva ho deciso di chiedere ai partecipanti quanto fosse stato difficile per loro riconoscere se i video fossero reali o falsi. Dalle risposte si evince che per il 70,7 % è stato difficile, mentre solo per il 16% non lo è stato.

Inoltre alcuni anonimi affermano "Abbastanza" oppure "A volte era complicato capirlo". Altri sostengono "Ho cercato da intuire dagli sfondi, dai colori e dalle movenze. Ma non è facile" oppure "Bisogna porre molta attenzione ai particolari" o ancora "Non lo so, sono andata a sensazione", un altro anonimo afferma "Ho l'impressione che i video fake siano più statici e il volto delle persone ha un qualcosa di falso". Molti si sono impegnati e si sono concentrati sui dettagli per cercare di rispondere correttamente alle domande.

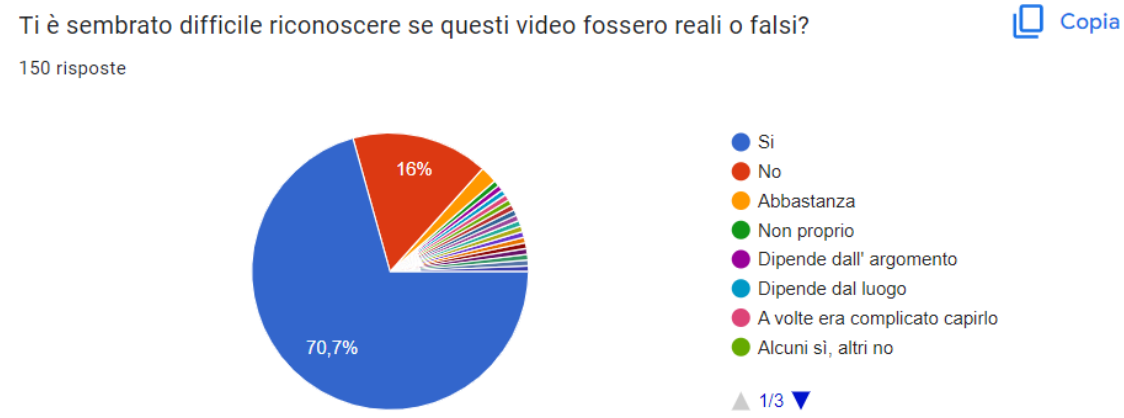

Figura 20 Domanda 8

Infine, ho posto un'ulteriore domanda per capire quali fossero stati i video più difficili da riconoscere per i rispondenti. La difficoltà maggiore è stata riscontrata nel video 7, ovvero quello di Morgan Freeman, che è stato selezionato dal 22%, il quale però è stato indovinato dalla maggior parte dei partecipanti; segue il video numero 3, ovvero quello della falsa copia di Tom Cruise che parla di Gorbačëv, con il 19,3%, che è stato infatti sbagliato da quasi la metà dei rispondenti.

Al terzo posto vi è il video 1 di Elon Musk, ovvero quello reale, che raggiunge il 16,7%, una percentuale abbastanza elevata per essere un video non fake. Per il 12,7% invece il video più difficile è stato il numero 5, ovvero quello di Barack Obama, nonostante anche qui la maggior parte dei rispondenti abbia indovinato l'inganno. Segue con il 10,7% il video 2 di Bill Hader che impersona Arnold Schwarzenegger, infatti pochi hanno svelato la manipolazione in questo video; infine il numero 4 di Tom Cruise che è stato selezionato dal 10% e il numero 6 di Bill Hader che è stato scelto solo dall'8,7%.

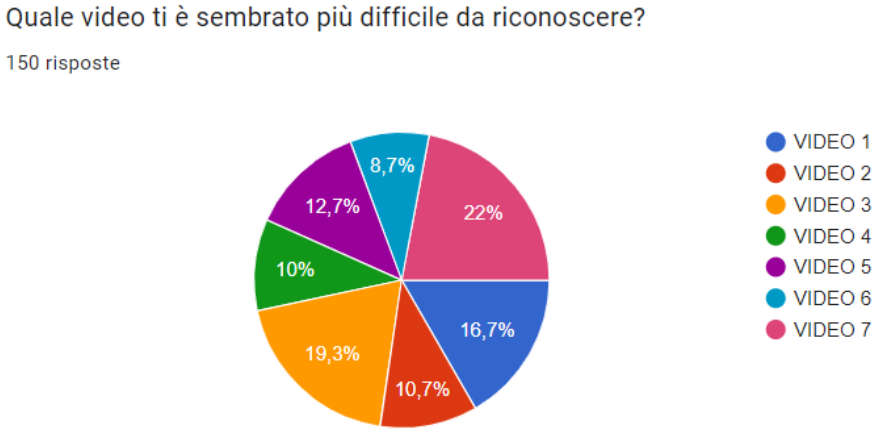

Figura 21 Domanda 9

Molti rispondenti, successivamente la compilazione del questionario, si sono rivelati interessati scrivendomi in privato per scoprire le risposte corrette. Dopo avergli mostrato la correzione, sono rimasti molto sorpresi, dichiarando di aver risposto in modo errato a diverse domande e di aver riscontrato molta difficoltà nella rilevazione dei video deepfake.

Questo sondaggio aveva come scopo principale quello di dimostrare che per qualsiasi persona che non sia dotata di strumenti adatti è difficile riconoscere un video deepfake e che questo influenza anche la credibilità dei video reali. I risultati in certi casi

hanno persino superato le aspettative, soprattutto nel secondo video in cui il 76,3% ha sbagliato la risposta.

La tecnologia deepfake si sta diffondendo sempre di più a causa dell'avanzamento tecnologico e questo rischia di diventare una minaccia sempre più grande per la nostra società, la sicurezza nazionale e la fiducia dei cittadini nei confronti dell'informazione giornalistica e delle autorità. I risultati del mio sondaggio, seppur rappresentativi di un piccolo campione, dimostrano che già oggi per molti è difficile distinguere il vero dal falso e che in futuro sarà praticamente impossibile se non vi saranno delle soluzioni adeguate che limitino la diffusione di questi video se creati a scopo disinformativo e lesivo della persona.

Capire quanto le persone riescano a riconoscere i deepfake è importante, ma lo è anche capire come li riconoscono gli algoritmi di rilevamento. Ritengo, dunque, che bisognerebbe impegnarsi nel gestire questo fenomeno al meglio: nel panorama delle leggi italiane non vi sono ancora provvedimenti che si occupano specificatamente della diffusione dei deepfake e delle loro conseguenze, inoltre non appaiono ancora nel panorama tecnologico degli algoritmi che riescano a rilevare in modo estremamente accurato questi video.

CAPITOLO 4
COME COMBATTERE IL DEEPFAKE

4.1 Legislazione anti-deepfake

In quest'ultimo capitolo cercheremo di osservare se vi sono in atto delle soluzioni per scalfire la portata dei deepfake e quali potrebbero essere apportati per limitare i danni da essi compiuti. Innanzitutto, potrebbero esserci due mezzi attraverso cui limitare i danni apportati dalla diffusione dei deepfake: attraverso una legislazione adeguata che si concentri esclusivamente sulla limitazione della portata dei danni di questo fenomeno e attraverso una maggiore istruzione riguardo il riconoscimento di questi video, in particolare per le fasce più fragili (tendenzialmente quelle più adulte).

Il panorama legislativo italiano e di istruzione riguardo questo fenomeno viene approfondito per la prima volta in maniera esaustiva al Convengo denominato "La minaccia del Deepfake: non basta più vedere per credere"[17], un dibattito organizzato da Videocittà Rassegna culturale nel 2019. In questo Convegno, Claudio Galoppi, consigliere per gli Affari Giuridici ed Istituzionali del Presidente del Senato, afferma:

> "Interrogarsi sulla minaccia del deepfake significa innanzitutto avviare un percorso di conoscenza sulle specificità di una tecnologia che rappresenta una delle più avanzate applicazioni dell'intelligenza artificiale. Generare e diffondere video falsamente credibili colpisce e sconvolge la nostra capacità di lettura e di comprensione della realtà, ma soprattutto esige che ciascuno di noi cominci ad acquisire strumenti nuovi di valutazione e di giudizio. Quindi si tratta di una nuova frontiera, del nuovo inizio di ulteriori e imprevedibili cambiamenti: basti pensare alle applicazioni e alle conseguenze di queste tecnologie, ad esempio in ambito economico e commerciale – alle conseguenze sociali, basti pensare sotto

[17] https://www.radioradicale.it/scheda/588781/la-minaccia-del-deepfake-non-basta-piu-vedere-per-credere

questo profilo al tema della protezione dell'identità personale e della violazione della privatezza della persona, alle conseguenze istituzionali e politiche – e alle conseguenze culturali. Sotto questo profilo occorre sottolineare che non si tratta più e solamente di confrontarsi con una realtà virtuale, ma di vedere completamente alterato e falsato il rapporto tra l'individuo e la realtà. La complessità di ognuna di queste riflessioni rende ancora più preziosa questa occasione di approfondimento che si inserisce in un percorso sostenuto e condiviso dal Presidente del Senato" (Galoppi C., 2019).

Il consigliere ricorda poi che seguirà anche un altro incontro il 9 dicembre in Senato nel quale sarà trattato e approfondito per la prima volta questo tema. Inoltre sottolinea:

"La prospettiva che interessa è duplice: da un lato formazione, conoscenza e consapevolezza del problema e delle sue implicazioni, dall'altro la necessità di fornire un contributo per l'elaborazione di un'efficace strategia di tutela e di protezione non solo dal punto di vista tecnologico e culturale, ma anche e soprattutto dal punto di vista normativo. Occorre cioè sensibilizzare anche in termini propositivi il legislatore perché da una situazione di pressoché totale vuoto normativo si passi a una regolazione mirata ed efficace del fenomeno" (Galoppi C., 2019).

Galoppi quindi evidenzia non solo il problema dell'assenza di una legge che tuteli l'individuo dall'esposizione al deepfake, ma anche della mancata istruzione e consapevolezza nei confronti del fenomeno: infatti, quando Rutelli, Presidente dell'ANICA, all'inizio del Convegno chiede per alzata di mano quanti siano a conoscenza del termine, pochissimi si fanno avanti, proprio come ha dimostrato anche il questionario analizzato nel terzo capitolo.

Durante il suo intervento, inoltre, il direttore del Servizio di Polizia Postale e delle Comunicazioni Nunzia Ciardi, afferma:

"Coniugare un'utilizzazione fruttuosa del mezzo informatico a una sicurezza accettabile costituisce la vera sfida per lo sviluppo economico e sociale di un paese moderno. Qui abbiamo un duplice problema: rendere sicuro l'ecosistema digitale e questo aspetta alle istituzioni, alle grandi aziende, alle amministrazioni pubbliche, e rendere sicuro l'ecosistema del singolo. Non basta rendere sufficientemente sicuro l'ecosistema globale se anche il singolo non è adeguatamente culturalmente attrezzato ad affrontare questo mondo" (Ciardi N., 2019).

Inoltre, alla domanda di Rutelli "La tecnologia non ha nazione, ha un territorio globale. Come fate a gestire qualcosa che trascende il territorio italiano, una minaccia che ci riguarda tutti e che ci riguarda a livello planetario?", Ciardi risponde:

> "I fenomeni patologici della rete oggi non possono essere gestiti con una legislazione nazionale. È miope pensare che una norma nazionale riesca in qualche modo ad essere un argine accettabile al crimine *cyber*. Questo perché è evidente che la rete ha sbriciolato ogni confine spazio-temporale, per cui se io individuo un reato e scopro che il server che ospita un sito pedopornografico o filmati deepfake in Nigeria, io non posso andare lì e sequestrare il sito. Quando noi veniamo a conoscenza di un sito che contiene orrende immagini pedopornografiche e sappiamo che non è in Italia, io non posso far altro che avvisare con i canali internazionali e di cooperazione l'altro paese, sperando che collabori, ma il massimo che posso fare è ordinare ai miei *provider* di mettere quegli indirizzi ip in una *black list*, una lista nera che renda irraggiungibile dall'Italia quel sito, ma quel sito continuerà ad esistere perché io non ho giurisdizione sul paese che ospita quel sito, quindi posso limitarmi a renderlo irraggiungibile dall'Italia" (Ciardi N., 2019).

Quindi anche il direttore sostiene che vi sia un vuoto dal punto di vista normativo che dev'essere gestito nei migliori dei modi. In questo caso, però, entra in gioco un altro fattore importante: non basta formulare un disegno di legge che riguardi la protezione dal deepfake nel nostro Paese, perché il web è accessibile a tutti e tutti possono navigare in siti stranieri in cui vengono diffusi video deepfake. Un passo importante per la difesa da questo fenomeno potrebbe essere quindi quello di chiudere agli italiani l'accesso a determinati siti diffusori di video deepfake dannosi attraverso l'inserimento degli stessi in una *black list* che impedisca l'accessibilità e la condivisione.

Per quanto riguarda il tema dell'istruzione il commissario dell'Autorità per le Garanzie nelle Comunicazioni Mario Morcellini sostiene:

> "La parola sostenibilità è importante perché significa che le innovazioni devono in qualche misura essere in grado di raggiungere soggetti sociali che sono molto disuguali dal punto di vista delle opportunità culturali. Noi vediamo quelli astuti e abili per i quali tutte le innovazioni si trasformano in una vitamina della conoscenza, ma non è così per tutti. La seconda parola chiave è *media education*, è pazzesco che questo paese non si doti di una struttura a basso costo che risolverebbe in buona misura i problemi della disinformazione.

> Non ha senso che la scuola contemporanea, tranne i generosi episodi di scuola digitale, non considerino la comunicazione il vero libro di testo su cui costruire l'atlante dei saperi: in questo modo il docente finisce per apparire arretrato anche quando non lo è e parla di mondi scomparsi, mentre la funzione del docente non è di essere aggiornato quanto il suo allievo ma di aiutarlo a mettere in fila i saperi, cosa che può fare solo un insegnante adulto" (Morcellini M., 2019).

Quindi dal punto di vista educativo è importante portare a livello scolastico una nuova forma di conoscenza e di apprendimento legato al mondo della comunicazione in modo da istruire i giovani fin da subito e renderli in grado di non farsi ingannare da un video deepfake. Questo però non è ancora presente nello scenario italiano in cui prevalgono principalmente la scelta di materie teoriche e umanistiche che non includono materie legate alla comunicazione, a meno che non parliamo di università o corsi extra, motivo per cui sarebbe significativo introdurle. Tuttavia, non possiamo considerare solo coloro che sono abili nell'utilizzo della tecnologia e del web, ma anche coloro che non ne sono a stretto contatto e che di conseguenza risultano essere fragili nel riconoscimento di un video manipolato.

Ci troviamo quindi di fronte a un fenomeno sotto certi punti di vista estremamente pericoloso, che ha bisogno di un intervento da parte del legislatore. A questo proposito gli studenti del Liceo Don Carlo la Mura hanno provato a creare un disegno di legge orientato esclusivamente al deepfake che garantirebbe una tutela e una protezione delle persone coinvolte.

Il disegno è stato presentato ai senatori e si divide in quattro articoli che riguardano rispettivamente l'ambito di applicazione, i soggetti destinatari, la procedibilità e i diritti della persona offesa. Secondo l'articolo 1:

> "1. La presente legge si applica a video manipolati, detti deepfakes, che, realizzati, pubblicati e divulgati senza il consenso e l'autorizzazione della persona riportata nel video fake, violano i diritti alla privacy, all'onore, all'immagine, al decoro, alla reputazione ed ogni altro diritto della persona stessa.
>
> 2. Ai fini della presente legge si intende per deepfake: qualsiasi immagine e/o video, in ogni modo realizzato, che combini e/o sovrapponga immagini e/o video di una persona su altre immagini e/o video di altra persona, al fine di generare un video realistico, ma finto o fake" (Liceo don Carlo La Mura, 2020).

Nell’articolo 2 si afferma invece:

“1. Chiunque, senza il consenso della persona interessata, realizza ed invia ovvero consegna, cede, diffonde o pubblica, con qualsiasi mezzo, un deepfake di cui all’art. 1, salvo che il fatto costituisca più grave reato, è punito con la reclusione fino a 3 anni e la multa fino ad € 5.000.

2. Salvo che il fatto costituisca più grave reato, alla stessa pena soggiace chiunque, avendo ricevuto, o in qualsiasi modo acquisito, un deepfake, a sua volta lo invia, consegna, cede, diffonde, inoltra o pubblica, con qualsiasi mezzo, senza il consenso della persona interessata.

3. Le pene di cui ai commi 1 e 2 sono aumentate della metà se il fatto è commesso in danno di persona minore, di persona con limitazione psico-fisica, o di persona con disabilità come definita dalla legge 5 febbraio 1992, n. 104” (Liceo don Carlo La Mura, 2020).

Nell’articolo 3 si prosegue con la procedibilità, dove si dichiara che:

“1. Il reato è perseguibile a querela di parte ad eccezione delle ipotesi di cui all’art. 2 comma 3 e quando è connesso ad altro reato perseguibile d'ufficio.

2. Il termine per la proposizione della querela è di sei mesi.

3. Si procede d'ufficio nei casi di cui all'articolo 2, comma 3, nonché quando il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio” (Liceo don Carlo La Mura, 2020).

Infine, nell’articolo 4 si afferma:

“1. Nel caso in cui la fattispecie prevista dalla presente legge integri anche il reato di diffamazione, ovvero altro reato, la persona offesa può chiedere il risarcimento dei danni ai sensi dell’articolo 185 del codice penale, determinato in relazione alla gravità dell’offesa e alla diffusione del deepfake.

2. La persona offesa può altresì chiedere la rimozione dalle piattaforme telematiche, dai social network, dai siti internet, dai motori di ricerca e da qualsiasi altro canale di diffusione, dei deepfake di cui all’art. 1 della presente legge.

3. La persona interessata, in caso di rifiuto o di omessa cancellazione del video, ai sensi dell’articolo 14 del decreto legislativo 9 aprile 2003, n. 70, può chiedere al giudice di ordinarne la rimozione, dalle piattaforme telematiche, dai social network, dai siti internet, dai

motori di ricerca e da qualsiasi altro canale di diffusione, ovvero di inibirne l'ulteriore diffusione.

4. In caso di morte dell'interessato, le facoltà e i diritti di cui ai commi precedenti possono essere esercitati dagli eredi o dal convivente" (Liceo don Carlo La Mura, 2020).

In realtà nel panorama delle leggi italiane sono stati presentati diversi disegni di legge per contrastare questo fenomeno che lega il tema delle fake news con quello della tecnologia. Per esempio, nel 2017 ne è stato presentato uno che riguarda "disposizioni per prevenire la manipolazione dell'informazione online, garantire la trasparenza sul web e incentivare l'alfabetizzazione mediatica" che propone l'introduzione di una nuova infrazione nel codice penale, inserendo l'articolo 656-bis "pubblicazione o diffusione di notizie false, esagerate o tendenziose, atte a turbare l'ordine pubblico, attraverso piattaforme informatiche" includendo anche la "diffusione di notizie false che possono destare pubblico allarme o fuorviare settori dell'opinione pubblica" e la "diffusione di campagne d'odio volte a minare il processo democratico" con i quali ci si rivolge ai casi in cui la diffusione di fake news influenzi l'opinione pubblica, portando a casi di allarme pubblico (Liceo don Carlo La Mura, 2020).

Lo stesso anno ha visto una nuova iniziativa parlamentare che riguarda "Norme generali in materia di social network e per il contrasto della diffusione su internet di contenuti illeciti e delle fake news", che oltre a contrastare le fake news ha come scopo quello di regolamentare il fenomeno soprattutto dal punto di vista dei social, attribuendo determinate responsabilità ai fornitori di servizi del web.

Nel 2019 segue una prima proposta di legge che riguarda la "diffusione seriale e massiva di contenuti illeciti e di informazioni false attraverso la rete internet, le reti sociali telematiche e le altre piattaforme digitali", che parte dal presupposto che le fake news sono diventate un fenomeno molto diffuso che rischia di influenzare la politica soprattutto durante i periodi di elezione e di distruggere la reputazione non solo di figure pubbliche, ma anche dei normali cittadini.

In conclusione, negli ultimi anni, a causa dell'aumento incessante delle notizie false che influenzano l'opinione pubblica e che creano una forte instabilità sociale dal punto di vista della credibilità e della fiducia dei cittadini, sono stati attuati diversi tentativi con lo scopo di definire un limite alla diffusione del fenomeno. Tuttavia, manca

ancora nella sfera legislativa italiana una disposizione che regoli in modo specifico il problema dei deepfake e che potrebbe essere utile per limitare i danni da esso conseguiti.

4.2 Social e tecnologia per il contrasto ai deepfake

Introdurre una legislazione efficace contro la diffusione dei deepfake e un'istruzione che permetta di inserire i cittadini in un contesto di apprendimento mirato, però, non è l'unico modo. Potrebbero essere altri due i mezzi per impedire la divulgazione di video deepfake a scopo lesivo: attraverso un'accurata tecnologia anti-deepfake che permetta agli algoritmi di riconoscere se un video è falso o no, e, infine attraverso una maggiore presa di posizione da parte dei social che sono i principali mezzi di veicolazione di questo fenomeno.

La possibilità che chiunque possa oggi produrre un video deepfake attraverso l'utilizzo di un semplice software che tutti possono trovare sul web, rende necessario lo sviluppo di sistemi di rilevamento che siano in grado di riconoscerli in modo accurato. Nonostante gli esperti di *digital forensics* sappiano analizzare singoli video manipolatori ad alto impatto, non vale lo stesso per le migliaia di video che vengono condivisi sul web o sulle piattaforme social ogni giorno (Dolhansky B., Bitton J., Pflaum B., Lu J., Howes R., Wang M., Ferrer C., 2020). Il rilevamento di deepfake su larga scala richiede una serie di metodi di visione computerizzata o modelli multimodali che richiedono determinati dati di addestramento (Ibid., 2020). A questo proposito, sono state sviluppate diverse tecniche di rilevamento fino ad oggi, basate principalmente su tecniche di *machine learning*.

Lo studioso Peng Chen, per esempio, ha sviluppato "FSSPOTTER", un framework che si occupa di indagare le caratteristiche spaziali all'interno di un unico fotogramma con l'aiuto di uno *Spatial Feature Extractor* (SFE) insieme a un *Temporal Feature Aggregator* (TFA) che estrae le incongruenze tra le cornici (Saikia P., Dholaria D., Yadav P., Patel V., Roy M., 2022). Digvijay Yadav, invece, considera il battito di ciglia come una delle caratteristiche fondamentali per riconoscere il deepfake, attraverso la rilevazione di incongruenze temporali nei cambiamenti nelle cornici; Irene Amerini, è un'altra studiosa che ha introdotto un metodo per sfruttare le incongruenze temporali dei video con campi di flusso ottici tra due frame consecutivi per distinguere i video originali

da quelli falsi. Shivangi, invece, ha proposto un approccio basato sull'apprendimento del trasferimento, denominato come "*Deep Distribution Transfer*" per il rilevamento di falsificazioni facciali (Saikia P., Dholaria D., Yadav P., Patel V., Roy M., 2022). Tipicamente, gli algoritmi CNN sono ampiamente esplorati in letteratura per apprenderne ogni fotogramma della sequenza video. Gli algoritmi si concentrano principalmente sull'estrazione delle funzionalità *intra-frame* per il rilevamento dei deepfake. Tuttavia, anche la funzionalità *inter-frame* per sfruttare le incongruenze temporali può rilevarsi una direzione promettente nella ricerca del rilevamento dei deepfake (Ibid., 2022).

Un'iniziativa molto importante per la rilevazione di deepfake è stata lanciata nel 2019 da Amazon Web Services, Facebook, Microsoft, il Comitato direttivo per l'integrità dei media di *Partnership on AI* e altri accademici: l'iniziativa prende il nome di "*Deepfake Detection Challenge"* e si tratta di una vera e propria sfida che ha avuto come obiettivo quello di "stimolare i ricercatori di tutto il mondo a costruire nuove tecnologie innovative che possano aiutare a rilevare deepfake e media manipolati" (Kaggle, 2019).

La sfida si è chiusa nel 2020 e prevedeva anche dei premi in denaro che andavano da $ 500.000 USD a $ 40.000 USD e che sono stati assegnati ai primi cinque posti in classifica. Come riporta la descrizione di Kaggle, i partecipanti alla sfida erano tenuti a inviare il loro codice in un ambiente "*black box*" per il test e avevano inoltre la possibilità di aprire o chiudere la propria iscrizione al momento dell'accettazione del premio. Le proposte aperte erano tenute a rispettare i termini di licenza open source per raggiungere i premi della classifica. Le proposte chiuse, invece, sono di proprietà e non hanno potuto accettare i premi. Indipendentemente dalla scelta, tutti gli invii sono stati valutati allo stesso modo e i risultati sono stati pubblicati nella classifica su Kaggle.

Inoltre,

> "il comitato direttivo del PAI ha sottolineato la necessità di garantire che tutti gli sforzi tecnici incorporino l'attenzione su come il codice risultante e i prodotti basati su di esso possono essere resi il più accessibili e utili possibile ai principali difensori della qualità dell'informazione in prima linea come giornalisti e leader civici intorno al mondo. I risultati del DFDC daranno un contributo a questo sforzo e costruiranno una solida risposta alla minaccia emergente rappresentata dai deepfake a livello globale" (Kaggle, 2019).

A partecipare a questa iniziativa sono state 2.265 squadre e il primo posto è stato aggiudicato all'ingegnere Selim Seferbekov che ha utilizzato un approccio di classificazione fotogramma per fotogramma (Kaggle, 2019). Selim, inoltre, assieme ad altri studiosi, si è concentrato sull'identificazione dei limiti e delle carenze dei *framework* di rilevamento dei deepfake esistenti (Das S., Seferbekov S., Datta A., Islam Md. S., Amin Md. R., 2021). Insieme hanno identificato alcuni problemi chiave relativi al rilevamento dei deepfake attraverso l'analisi quantitativa e qualitativa dei metodi e dei set di dati esistenti e hanno scoperto che i set di dati deepfake sono ampiamente sovracampionati, causando un facile *overfitting* dei modelli.

I set di dati vengono creati utilizzando un piccolo insieme di volti reali per generare più campioni falsi. Quando vengono addestrati su questi set di dati, i modelli tendono a memorizzare i volti e le etichette degli attori invece di apprendere le caratteristiche false. Per mitigare questo problema, Selim e gli altri studiosi hanno proposto un semplice metodo di aumento dei dati chiamato "*Face-Cutout*": questo metodo ritaglia dinamicamente le regioni di un'immagine utilizzando le informazioni sul punto di riferimento del viso e aiuta il modello ad occuparsi selettivamente solo delle regioni rilevanti dell'input.

I loro esperimenti di valutazione dimostrano, inoltre, che *Face-Cutout* può migliorare con successo la variazione dei dati e alleviare il problema dell'*overfitting*, ottenendo una riduzione di *LogLoss* (la più importante metrica di classificazione basata sulle probabilità) dal 15,2% al 35,3% su diversi set di dati, rispetto ad altre tecniche basate sull'occlusione. Inoltre, Selim propone anche una linea guida generica per la pre-elaborazione dei dati per addestrare e valutare le architetture esistenti consentendo di migliorare la generalizzabilità di questi modelli per il rilevamento di deepfake (Das S., Seferbekov S., Datta A., Islam Md. S., Amin Md. R., 2021).

Un contributo importante per la rilevazione dei deepfake è stato quindi dato dalla *Deepfake Detection Challenge* che ha permesso di stimolare la ricerca di nuove forme di difesa dai deepfake. In questo caso il ruolo di un social in particolare è stato fondamentale: Facebook ha infatti stanziato più di 10 milioni di dollari per incoraggiare la partecipazione in questa sfida, tanto che il CTO di Facebook Mike Schroepfer ha affermato:

> "Le tecniche deepfake, che presentano video realistici generati dall'AI ritraenti persone reali che fanno e dicono cose fittizie, hanno significative implicazioni nel determinare la legittimità delle informazioni presentate online. Il settore non ha ancora un ampio set di dati con cui confrontarsi per poterli individuare. La speranza è di realizzare una tecnologia che tutti possano usare per meglio rilevare quando l'AI viene usata per alterare un video per ingannare lo spettatore" (Bai A., 2019).

La domanda da porsi ora è: qual è il ruolo dei social nei confronti di questo fenomeno? Attualmente, non sono molte le aziende social che combattono con forza il deepfake: Reddit e Pornhub hanno vietato il porno deepfake e altra pornografia che non prevede il consenso sei soggetti coinvolti, inoltre agiscono sulle segnalazioni da parte di utenti che reclamano questo tipo di materiale.

Facebook impedisce a qualsiasi contenuto identificato come falso o fuorviante di pubblicare annunci e guadagnare di conseguenza soldi, inoltre l'azienda collabora con oltre 50 organizzazioni di verifica dei fatti, accademici, esperti e responsabili politici per trovare nuove soluzioni (come abbiamo visto per la *Deepfake Detection Challenge*).

Anche gli algoritmi di Instagram cercano di limitare la visualizzazione di contenuti contrassegnati come "falsi" dai *fact checker* (Westerlund M., 2019). Nonostante ciò, sono ancora poche le piattaforme che hanno applicato politiche di contrasto al deepfake, infatti attualmente molto spesso diverse aziende non rimuovono i contenuti contestati dagli utenti o ritenuti fake, ma li rendono semplicemente più difficili da trovare, rendendoli meno visibili nei feed di notizie degli utenti.

Spesso questi contenuti non vengono completamente eliminati perché gli elementi fake che fanno clamore tra gli utenti raggiungono migliaia di visualizzazioni, e questo dal punto di vista finanziario risulta essere un successo per le aziende social, perché in questo modo si massimizza il tempo di coinvolgimento per la pubblicità. Per esempio, quando nel 2019 è girato in diversi social il video fake di Nancy Pelosi che sembrava apparire ubriaca, mentre piattaforme come Youtube e Twitter lo hanno rimosso per "violazione degli standard di correttezza", Facebook ha deciso di lasciare il video sulla piattaforma. Come ha affermato Monika Bicker, vice presidente di Facebook e responsabile per la tutela del prodotto e la difesa dalla propaganda in favore del terrorismo, gli utenti erano stati avvisati di trovarsi di fronte a un materiale fake. E ha aggiunto "non lo abbiamo

oscurato, perché pensiamo che le persone abbiano il diritto di vederlo e di farsi la propria opinione" (Sarcina G., 2019).

Questo ha ovviamente sollevato diverse critiche, per esempio il giornalista Anderson Cooper ha commentato "la forza delle immagini è più potente di qualsiasi cosa voi possiate scrivere a commento" (Sarcina G., 2019). Per questo motivo è importante che le aziende social collaborino per impedire che le loro piattaforme vengano utilizzate come armi per la disinformazione, applicando politiche trasparenti e accurate per bloccare e rimuovere i deepfake (Westerlund M., 2019).

Una soluzione potrebbe essere quella di introdurre un "*truth layer*", un sistema automatizzato su Internet che fornirebbe una rilevazione del livello di fake presente nel video rispetto a quello autentico, in modo tale che ogni video pubblicato su un social passi attraverso un processo di autenticazione (Westerlund M., 2019). Per esempio, il software incorporato nelle fotocamere degli smartphone potrebbe essere utilizzato per creare un'impronta digitale al momento della registrazione di un video, in modo tale che durante la riproduzione del filmato, la sua filigrana può essere confrontata con l'impronta digitale originale per verificare la corrispondenza e fornire all'utente un indice della probabilità di manomissione del video.

Effettivamente, la filigrana digitale può fornire a un video una breve serie di numeri che viene persa nel caso in cui il video sia manipolato. Questo potrebbe anche fornire una dimostrazione autenticata e reale per personaggi pubblici, dato che spesso sui social registrano video dove indicano il luogo in cui si trovano e cosa stanno facendo. Anche la tecnologia *Blockchain* può aiutare a verificare le origini e la distribuzione dei video: infatti essa si tratta di una "tecnologia basata su una catena di blocchi che registrano e gestiscono le operazioni contabili accessibili solo agli utenti di ciascun nodo, per assicurarne la tracciabilità" (Treccani, 2018).

Le piattaforme di social media dovrebbero quindi promuovere delle politiche aziendali che si occupino della verificabilità dei video e della rilevazione dei deepfake, dato che attualmente sono diverse le aziende che non si impegnano a gestire con maggiore serietà e approfondimento questo fenomeno, ma preferiscono sostenere i media falsi piuttosto che gestire una rapida rimozione dei contenuti segnalati dagli utenti sulle piattaforme.

Tuttavia, per raggiungere una vera soluzione è necessario innanzitutto comprendere in profondità il problema e la sua capacità di influenzarci: solo dopo averlo fatto diventa possibile sviluppare realmente delle soluzioni in grado di risolverlo. Inoltre, nessuna tecnologia, da sola, può combattere il deepfake e il "soluzionismo tecnologico" (che ogni problema ha una soluzione tecnologica), anzi, può persino portare la discussione riguardante questo fenomeno verso domande più esistenziali sul perché esistono i deepfake e quali altre minacce l'intelligenza artificiale può imporre. Il modo più efficace per combattere la diffusione e le conseguenze dei deepfake, quindi, implica un'applicazione di un misto di soluzioni legali, educative e sociotecniche (Westerlund M., 2019).

4.3 Il futuro che ci aspetta

In questa tesi è stata esplorata la storia e le origini del deepfake, i casi più eclatanti che hanno scosso la popolazione, le tecniche messe in campo per contrastarlo e le sue applicazioni positive o negative che siano. L'ultimo passo, ora, è quello di affrontare il futuro distopico a cui ci potrebbe condurre l'evoluzione di questo fenomeno.

Antonio Santangelo approfondisce questo tema nel libro intitolato "Il metavolto", in cui cerca di immaginare quale possa essere il futuro del volto nell'era dei deepfake. Secondo l'autore, oggi le tecnologie deepfake non sono ancora abbastanza sviluppate per non poterle riconoscere totalmente, in quanto presentano alcuni difetti che rendono questi video quasi "divertenti", dal momento in cui fanno pronunciare a soggetti parole che non direbbero mai (Santangelo A., 2022).

Tuttavia diversi studiosi, come Danielle Citron e Bobby Chesney (Schwartz O., 2018), delineando l'entità del potenziale pericolo, sostengono che oltre al problema dell'aumento della minaccia alla privacy, alla reputazione e alla sicurezza nazionale, una grande preoccupazione è anche legata a una distruzione catastrofica della fiducia sociale nei confronti dei mezzi di informazione all'interno di un clima politico che di per sé è già polarizzato (Schwartz O., 2018).

Inoltre, i due studiosi hanno previsto lo sfruttamento di deepfake da parte di fornitori di fake news: in pratica, chiunque sia abile ad utilizzare questa tecnologia, dai propagandisti approvati dallo stato ai *troll*, sarebbe in grado di distorcere le informazioni

che circolano in rete e nei media tradizionali, di manipolare le convinzioni dell'opinione pubblica e di spingere in questo modo le comunità online ideologicamente opposte nelle proprie realtà soggettive (Schwartz O., 2018).

Anche il comico Jordan Peel, come si è visto nel secondo capitolo, ha pubblicato un video deepfake di Barack Obama su BuzzFeed in cui definiva l'ex Presidente degli Stati Uniti Trump un "totale e completo imbecille", per aumentare la consapevolezza sulla futura potenzialità dell'intelligenza artificiale di manipolare e distorcere la realtà. Riguardo a ciò, tre membri del Congresso americano hanno inviato una lettera al direttore dell'intelligence nazionale, sottolineando quanto i deepfake possano incrementare le campagne di disinformazione durante le elezioni, attraverso la diffusione di video deepfake che coinvolgono il volto di politici che fanno e dicono cose che nella realtà non direbbero mai (Schwartz O., 2018).

A questo proposito, Citron afferma:

> "Quello che mi tiene sveglio la notte è uno scenario ipotetico in cui, prima del voto in Texas, qualcuno pubblica un falso profondo di Beto O' Rourke che fa sesso con una prostituta, o qualcosa del genere. Ora, so che questo sarebbe facilmente confutabile, ma se questo dovesse cadere la notte prima, non puoi sfatarlo prima che si siano diffusi gravi danni" (Schwartz O., 2018).

Inoltre, aggiunge: "Sto iniziando a vedere come un falso profondo tempestivo potrebbe benissimo interrompere il processo democratico" (Schwartz O., 2018). Altri studiosi come Tim Hwang, tuttavia, sostengono che questo fenomeno debba fare ancora molta strada prima di diventare un vero problema. Hwang ha studiato per diversi anni la diffusione delle fake news e della disinformazione e, a parte alcune eccezioni, non ha riscontrato incidenti eclatanti. Nonostante ciò, sostiene che probabilmente in un futuro in cui i deepfake diventeranno sempre più realistici e facili da produrre, si creerà una "perfetta tempesta di disinformazione".

Quindi dal punto di vista politico e dell'informazione si prevede una disinformazione sempre più diffusa, che porterà le persone a non credere più non solo a ciò che leggono, ma anche a ciò che vedono e sentono.

Secondo diversi studiosi, sotto diversi punti di vista, la tecnologia deepfake crescerà sempre di più e verrà perfezionata per essere applicata in diversi aspetti della

nostra vita, sia in modo positivo, sia in modo negativo. Il problema, però, è che se oggi siamo ancora abbastanza in grado di distinguere dei video deepfake in base a una serie di fattori visibili (meccanicità del movimento corporeo, artificialità di luci e colori, anomalo battito di ciglia, labiale mal sincronizzato, pelle eccessivamente liscia, ecc.) questo porterà a creare delle versioni totalmente false che saranno completamente indistinguibili a occhio nudo.

Antonio Santangelo, per esempio, immagina che in un futuro saranno creati per tutti dei meta-volti digitali falsi che verranno rappresentati virtualmente e che potranno potenzialmente farci perdere il controllo sulla diffusione della nostra immagine e parola, mostrandoci in contesti che non ci appartengono (Santangelo A., 2022).

La preoccupazione principale è che nel futuro, scrollando sulla *home page* dei social, ci troveremo di fronte a video che vedono come protagonisti *opinion leader*, politici, personaggi famosi o addirittura noi stessi che diciamo e facciamo cose che nessuno sarà più in grado di valutare come vere o false, e questo potrebbe portare persone ostili ad alimentare il dubbio confondendo e influenzando l'opinione pubblica, producendo sfiducia nei confronti delle istituzioni e delle fonti ufficiali di informazione, danneggiando la reputazione e la privacy degli individui, minacciando e ricattando le persone, mettendo a rischio la sicurezza pubblica e alimentando la disinformazione attraverso la pubblicazione e la diffusione di video deepfake.

Tuttavia, l'evoluzione del deepfake potrà essere utilizzata anche per scopi più positivi: Christian Theobalt, per esempio, immagina che in un futuro questa tecnologia sarà utilizzata in modo più efficace per applicare dei doppiaggi più accurati e sincronizzati nei film stranieri, per utilizzare tecniche avanzate di *editing* facciale per la post-produzione nei film e per incrementare gli effetti speciali (Schwartz O., 2018).

Questo di conseguenza porterà a sintetizzare volti che sembreranno quasi indistinguibili da quelli reali, il che potrebbe costituire una evidente differenza per l'industria dell'intrattenimento cinematografico e visivo. Santangelo, inoltre, sostiene che in ambito politico il deepfake potrà essere utilizzato dai candidati che non hanno il tempo necessario per rivolgersi al popolo o che non conoscono la lingua di coloro a cui si rivolgono (come i migranti), per creare degli "alter ego digitali" che si occupino dei loro discorsi in base ai valori da essi sostenuti (Santangelo A., 2022).

Un'altra applicazione potrebbe essere invece di stampo tanatologico, permettendo di produrre delle versioni digitali delle persone care decedute con cui poter sostenere nuovamente una conversazione attraverso un video deepfake, rendendo così reale la puntata di *Black Mirror* di cui si è parlato nel secondo capitolo.

Oppure, si potrebbe immaginare di creare un avatar che imiti perfettamente il nostro volto e la nostra voce, dicendo o facendo pubblicamente cose che noi vorremmo dire o fare, ma che non ci sentiamo in grado di attuare direttamente, magari per timidezza o per motivi più profondi.

Un giorno magari, in campo medico, sarà possibile inoltre restituire la voce a chi l'ha persa per un intervento attraverso l'apprendimento automatico del *deep learning*, che potrebbe essere addestrato per studiare la nostra voce in una precedente registrazione audio per poi restituircela attraverso dei particolari macchinari. Oppure potrebbe dare una nuova voce a chi non l'ha mai avuta, attraverso lo stesso procedimento ma assunto da una registrazione audio di qualcun altro. In questo modo le conseguenze del deepfake verrebbero apportate in modo benefico anche sotto il punto di vista psicologico.

In conclusione, in questo capitolo sono stati osservati i mezzi principali attraverso cui combattere il deepfake, che vanno da una legislazione improntata sul fenomeno a una tecnologia anti-deepfake in grado di rilevare eventuali manomissioni dei video. Abbiamo inoltre esplorato delle nuove tecniche che potrebbero essere funzionali per il contrasto al deepfake ma che non sono ancora state messe in atto. Risulta quindi che questo fenomeno non sia ancora ritenuto così rilevante da farne un problema serio su cui applicare una legislazione specifica e le tecnologie anti-deepfake non sono ancora abbastanza sviluppate per riconoscere video manipolati.

Inoltre, vi è una scarsa istruzione riguardo al fenomeno e i social non si mobilitano a sufficienza per difendere gli utenti dall'intrusione dei video deepfake, alimentando così la sfiducia sociale. Infine, si è cercato di immaginare quale sarà il futuro che ci aspetta, sia negativamente che positivamente. Attualmente i deepfake presentano una minaccia molto grave soprattutto dal punto di vista della disinformazione e della privacy, ma in futuro potranno diventare problematici anche dal punto di vista politico, soprattutto in vista delle elezioni. Appare quindi chiaro che sia necessario un intervento tempestivo da parte delle istituzioni, delle politiche aziendali e dei ricercatori in modo da non giungere

al punto in cui neanche la tecnologia stessa sarà in grado di distinguere il vero dal falso, non potendo così più tornare indietro.

Conclusioni

Attraverso la stesura di questa tesi, si è cercato di descrivere al meglio questo fenomeno ancora poco conosciuto e che prenderà sempre più piede nella nostra società.

Fin dal primo capitolo è stato analizzato l'argomento attraverso l'approfondimento delle fake news, che sono ormai una minaccia reale e intrinseca della nostra società. Per analizzare l'effetto e il comportamento delle persone nei loro confronti, si è partiti dalla teoria di Daniel Kahneman e dai suoi studi sui processi decisionali che hanno portato allo sviluppo dell'idea secondo cui gli esseri umani sono dotati di due sistemi mentali che influenzano il nostro comportamento nei confronti delle notizie false.

In seguito sono state descritte le origini del deepfake, che vanno dalla prima pubblicazione di un video falso su Reddit alla sua applicazione nell'industria cinematografica. Sono state inoltre esplorate le caratteristiche principali del fenomeno che viene generato attraverso le GAN, reti generative contradditore, introdotte dal ricercatore Ian Goodfellow, per poi passare dall'approfondimento del *face swapping*, ovvero la modifica del volto nei video deepfake, a quello dell'*audio editing*, ovvero l'imitazione di una voce che avviene attraverso dei particolari strumenti chiamati sintetizzatori. A questo proposito si è fatto riferimento all'ipotesi della *Uncanny Valley* esposta da Masahiro Mori, secondo cui la sensazione di familiarità nei confronti dei robot aumenta in base alla loro somiglianza con la figura umana, fino a raggiungere un punto in cui, però, si riducono le reazioni emotive positive e si cade nella valle perturbante (*uncanny valley*) nella quale si generano emozioni inquietanti e distaccate. Per completare il primo capitolo, si è anche cercato di analizzare gli autori e le motivazioni generali che stanno dietro alla creazione dei deepfake, che vanno da semplici ragioni ludiche alla truffa.

Successivamente nel secondo capitolo sono stati esplorati gli aspetti positivi e negativi del fenomeno in maniera approfondita, con l'obiettivo di comprendere al meglio le sue funzionalità e implicazioni. Sono stati riscontrate diverse conseguenze dannose,

come la diffusione persistente della disinformazione, l'infodemia, il furto d'identità e della privacy, il *cyberbullismo* e il *deepnude*. Tuttavia, vi sono anche degli utilizzi etici legati al fenomeno, come per esempio nel campo dell'industria cinematografica, dei giochi multiplayer, del commercio e della moda.

Inoltre, attraverso l'analisi del documentario "*Welcome to Chechnya*", si è osservato come il deepfake può essere utilizzato paradossalmente per tutelare la privacy delle persone che necessitano per cause di forza maggiore di nascondere la propria identità. Infine, è stano analizzato un fenomeno ambiguo, incollocabile tra gli aspetti positivi o negativi, in quanto solleva giudizi di natura diversa: la Deep nostalgia. Al termine dell'analisi di questi aspetti, si può quindi affermare che è vero che il deepfake può aiutare alcuni settori attualmente in prevalenza commerciali, come quello cinematografico o della moda, tuttavia questo non toglie che il rapporto rischi - benefici è decisamente sbilanciato verso i rischi, per tale motivo si ritiene necessario introdurre delle legislazioni efficaci e un'utile istruzione, assieme a una accurata tecnologia anti-deepfake e a delle politiche aziendali che vengono analizzate più approfonditamente nel quarto capitolo.

Il terzo capitolo è stato dedicato al questionario che è stato somministrato a 160 partecipanti. L'obiettivo preposto era quello di intercettare la tendenza dei rispondenti nel riconoscere i video deepfake. In seguito a una serie di domande generali, sono quindi stati inseriti cinque video manipolati e due reali. L'aspettativa era che alcuni dei rispondenti non riuscissero a riconoscere la manipolazione dei video fake, dimostrando così che per alcuni sia pericolosa l'esposizione al fenomeno.

I risultati, però, sono stati decisamente sorprendenti, in quanto in diversi video la maggior parte delle persone non ha riconosciuto la manipolazione (raggiungendo fino al 76,3% di errore). Inatteso anche il non riconoscimento da parte di una buona percentuale dei rispondenti dei video reali: questo dimostra che la diffusione dei video deepfake provoca confusione e incertezza rischiando di ridurre la credibilità nei confronti di qualsiasi fonte di informazione. Nonostante il questionario sia quindi rappresentativo di una piccola percentuale di persone, si è rivelato comunque molto utile per mostrare l'impatto che il fenomeno può avere sulla nostra società.

L'ultimo capitolo si è concentrato sulle soluzioni possibili per combattere gli usi dannosi del deepfake. Sono stati analizzati quattro metodi principali: una legislazione

specifica che protegga dal deepfake, una maggiore istruzione riguardo al fenomeno, una tecnologia anti-deepfake più accurata e le politiche aziendali dei social media, che risultano essere i maggiori mezzi di veicolazione dei video manipolati. Da questa ricerca si è potuto comprendere che non solo questi aspetti vanno migliorati, ma che vanno anche applicati tutti assieme affinché siano efficaci.

Si è infine cercato di immaginare quale sia un eventuale futuro che ci aspetta attraverso gli studi di alcuni ricercatori: questo fenomeno crescerà sempre di più e verrà applicato in settori sempre più diversi. Si dovrà quindi cercare di implementarlo maggiormente in ambiti in cui può rivelarsi utile, da quello medico a quello cinematografico, ma continuare a limitarlo nei casi in cui rischia di diffondere disinformazione (soprattutto dal punto di vista politico) o di danneggiare la privacy e la reputazione delle persone.

In conclusione, questa tesi ha permesso di esplorare le caratteristiche intrinseche di questa tecnologia attualmente ancora poco conosciuta e di capire l'impatto che può avere sulla nostra società. Si è cercato di descriverle a partire da alcuni studi che fino ad oggi si sono impegnati nella ricerca e nella comprensione di questo fenomeno, nonostante sia ancora molta la strada da fare per comprenderlo più a fondo. L'auspicio è che questa tesi si riveli utile per esserne maggiormente a conoscenza e per capirlo più approfonditamente, nonché per fornire uno spunto per eventuali riflessioni future.

Bibliografia

Ajder H., Patrini G., Cavalli F., Cullen L., 2019, *The state of deepfakes, landscape, threats, and impact*, Deeptrace, settembre 2019, pp. 1-27.

Arruzzoli F., 2019, *Perché il deep fake preoccupa l'intelligence? Disinformazione e attacchi psicologici con l'uso illecito dell'IA*, ICT Security Magazine, 3 giugno 2019, https://www.ictsecuritymagazine.com/articoli/perche-il-deep-fake-preoccupa-lintelligence-disinformazione-e-attacchi-psicologici-con-luso-illecito-dellia/.

Barbero N., 2021, *Deep Nostalgia: Deepfakes con Fines Nostálgicos. Riesgos de la IA Desde la Teoría del Bildakt*, Nuevos enfoques y perspectivas de la Antropología Visual, I quadrimestre 2021.

Bredekamp H., 2015, *Immagini che ci guardano. Teoria dell'atto iconico*, Raffaello Cortina Editore, Milano.

Brighi C., Chiara P.G., 2021, *La cybersecurity come bene pubblico: alcune riflessioni normative a partire dai recenti sviluppi nel diritto dell'Unione Europea*, Federalismi.it, 21, 8 settembre 2021, pp. 18-42, http://hdl.handle.net/10993/48032.

Cherchi F., 2022, *Deepfake: la "uncanny valley" di Mori e il perturbante*, State of Mind, 21 aprile 2022, https://www.stateofmind.it/2022/04/uncanny-valley-deepfake/.

Cosentino G., 2017, *L'era della post-verità, media e populismi dalla Brexit a Trump*, Rizzoli Libri, Reggio Emilia.

Das S., Seferbekov S., Datta A., Islam Md. S., Amin Md. R., 2021, *Towards Solving the DeepFake Problem: An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation*, arXiv:2102.09603.

Güera D., e Delp E., 2018, *Deepfake Video Detection Using Recurrent Neural Networks*, 15th IEEE International Conference on Advanced Video and Signal Based Surveillance, 27-30 novembre 2018, pp. 1-6.

Esiodo, 2016, *Le opere e i giorni*, Raffaelli Editore, Rimini.

Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2020, *Generative adversarial networks*, ACM, vol. 63, pp. 139–144.

Grandis N., 2021, *Voice cloning, cos'è, come funziona, prospettive e problematiche*, AI4Business, 23 novembre 2022, https://www.ai4business.it/intelligenza-artificiale/voice-cloning-cose-come-funziona-prospettive-e-problematiche/.

Iozzia G., 2022, *Visual Deepfakes: come vengono generati e come possono essere identificati*. AI4Business, 22 novembre 2022, https://www.ai4business.it/intelligenza-artificiale/visual-deepfakes-come-vengono-generati-e-come-possono-essere-identificati/.

Kahneman D., 2012, *Pensieri lenti e veloci*, Mondadori, Milano.

Korshunov P., Marcel S., 2020, *Deepfake detection: humans vs. machines*, arXiv:2009.03155.

Moravec P., Randall K. Minas and Alan R. Dennis, 2018, *Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense at All*, Kelley School of Business research paper, gennaio 2018, pp. 8-87, doi:10.2139/ssrn.3269541.

Mori M., MacDorman KF., e Kageki N., 2012, *The Uncanny Valley*, IEEE Robotics & Automation Magazine, vol. 19, n. 2, giugno 2012, pp. 98-100, doi: 10.1109/MRA.2012.2192811.

Pan American Health Organization, 2020, *Understanding the Infodemic and Misinformation in the fight against COVID-19*, Washington, pp 2-4.

Peters M., 2017, *Education in a post-truth world. Educational Philosophy and Theory*, Routledge, Taylor & Francis Group, vol. 49:6, pp. 563-566, doi: 10.1080/00131857.2016.1264114.

Pinotti A., Somaini A., 2016, *Cultura visuale. Immagini, sguardi, media, dispositivi*, Einaudi, Torino.

Piras A., 2020, *Fake news e nuove tecnologie: la Blockchain può realmente essere la nuova frontiera della lotta alla disinformazione in rete?,* ANDIG.it, 10 luglio 2020, https://www.andig.it/saggi/fake-news-e-nuove-tecnologie-la-blockchain-puo-realmente-essere-la-nuova-frontiera-della-lotta-alla-disinformazione-in-rete.

Quattrociocchi W. e Vicini A., 2016, *Misinformation: guida alla società dell'informazione e della credulità*, FrancoAngeli, Milano.

Rehak B., 2011, *Computer-Generated Imagery,* Cinema And Media Studies, doi: 10.1093/obo/9780199791286-0068.

Santangelo A., 2022, *Il futuro del volto nell'era del deepfake,* in Leone M. (a cura di), *Il metavolto*, FACETS Digital Press, Torino, pp 6-38.

Sarcina G., 2019, *Il video di Nancy Pelosi ubriaca: falso, ma Facebook non lo blocca*, Corriere della Sera, 25 maggio 2019, https://www.corriere.it/video-

articoli/2019/05/25/video-nancy-pelosi-ubriaca-falso-ma-facebook-non-blocca/c82b0940-7ed1-11e9-a444-6e83400b8609.shtml.

Saikia P., Dholaria D., Yadav P., Patel V., Roy M., 2022, *A Hybrid CNN-LSTM model for Video Deepfake Detection by Leveraging Optical Flow Features*, arXiv:2208.00788.

Schwartz O., 2018, *You thought fake news was bad? Deep fakes are where truth goes to die*, The Guardian, 12 novembre 2018, https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth.

Tonioni F., 2014, *Cyberbullismo, come aiutare le vittime e i persecutori*, Mondadori, Milano.

Vaccari C., Chadwick A., 2020, *'Deepfakes' are here. These deceptive videos erode trust in all news media*, The Washington Post, 28 maggio 2020, https://www.washingtonpost.com/politics/2020/05/28/deepfakes-are-here-these-deceptive-videos-erode-trust-all-news-media/.

Westerlund M., 2019, *The Emergence of Deepfake Technology: A Review*, Technology Innovation Management Review, vol. 9, cap. 40, pp 40-53.

Zanon Giusto M., 2021, *La voce surreale di Anthony Bourdain*, Il Foglio, 31 luglio 2021, https://www.ilfoglio.it/cultura/2021/07/31/video/la-voce-surreale-di-anthony-bourdain-2727279/.

Sitografia

Bai A., *Deepfake Detection Challenge, i big della tecnologia per supportare lo sviluppo di tecnologie anti-deepfake*, https://www.hwupgrade.it/news/web/deepfake-detection-challenge-i-big-della-tecnologia-per-supportare-lo-sviluppo-di-tecnologie-anti-deepfake_86041.html, consultato il 02/11/2022.

Carboni K., *Il video deepfake in cui Zelensky annuncia la resa è stato rimosso*, https://www.wired.it/article/guerra-russia-ucraina-video-deepfake-zelensky-resa/, consultato il 17/10/2022.

Coppola J., *TRY IT ON: una nuova concezione del camerino virtuale*, https://www.sautechgroup.com/try-it-on-una-nuova-concezione-del-camerino-virtuale/, consultato il 20/10/2022.

Cosimi S., @Roman, *la chatbot del defunto è una noia mortale*, https://www.wired.it/attualita/tech/2016/10/10/ho-chattato-col-bot-di-un-morto/#:~:text=La%20startupper%20russa%20Eugenia%20Kuyda,limitata%20e%20a%20tratti%20senza%20logica., consultato il 21/10/2022.

Digital Guide Ionos, *Deepfake: la contraffazione della prossima generazione*, https://www.ionos.it/digitalguide/online-marketing/social-media/deepfakes/#:~:text=A%20partire%20da%20questi%20dati,rigenerarla%20autonomamente%2C%20anche%20in%20movimento, consultato il 01/10/2022.

Enciclopedia Treccani, https://www.treccani.it/enciclopedia/, consultato il 23/11/2022.

Eud International Foundation C.I.C., *Fake News: cosa sono e perchè è importante combatterle*, https://eudfoundation.it/fake-news-cosa-sono-e-combatterle/, consultato il 01/10/2022.

Garante per la protezione dei dati personali, *Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9512278, consultato il 06/10/2022.

Kaggle, *Deepfake Detection Challenge*, https://www.kaggle.com/c/deepfake-detection-challenge, consultato il 03/11/2022.

Liceo don Carlo la Mura, *Disposizioni in materia di contrasto ai deepfakes*, https://www.senatoragazzi.it/iniziative/disegno-di-legge/103/, consultato il 31/10/2022.

Mattarella F., *L'era della post-verità: un cambiamento socio-culturale pericoloso*, https://www.pensierocritico.eu/l-era-della-post-verita.html, consultato il 01/10/2022.

My Heritage, https://www.myheritage.it/deep-nostalgia, consultato il 21/10/2022.

Oxford Languages, https://languages.oup.com/, consultato il 23/11/2022.

Oxford Learner's Dictionary, https://www.oxfordlearnersdictionaries.com/definition/english/post-truth, consultato il 23/11/2022.

Rutelli F., Galoppi C., Jacona A., Ciardi N., Morcellini M., Giorgino F., Manzella G.P., Passarelli P., Sesta A., De Nucci J., *La minaccia del Deepfake: non basta più vedere per credere*, https://www.radioradicale.it/scheda/588781/la-minaccia-del-deepfake-non-basta-piu-vedere-per-credere, consultato il 31/10/2022.

Scarselli L., *Rogue One: A Star Wars Story: Peter Cushing e Carrie Fisher ricreati in digitale per il film*, https://movieplayer.it/news/rogue-one-a-star-wars-story-peter-cushing-carrie-fisher_80863/, consultato il 17/11/2022.

Sipucin, *Cosa sono i bias?*, https://www.noemahr.com/cosa-sono-i-bias/, consultato il 06/10/2022.

Torrini F., *I rischi e i pericoli del deepfake: tutto quello che devi sapere*, https://universeit.blog/deepfake/, consultato il 01/10/2022.

Viviani M., *Deepnude e le foto bufala: immagini di donne nude mai scattate da loro*, https://www.iene.mediaset.it/2019/news/deepnude-foto-donne-nude_578538.shtml, consultato il 17/10/2022.

World Economic Forum, *Rischi globali 2013*, Ottava edizione, https://reports.weforum.org/global-risks-2013/, consultato il 06/10/2022.

Ringraziamenti

Vorrei dedicare uno spazio a chi ha contribuito con pazienza alla conclusione di questo splendido percorso.

Ringrazio il mio relatore, Bruno Mastroianni, che grazie alle sue utili indicazioni mi aiutato e spronato nella realizzazione di questo elaborato.

Ringrazio mia sorella, Rossella, che c'è sempre stata e ha sempre speso dei buoni consigli per me per aiutarmi a concludere in maniera positiva la mia carriera universitaria.

Ringrazio Marco, che ha saputo sempre strapparmi un sorriso e che mi è sempre stato accanto, facendomi vivere con più leggerezza e positività questo percorso.

Infine, non posso che ringraziare di cuore i miei genitori, Giuseppe e Laura.

Grazie di cuore papà, per gli interi pomeriggi passati a interrogarmi, per la pazienza con cui mi hai ascoltato per ore, per le partite di calcio che hai dovuto seguire in silenzio per lasciarmi studiare, per avermi supportato sempre e per ogni parola che hai speso per tranquillizzarmi e spronarmi prima di un esame. Senza di te non sarebbe stato lo stesso.

Grazie di cuore mamma, per aver esultato con me ad ogni valutazione, per aver creduto in me ogni giorno, per aver atteso ogni mio risultato come se fosse tuo, per essermi stata sempre vicina e per avermi incoraggiato dall'inizio alla fine. Senza di te non sarei riuscita a concludere questo percorso nei migliori dei modi.

Il raggiungimento di questo traguardo è anche vostro.

Grazie di cuore.