



Dottorato di Ricerca in Visual Studies Ciclo XXXV
Curriculum in Film and Media Studies

IT'S NOT ME, IT'S A DEEPPFAKE

Un percorso d'orientamento tra i media manipolati attraverso
l'intelligenza artificiale

Lafiandra Pietro

1012115

Tutor: Prof. Gian Battista Canova

Cotutore: Prof. Pietro Montani

Coordinatore: Vincenzo Trione

ANNO ACCADEMICO 2021/2022

A Marcello

Indice

Nota metodologica.....	1
Deepfake: un framework teorico assente.....	3
<i>Grafia</i>	5
<i>Deep learning</i>	8
Per una tassonomia dei <i>deepfake</i>.....	16
<i>Sommario</i>	16
<i>Immagini, audio, testi, video</i>	17
Immagini deepfake.....	18
DeepNude.....	19
Questecosenonesistono.....	22
Audio deepfake.....	26
Testi deepfake.....	28
GPT-3.....	30
Video deepfake.....	35
<i>Cheapfake vs deepfake</i>	41
Cheapfake (shallowfake).....	41
Deepfake.....	51
GAN.....	55
DeepFaceLab.....	58
<i>Conclusioni capitolo</i>	63
<i>Tornare alla mimesis</i>.....	65

<i>Sommario</i>	65
<i>Doppi, gemelli, cloni</i>	65
Avatar	65
Corpi digitali	68
Clonophobia	71
<i>Mimesis: un framework teorico</i>	74
GAN e mimesis	74
Mimesis come indebolimento ontologico di un originale	77
Mimesis come raddoppiamento chiarificante	86
Mimesis come “incremento d’essere”	91
<i>Conclusioni capitolo</i>	97
Studi di caso	99
<i>The Eternal feat. Ian Curtis: deepfake come indebolimento ontologico di un originale</i>	99
<i>Marilyn Monroe gokkun: deepfake come raddoppiamento chiarificante</i>	114
<i>The Andy Warhol Diaries: deepfake come “incremento d’essere”</i>	133
Riferimenti bibliografici	151
Ringraziamenti	168

Nota metodologica

Considerata la frammentarietà dell'argomento trattato, l'approccio adottato nella seguente tesi è stato di tipo multidisciplinare, nel tentativo di mettere in dialogo la letteratura a oggi esistente attorno ai *deepfake* con la teoria del cinema, i *porn studies*, i *celebrity studies*, la filosofia dell'immagine e gli studi nel campo dell'intelligenza artificiale, cercando di delineare traiettorie e possibili percorsi di una pratica ancora in via di sviluppo. I testi di riferimento sono stati in particolar modo *Cloning Terror. The War of Images, 9/11 to the Present* (2011) di W.J.T Mitchell, *L'estetica della mimesis* (2009) di Stephen Halliwell, *Deepfakes and the Infocalypse* (2019) di Nina Schick, *Trust No One: Inside the World of Deepfakes* (2022) di Michael Grothaus, *Arte e intelligenza artificiale. Be my GAN* (2020) a cura di Alice Barale e *Software Takes Command* (2013) di Lev Manovich. L'aggiornamento sui video *deepfake* è stato costante e accompagnato dal dialogo con noti *deepfake designer* italiani come Gerardo Bosso e Salvatore Fortunato, oltre che dalle interviste fornitemi da Ctrl Shift Face e Francesco Angeli, produttore dei video *deepfake* per Striscia la Notizia. Alla ricerca teorica si è inoltre unita la sperimentazione diretta nella produzione dei video *deepfake*, in particolar modo con il video *The Eternal feat. Ian Curtis*, il quale sarà trattato come studio di caso nel capitolo conclusivo della tesi.

Deepfake: un framework teorico assente

Nella letteratura dedicata i *deepfake* sono stati definiti come “video sintetici che assomigliano molto a video reali”¹, “video iper-realistici nei quali il volto di una persona è stato analizzato da un algoritmo di *deep learning* e successivamente sovrapposto sul volto di un attore in un video”², “video che sovrappongono il volto di una persona sul corpo di un’altra [...] resi possibili da algoritmi gratuiti e facili da usare”³, “video manipolati, o altre rappresentazioni digitali prodotte da una sofisticata intelligenza artificiale, che produce suoni e immagini finzionali che appaiono come reali”⁴, “un tipo di media sintetici, ovvero media (che includono immagini, audio e video) che sono o manipolati o completamente generati attraverso l’intelligenza artificiale”⁵, “video falsificati attraverso l’intelligenza artificiale che presentano nuove strade multi-sensoriali per manipolare la percezione della realtà”⁶, “una categoria di operazioni di *face-swapping* che

¹ C. Vaccari, A. Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, in “Social Media + Society”, 2020, p.1 (traduzione mia).

² C. Öhman, *Introducing the pervert’s dilemma: a contribution to the critique of Deepfake pornography*, in “Ethics and Information Technology”, n. 22, 2020, cit. p. 133 (traduzione mia).

³ V. C. Martinez, G.P. Castillo, *Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso*, in “Historia y comunicación social”, n. 24(2), cit. p. 505 (traduzione mia).

⁴ A. Karandikar, V. Deshpande, *Deepfake Video Detection Using Convolutional Neural Network*, in “International Journal of Advanced Trends in Computer Science and Engineering”, Vol. 9, 2020, n.2, cit. p. 1311 (traduzione mia).

⁵ N. Schick, *Deep Fakes and the Infocalypse: What You Urgently Need to Know*, Octopus Publishing Group, Londra 2020, cit. p.8 (traduzione mia).

⁶ S. Woolley, *The Reality Game: How the next wave of technology will break the truth and what we can do about it*, Octopus Publishing Group, Londra 2020, cit. p.104 (traduzione mia).

sfrutta modelli di *machine learning* come *autoencoder* o *generative adversarial network*"⁷ o come "il prodotto di applicazioni di intelligenza artificiale o di *machine learning* che fondono, mescolano, rimpiazzano o sovrappongono immagini e clip video all'interno di un video, creando un video falso che appare come autentico"⁸.

Pur facendo quasi tutte riferimento all'ambito dell'intelligenza artificiale, le definizioni sopracitate (solo una collezione esemplare delle possibili definizioni fornite) propongono una terminologia di volta in volta differente in merito agli strumenti di produzione dei *deepfake* (*autoencoder*, *machine learning*, *deep learning*, *intelligenza artificiale*, *generative adversarial network*, *face-swapping*) e fanno più volte esplicito riferimento a media audiovisivi, ma trovando anche definizioni contraddittorie o che ne ampliano la sfera di riferimento.

Le difficoltà nel definire i *deepfake* in maniera univoca sono state evidenziate da Nina Schick in *Deep Fakes and the Infocalypse: What You Urgently Need to Know*, primo saggio interamente dedicato ai *deepfake* in cui l'autrice scrive:

Ho visto il termine *deepfake* usato per indicare tutti i media sintetici, per indicare i *face-swap*, e per indicare i *face-swap* nei video pornografici. [...] Nonostante ci siano dozzine di persone e organizzazioni all'interno di questo spazio, non c'è ancora un framework

⁷ C. Yang, L. Ding, *Defending against GAN-based Deepfake Attacks via Transformation-aware Adversarial Faces*, arXiv, 2020.

⁸ M. Maras, A. Alexandrou, *Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos*, in "The International Journal of Evidence & Proof", Vol. 23 (3), p. 255 2018.

concettuale condiviso e una tassonomia a cui fare riferimento, specialmente in relazione ai *deepfake*⁹.

Le parole di Schick evidenziano la necessità di delineare un percorso d'orientamento all'interno del dispersivo mondo dei *deepfake*, partendo dal presupposto che il loro sviluppo sia ancora *in fieri* e che, nel delineare direzioni, strade e percorsi intrapresi, si dovrà astenersi da ogni tipo di sintesi poiché, vista la recente nascita dell'argomento trattato, nel periodo di ricerca affrontato alcuni degli assunti dati per assodati sono stati presto smentiti e le possibili direzioni di questa categoria di media manipolati sono numerose e tuttora incerte.

Questa considerazione non ci dispensa però dal provare a tracciare quelle che si pensa possano essere alcune chiare indicazioni su cosa i *deepfake* siano, su quale sia la tecnologia che li rende possibili e, sotto l'invito delle parole di Schick, dal tracciare una prima tassonomia che perimetri il territorio di movimento, nella speranza di fornire un *vademecum*, una bussola per l'orientamento che possa restituire utili indicazioni per future ricerche.

Grafia

Nelle prime pubblicazioni sul tema risalenti al periodo 2018/2020, sembra esserci poca chiarezza attorno alla grafia del vocabolo. Non è

⁹ N. Schick, *op.cit.*, p.91 (traduzione mia).

insolito trovare la forma *deep fake*¹⁰¹¹¹² o, più raramente, *deep-fake* (o *Deep-Fake*)¹³ e *DeepFake*¹⁴.

Samantha Cole, redattrice di Vice e pioniera della ricerca sui *deepfake*, nel suo articolo *For the Love of God, Not Everything is a Deepfake*, lamentava l'utilizzo del termine scritto come "*deep fake*" da parte del collega David Frum in un articolo apparso sul The Atlantic¹⁵.

Ironicamente, quello che non è in grado di riconoscere il vero è Frum, che chiama un'immagine manipolata su una stupida app un *deepfake* (e scrive anche correttamente "*deep fake*", attraverso due parole). Vale la pena essere pedanti, e differenziare tra *deepfake* e altri media manipolati. I *deepfake* con scopi calunniosi sono fatti con lo scopo di ingannare - una distinzione che molte piattaforme social mainstream fanno ora nei loro termini di utilizzo - ma sono anche prodotti attraverso l'uso di algoritmi e machine learning, non burattinati da un dito su un iphone¹⁶.

¹⁰ H. Smith, K. Mansted, *Weaponised Deep Fakes: National Security and Democracy*, 2020, <https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-04/Weaponised%20deep%20fakes.pdf?lgwT9eN66cRbWTovhN74WI2z4zO4zJ5H> (ultima consultazione il 14 febbraio 2023) (traduzione mia).

¹¹ G. Barcellona, *Impressions, l'app per creare deep fake credibili. Un pericoloso salto di qualità per i falsi*, 2020, https://www.repubblica.it/tecnologia/social-network/2020/04/26/news/impressions_app_deep_fake-254939338/?refresh_ce (ultima consultazione il 14 febbraio 2023) (traduzione mia).

¹² R. Eugeni. *Capitale algoritmico: cinque dispositivi postmediali (più uno)*, Scholè, Brescia 2021.

¹³ S. Agarwal, T. El-Gaaly, H. Farid, S. Lim, *Detectin Deep-Fake Videos From Phoneme-Viseme Mismatches*, in "*IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*", pp. 660-661 2020.

¹⁴ P. Korshunov, S. Marcel, *DeepFakes: a New Threat to Face Recognition? Assessment and Detection*, arXiv:1812.08685v1 2018.

¹⁵ D. Frum, *The Very Real Threat of Trump's Deepfake*, 2020, <https://www.theatlantic.com/ideas/archive/2020/04/trumps-first-deepfake/610750/> (ultima consultazione il 14 febbraio 2023) (traduzione mia).

¹⁶ S. Cole, *For the Love of God, Not Everything Is a Deepfake*, 2020, <https://www.vice.com/en/article/7kzgg9/joe-biden-tongue-gif-twitter-deepfake> (ultima consultazione il 14 febbraio 2023) (traduzione mia).

Non entrando nel merito della problematica descrizione attuata da Cole di *malicious deepfake* come di “media prodotti a scopo di inganno” (il che implicherebbe un’analisi dei processi percettivi dello spettatore del *deepfake*¹⁷), ciò che qui preme sottolineare è come il corretto uso del termine implichi una stretta correlazione tra medium prodotto, tra l’output, e ricorso all’intelligenza artificiale, o, più precisamente, al *machine learning*.

La grafia corretta del vocabolo è, infatti, *deepfake* dal momento che il termine è un portmanteau, un neologismo sincratico derivante da “*deep learning*” e “*fake*”¹⁸, adottato dal nome dell’utente di Reddit (*deepfakes*) che il 2 novembre 2017 creò sul social il forum di discussione “r\deepfakes”, dedicato a video pornografici di celebrità hollywoodiane che il redditor aveva prodotto attraverso l’intelligenza artificiale¹⁹.

Si ritiene pertanto cruciale, e non a puro scopo di pedanteria, sottolineare come il termine *deepfake* sarebbe da intendersi non in qualità di “falso profondo” ma, piuttosto, come di “falso prodotto attraverso il *deep learning*”. Queste due declinazioni differenti comporterebbero dei distinguo nelle accezioni dei prodotti: intenderli come *deep fake* implicherebbe l’attribuzione al medium di un’aprioristica capacità di ingannare lo spettatore, il che appiattirebbe la definizione di *deepfake* al suo

¹⁷ Si ritiene comunque corretto sottolineare come l’attribuzione qualitativa del termine *malicious* abbia il merito di ricordare come non sia connaturato ai *deepfake* il tentativo di mistificare la realtà, ma come il suo valore sia dettato da differenti fattori: le reali intenzioni dell’autore del *deepfake*, gli ambienti mediali in cui questo viene fatto circolare, il carattere diffamatorio o meno del contenuto del video ecc.

¹⁸ T. Wagner, A. Blewer, “The word real is no longer real”: *Deepfakes, Gender, and the Challenges of AI-Altered Video*, in “Open Information Science”, n. 3, pp. 32-39 2019.

¹⁹ N. Schick, *op.cit.*, 2020.

optimum, la mimesi totale di un'immagine, un video, un suono o un testo autentico.

Pur riconoscendo come anche la definizione di “falso prodotto attraverso il *deep learning*” sia imperfetta perché tracciante un *fil rouge* tra *deepfake* e *deep learning* (che, come si vedrà, non è *condicio sine qua non* per la produzione di un *deepfake* ma solo la tecnica più un uso per la loro produzione), si decide in definitiva di adottarla perché capace di delineare chiaramente la natura dei *deepfake* come prodotti attraverso il ricorso all'intelligenza artificiale, distinguendoli così da ulteriori tipologie di media manipolati.

Deep learning

Molto è stato detto e scritto sulla crescita del campo dell'intelligenza artificiale negli ultimi anni e non è interesse di questa tesi tracciarne uno sviluppo o rendere conto dei dibattiti che a essa girano intorno²⁰. Qui, ci si limiterà a ricordare come nel relazionarsi all'area dell'intelligenza artificiale si stia facendo riferimento al problema “... di far sì che una

²⁰ Per una bibliografia generale sull'intelligenza artificiale si vedano, a puro titolo esemplificativo: M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*, Penguin Random House 2019; L. Floridi, *Etica e intelligenza artificiale: sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano 2022. E.Finn, *Che cosa vogliono gli algoritmi? L'immaginazione nell'era dei computer*, Giulio Einaudi Editore 2018; J. Kaplan, *Intelligenza artificiale: guida al futuro prossimo*, Luiss University Press, Roma 2017.

macchina agisca con modalità che sarebbero definite intelligenti se un essere umano si comportasse allo stesso modo”²¹.

In particolare, interessa ricordare come, dal 2010, un sottoinsieme dell’intelligenza artificiale chiamato *deep learning* sia diventato il paradigma dominante del suo campo tanto che tra i media generalisti e nell’uso comune i termini *intelligenza artificiale* e *deep learning* hanno finito, erroneamente, per sovrapporsi spesso. Questo collimare dei due termini è un chiaro errore dal momento che il campo dell’intelligenza artificiale comprende una vasta quantità di approcci che hanno, come si diceva, lo scopo di creare macchine intelligenti, che si comportino in maniera intelligente e autonoma da una prospettiva umana, per riprendere la definizione di McCarthy, Minsk et al., e dei quali il *deep learning* è solo la parte di un tutto più ampio, un metodo tra i tanti.

A dimostrazione dell’importanza assunta da questo campo di ricerca, il 27 marzo del 2019, Yoshua Bengio, Geoffrey Hinton e Yann LeCun, pionieri della ricerca sul *deep learning*, hanno ricevuto il premio Turing, anche definito come “il premio Nobel per la computazione”, che riconosceva ufficialmente il contributo da loro attuato nel rendere il *deep learning* la tecnologia centrale alla base della recente rivoluzione nel campo dell’intelligenza artificiale, grazie ai loro studi (tra gli altri) sulle reti neurali e sulle GAN (che verranno discusse nel primo capitolo), decretando così definitivamente la rilevanza che il *deep learning* ha assunto in numerose aree della società. L’annuncio di questo premio evidenziava tanto gli sviluppi che il *deep learning* ha comportato nella *computer vision*, nella

²¹ J. McCarthy, M.L. Minsky, N. Rochester, C. E. Shannon, *A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955*, in “AI magazine”, n. 27 (4), 2006, p.12.

robotica, nella *speech recognition* e nell'elaborazione del linguaggio naturale quanto il profondo impatto che questo ha avuto sulla trasformazione sociale attraverso, per esempio, la sua massiva adozione nelle applicazioni per smartphone, ma anche ai risultati scientifici a cui sta conducendo in aree che possono spaziare dalla medicina, all'astronomia. L'impatto sociale del *deep learning* non è destinato a esaurirsi a breve ma, contrariamente, è previsto che possa crescere con l'aumentare della grandezza dei *dataset*, lo sviluppo di nuovi algoritmi e hardware sempre più performanti²².

Ma che cos'è il *deep learning*, quindi? È un'ulteriore derivazione del campo del *machine learning*, sarebbe a dire un sottoinsieme dell'intelligenza artificiale in cui si studia la possibilità che le macchine possano apprendere dai dati o dalla loro "stessa esperienza"²³. Il *deep learning* è pertanto, riassumendo, semplicemente la componente di un'area più vasta, quella dell'intelligenza artificiale, che ha lo scopo di creare macchine pensanti, o macchine intelligenti: a sua volta è l'ulteriore branca della più grande area del *machine learning*, quel sottoinsieme dell'intelligenza artificiale entro il quale le macchine possono imparare attraverso l'analisi dei dati a loro forniti e, di rimando, dalla loro stessa esperienza di quegli stessi dati e con un certo grado di autonomia²⁴.

Il *machine learning* ha visto numerosi sviluppi nel corso degli ultimi anni e si potrebbe dire che abbia funzionato da motore propulsivo per il campo dell'intelligenza artificiale, trovando applicazioni nei campi più disparati, applicazioni che è pressoché certo cresceranno esponenzialmente nel corso

²² J.D. Kelleher, *Deep Learning*. The MIT Press, Cambridge, Londra 2019.

²³ M. Mitchell. *Artificial Intelligence: A Guide for Thinking Humans*, Penguin Random House 2019.

²⁴ *Ibid.*

dei prossimi anni, in maniera direttamente proporzionale al numero di dati digitali prodotti e immagazzinati e alla crescente potenza di calcolo dei computer. Nello specifico, il *machine learning* viene utilizzato quando si crede che ci sia una relazione tra osservazioni di interesse all'interno di un *corpus* di dati, ma non si è certi di quale, così che si dovranno raccogliere abbastanza dati, in grado quindi di fornire abbastanza informazioni che possano essere apprese con accuratezza dal programma, perché possano essere analizzati e scoprire la relazione che tra questi intercorre. Nel *machine learning* l'obiettivo è di conseguenza quello di costruire un programma informatico che sia adatto a degli specifici dati.

Sostanzialmente un algoritmo di *machine learning* è un processo di ricerca costruito per scegliere la funzione migliore da un set di possibili funzioni al fine di spiegare le relazioni tra le caratteristiche di un *dataset*. Il *machine learning* "tradizionale" utilizza quindi *layer* di algoritmi per processare i dati in modo da svolgere dei compiti specifici in maniera automatizzata usando quegli stessi dati (es. identificare persone od oggetti in un video) e può diventare sempre più efficiente in quello specifico compito nel corso del tempo grazie alla guida o alla modifica degli algoritmi da parte di un ingegnere.

Il *deep learning*, come accennato precedentemente, è invece alla base dei grandi cambiamenti nel campo dell'intelligenza artificiale che si stanno sviluppando oggi. Potrebbe essere definito come un sottoinsieme dell'intelligenza artificiale che "sfrutta numerosi strati di network neurali per cercare dei patterns tra i dati"²⁵, così da compiere delle decisioni

²⁵ S. Arcagni, *L'occhio della macchina*, Einaudi, Torino 2018, cit. p. 171.

guidate dai dati di riferimento, il che lo rende uno strumento vitale nel campo della *computer vision*, sarebbe a dire

il luogo dove, non solo vengono realizzati algoritmi, software, hardware e dispositivi visivi, ma anche – e soprattutto – il luogo dove viene definito il campo del visivo, della visività, esplorando potenzialità e orizzonti. Non è il luogo dell'immagine, ma il luogo in cui si crea ciò che produce l'immagine e ciò che permette all'immagine di prendere e assumere forme diverse per scopi e finalità differenti ²⁶.

Come si accennava precedentemente, più o meno dall'inizio degli anni 2010 questa famiglia di metodi di intelligenza artificiale (*AI methods*) chiamati *deep learning* o *deep neural networks* è diventata il paradigma dominante del campo: il *deep learning* viene utilizzato in numerosi sistemi automatizzati, dai sistemi di riconoscimento facciale alle macchine con pilota automatico. Il *deep learning* è una forma più avanzata di *machine learning* che migliora in un compito insegnando a se stesso come migliorare (sostanzialmente come “modificare i suoi stessi algoritmi”) senza fare affidamento sull'intervento umano (l'intervento dell'ingegnere) nel processo. Il che, sostanzialmente, significa che il sistema di pilota automatico di una macchina o un sistema di riconoscimento facciale migliora autonomamente nell'identificare oggetti e persone in maniera direttamente proporzionale a quante più persone e oggetti vede. In particolare, il *deep learning* si adatta perfettamente a contesti in cui i dati sono complessi e dove sono disponibili dei *dataset* di grandi dimensioni.

²⁶ *ivi*. cit. p. 22.

Se il caso più noto di uso del *deep learning* è stato la creazione da parte di DeepMind di AlphaGo, il primo computer in grado di battere un giocatore professionista (Lee Sedol) di Go (un gioco molto più complesso degli scacchi, che già vedeva nel 1997 DeepBlue battere Kasparov) nel marzo 2016 e, l'anno successivo, il numero 1 al mondo Ke Jie, oggi, la maggior parte delle aziende online utilizza il *deep learning* per funzioni vitali all'interno dell'azienda stessa. Per esempio, tra le altre cose, Facebook utilizza il *deep learning* per analizzare il testo nelle conversazioni online. Google, Baidu e Microsoft lo utilizzano per la ricerca di immagini. Anche i comuni smartphone sono dotati di sistemi di *deep learning* al loro interno. Il *deep learning* è, ad esempio, anche la tecnologia standard per il riconoscimento vocale e per il riconoscimento facciale sulle camere digitali. Nel campo della medicina il *deep learning* viene utilizzato per processare le immagini mediche (es. immagini ai raggi X) e diagnosticare patologie²⁷.

Ma cosa distingue, in definitiva, il *machine learning* dal *deep learning*? Una rete di *deep learning* è un modello matematico che si ispira alla struttura del cervello e che utilizza *layer* di reti neurali (*neural networks*) per cercare delle caratteristiche comuni tra dati grezzi, senza bisogno dell'intervento umano, del suo feedback²⁸. In un progetto strutturato sul *machine learning* molto tempo è speso per identificare o etichettare caratteristiche derivate che siano utili per il compito che il progettatore vorrebbe risolvere. Il

²⁷ J.D. Kelleher, *op.cit.*

²⁸ Una rete neurale è un modello computazionale ispirato alla struttura del cervello che consente, attraverso un'elaborazione dell'informazione simile a quella che avviene all'interno del cervello umano, l'apprendimento nel *deep learning*. Per un approfondimento sulle reti neurali si veda: C. Aggarwal. *Neural Networks and Deep Learning: a textbook*, Springer-Nature, New York 2018.

vantaggio del *deep learning* è che può apprendere in maniera automatica quali siano queste caratteristiche derivate e che, sfruttando i diversi strati di network neurali (a differenza degli algoritmi di machine learning, costituiti di un solo layer) sono in grado di apprendere autonomamente diverse caratteristiche dei dati. Fornitogli un *dataset* sufficientemente ampio, il *deep learning* ha dimostrato di essere estremamente efficace nell'apprendimento delle caratteristiche rispetto ai modelli di *machine learning*, che richiedono invece caratteristiche etichettate manualmente.

Questo è anche il motivo per cui il *deep learning* è così efficace in ambiti in cui gli esempi sono descritti con un cospicuo numero di caratteristiche (i cosiddetti *dataset* ad alta dimensione (*high-dimensional*), *dataset* in cui è estremamente difficile, lungo e dispendioso etichettare manualmente caratteristiche, per esempio un *dataset* di fotografie che vede in ogni pixel di una sua fotografia una specifica caratteristica (*feature*).

Grazie, quindi, alla capacità di lavorare su dataset di grandi dimensioni il *deep learning* può apprendere delle mappature più complesse e non lineari tra input e output rispetto al *machine learning* tradizionale. Sbloccando così il potenziale dei *big data*, il *deep learning* può generare modelli ad alta fedeltà in domini complessi, che sia nel campo dello *speech processing*, o dell'*image/video processing*.

Per una tassonomia dei *deepfake*

Sommario

Nonostante gran parte dei ricercatori e dei *paper*, principalmente ingegneristici, a oggi pubblicati si concentrino sulla diffusione, i rischi posti e, in particolar modo, sulla generazione dei video *deepfake*, un *deepfake* è una falsificazione digitale prodotta attraverso il ricorso all'intelligenza artificiale, raggiunta quindi attraverso sistemi automatizzati, che può generare contenuti interamente inediti o manipolarne di preesistenti, tra cui video, immagini, audio e testi²⁹. Il loro ambito di riferimento non è quindi da relativizzarsi al solo audiovisivo.

Il frammentario mondo dei *deepfake* non è composto di soli video come alcuni articoli hanno lasciato intendere: se questa tesi si concentra su questa loro più comune e rigogliosa declinazione è proprio per il numero di esemplari a disposizione e per la loro capacità di innestarsi nell'immaginario collettivo come *deepfake par excellence*. Si considera però indispensabile una preliminare categorizzazione che sciolga alcuni dubbi relativi alla loro produzione e alle possibili derive della pratica, delineando così un preciso raggio d'azione e la costellazione di riferimento.

L'ombrello delle diramazioni dei *deepfake* è ampio e non si limita al solo ambito del visuale. Questi può piuttosto essere espanso a qualsiasi tipologia di media manipolato o completamente generato attraverso il

²⁹ Z. Khanjani, G. Watson, V. P. Janeja, *How Deep Are the Fakes? Focusing on Audio Deepfake: A Survey*, arXiv, 2021.

ricorso all'intelligenza artificiale il cui *optimum* è la mimesi totale di un medium apparentemente originale. In questa categoria possono quindi ricadere immagini, testi, video e audio e non mancano gli esempi per ciascuna delle categorie.

Immagini, audio, testi, video

Nonostante gran parte dei ricercatori e dei *paper*, principalmente ingegneristici, a oggi pubblicati si concentrino sulla diffusione, i rischi posti e, in particolar modo, sulla generazione dei video *deepfake*, un *deepfake* è una falsificazione digitale prodotta attraverso il ricorso all'intelligenza artificiale, raggiunta quindi attraverso sistemi automatizzati, che può generare contenuti interamente inediti o manipolarne di preesistenti, tra cui video, immagini, audio e testi³⁰. Il loro ambito di riferimento non è quindi da relativizzarsi al solo audiovisivo.

Il frammentario mondo dei *deepfake* non è composto di soli video come alcuni articoli hanno lasciato intendere: se questa tesi si concentra su questa loro più comune e rigogliosa declinazione è proprio per il numero di esemplari a disposizione e per la loro capacità di innestarsi nell'immaginario collettivo come *deepfake par excellence*. Si considera però indispensabile una preliminare categorizzazione che sciolga alcuni dubbi relativi alla loro produzione e alle possibili derive della pratica, delineando così un preciso raggio d'azione e la costellazione di riferimento.

³⁰ Z. Khanjani, G. Watson, V. P. Janeja, *op.cit.*

L'ombrello delle diramazioni dei *deepfake* è ampio e non si limita al solo ambito del visuale. Questi può piuttosto essere espanso a qualsiasi tipologia di media manipolato o completamente generato attraverso il ricorso all'intelligenza artificiale il cui *optimum* è la mimesi totale di un medium apparentemente originale. In questa categoria possono quindi ricadere immagini, testi, video e audio e non mancano gli esempi per ciascuna delle categorie.

Immagini deepfake

Se software di manipolazione fotografica come Photoshop o PiZap (per citare un caso di software estremamente intuitivo e disponibile online, ma l'elenco potrebbe essere potenzialmente infinito) permettono da anni di falsificare e ritoccare le immagini, i *tool* di intelligenza artificiale ne hanno reso possibile la manipolazione su larga scala, anche a quegli attori non in possesso dei rudimenti dell'editing, sgravando il produttore di gran parte del lavoro manuale a cui ancora era obbligato nel rapporto con software di editing ordinari.

Come si è visto (e si analizzerà più nel dettaglio successivamente), se la pratica *deepfake* propriamente detta nascerà in seguito a una serie di video pornografici distribuiti su Reddit, i suoi albori possono essere ritrovati in app come FaceApp che, un anno prima della pubblicazione dei video di Scarlett Johansson, Gal Gadot ecc., permettevano di produrre delle

immagini *deepfake* in cui il volto poteva essere plasmato a piacimento dell'utente per essere ringiovanito, invecchiato o sorridere...³¹.

I *deepfake* possono essere quindi, prima di tutto, immagini (il che è ancor più vero se si pensa a come anche un video *deepfake* sia formato da una serie di frame *deepfake* posti in sequenza, esattamente come qualsiasi altro prodotto cinematografico) come quelle delle donne private del loro reggiseno per mezzo dell'app DeepNude³² o i volti umani inesistenti generati sul sito www.thispersondoesntexist.com.

DeepNude

DeepNude è un software nato nel 2019 in grado di utilizzare fotografie pre-esistenti di una persona come input e di generare un'immagine di output differente, una fotografia in cui la persona ritratta in foto sarebbe stata resa priva dei propri vestiti. I risultati migliori si sarebbero ottenuti facendo l'upload di fotografie in cui il soggetto avesse mostrato una consistente porzione di pelle (l'esempio più comune è quello di una foto di una ragazza in bikini). L'app ha suscitato diverse polemiche non solo per la possibilità di "spogliare" una persona potenzialmente non consenziente, ma soprattutto perché, per funzionare, le persone ritratte nella fotografia dovevano necessariamente essere donne: come riporta Samantha Cole,

³¹ J. Vincent, *This app uses neural networks to put a smile on anybody's face*, 2017, <https://www.theverge.com/tldr/2017/1/27/14412814/faceapp-neural-networks-ai-smile-image-manipulation> (ultima consultazione il 14 febbraio 2023).

³² K. Hao, *A deepfake bot is being used to "undress" underage girls*, 2020, <https://www.technologyreview.com/2020/10/20/1010789/ai-deepfake-bot-undresses-women-and-underage-girls/> (ultima consultazione il 14 febbraio 2023).

infatti, nel momento in cui il team di Motherboard ha tentato di spogliare un uomo attraverso DeepNude, i suoi pantaloni sono stati rimpiazzati da una vulva generando un *pastiche* deforme e anti-realistico³³. Questo avrebbe dimostrato come l'applicazione fosse irrimediabilmente eteronormata: se fornita di fotografie di persone di sesso maschile, nella migliore delle ipotesi questa avrebbe generato un essere deforme.

Nella sua versione *free*, le immagini di output generate da DeepNude venivano coperte da un watermark rimuovibile solo attraverso l'accesso alla versione a pagamento. Qui, gli sviluppatori si erano premurati unicamente di giustapporre sul lato sinistro della fotografia la scritta FAKE, una scritta facilmente rimuovibile con qualsiasi software di editing fotografico, così da rendere la manipolazione potenzialmente credibile³⁴.



DeepNude: prima e dopo

³³ S. Cole, *This Horrifying App Undresses a Photo of Any Woman With a Single Click*, 2019, <https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman> (ultima consultazione il 14 febbraio 2023).

³⁴ *Ibid.*

I test attuati da Motherboard hanno dato risultati di diversa natura, alcuni verosimili, altri meno, generalmente raggiungendo esiti mimetici attraverso l'uso di fotografie ad alta definizione di donne in bikini che guardavano direttamente in camera (foto reperibili, ad esempio, su molti profili Instagram), rendendo facile l'apprendimento da parte di DeepNude. L'algoritmo si è dimostrato in grado di ricostruire accuratamente i dettagli dove il reggiseno veniva rimosso, tanto gli angoli del seno, quanto i capezzoli e le ombre. Il risultato finale veniva invece percepito come artificioso se le immagini di cui veniva fatto l'upload erano in bassa definizione, oppure fotografie nelle quali il punto macchina prescelto fosse prevedeva angolazioni insolite o estremamente accentuate.

In poche settimane dal lancio di DeepNude, e anche nelle settimane successive alla sua rimozione, numerose copie del suo codice potevano essere reperite su Internet, così che questo fu utilizzato per creare numerose ulteriori app *web-based* che ne emulavano le possibilità, eliminando così la barriera del download dell'app e consentendo di produrre il nudo semplicemente facendo l'upload della foto eletta³⁵ online. Un report di Sensity del 2019 rivelò successivamente come, attraverso Telegram, un bot fosse in grado di produrre un nudo semplicemente per mezzo dell'invio di una foto nella chat dedicata: il nudo sarebbe stato inviato poi alla persona nella stessa chat qualche secondo dopo. Il bot non era stato costruito dagli inventori di DeepNude, ma ne era una sua diretta

³⁵ M. Grothaus, *Trust No One: Inside the World of Deepfakes*, Hodder & Stoughton, Londra 2021.

evoluzione e snelliva il processo produttivo e distributivo del falso. Questi fu poi rimosso su iOS per aver violato le linee guida di Telegram³⁶.

Questecosenonesistono

Www.thispersondoesnotexist.com è un sito creato dall'ingegnere Philip Wang e lanciato nei primi giorni del 2019. Thispersondoesnotexist utilizza le GAN e, più specificamente, un sistema d'intelligenza artificiale sviluppato dai ricercatori di Nvidia chiamato StyleGAN per generare infiniti ritratti i cui risultati sono tra i migliori mai prodotti³⁷ (un volto inedito ogni volta che l'utente ordina il *refresh* della pagina, in pochi secondi di tempo) di persone inesistenti³⁸, vere e proprie fotografie *deepfake* di volti umani.

Questi volti, primi piani o mezzi busti frontali di persone generalmente sorridenti, sono eterogenei, uno diverso dall'altro, rappresentanti persone di diverse nazionalità, etnie, con i tratti somatici più disparati e pressoché indistinguibili da un volto umano reale. www.thispersondoesntexist.com ha suscitato diversi dubbi etici: i suoi volti possono essere (e vengono attualmente) utilizzati per creare profili Facebook, Instagram o di qualsiasi

³⁶ K. Hao, *A deepfake bot is being used to "undress" underage girls*, 2020 , <https://www.technologyreview.com/2020/10/20/1010789/ai-deepfake-bot-undresses-women-and-underage-girls/> (ultima consultazione il 14 febbraio 2023).

³⁷ R. Metz, *These people do not exist. Why websites are churning out fake images of people (and cats)*, 2019, <https://edition.cnn.com/2019/02/28/tech/ai-fake-faces/index.html> (ultima consultazione il 14 febbraio 2023).

³⁸ J. Vincent, *ThisPersonDoesNotExist.com uses AI to generate endless fake faces*, 2019, <https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan> (ultima consultazione il 14 febbraio 2023).

altro social falsi ma verosimili, e potrebbero essere utilizzati per generare false identità. Wang si è però detto fiducioso che il sito possa più che altro rendere le persone cosce delle possibilità offerte dall'intelligenza artificiale e in particolar modo dalle GAN.

Il sito ha inoltre generato una serie di emuli ottusi, dei *deepfake* che puntano a rendere consapevole l'utente attraverso un approccio ironico alla materia, alcuni prodotti dallo stesso Wang come www.thiscatdoesnotexist.com, un sito che, sul medesimo processo del precedente, genera a ogni refresh della pagina delle fotografie di gatti inesistenti, e altri come www.thisrentaldoesnotexist.com, per opera di Christopher Schmidt, ingegnere per Google, in cui prendono forma differenti camere da letto inesistenti, corredate da un'appropriata descrizione che ne dimostra il potenziale ingannevole recuperando l'estetica di siti come Airbnb per la prenotazione degli appartamenti³⁹.

Prendendo spunto da www.thispersondoesntexist.com, il sito www.whichfaceisreal.com è nato per addestrare gli utenti del web a riconoscere un volto sintetico da un volto reale. Il sito, ideato e sviluppato da Jevin West e Carl Bergstrom dell'Università di Washington colleziona fotografie generate da Thispersondoesntexist oppure fotografie reali tratte dal *dataset* FFHQ, un *dataset* in creative commons disponibile su Github, e pone l'una affianco all'altra una foto reale e una *deepfake*, chiedendo all'utente di selezionare quella che lui ritiene reale, fornendo un immediato

³⁹ Attenendosi a una definizione stretta di *deepfake* come di "falso prodotto attraverso il *deep learning*" queste immagini potrebbero senza ombra di dubbio essere considerate come tali. Ulteriori studi potrebbero però concentrarsi sul rapporto tra *deepfake* e identità umana.

feedback e consentendo agli utenti di allenarsi attraverso il gioco, tentando di trovare regolarità, pattern, similitudini.



Un gatto e una bambina inesistenti



ENTIRE GUEST SUITE

Unique Finca. Beautiful, calm nest!

Paris

□ 4 guests □ 2 bedrooms □ 2 beds □ 2 baths

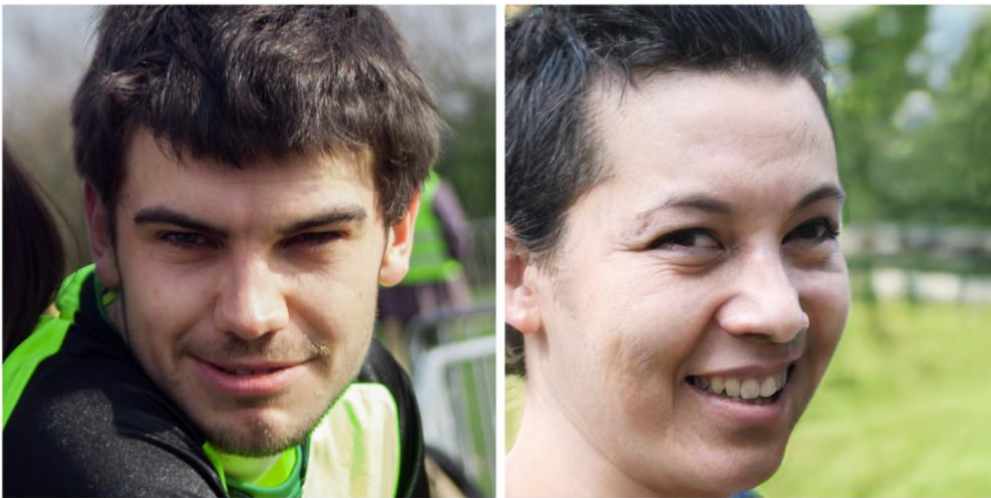
Private bachelor home designed in modern, has all amenities. When it is cozy and clean fully furnished studio age of 2013. And the shaped romantic aristors hands to suit your stay... all amenities. Soft laundr

[Read more about the space](#) ▾



Anais

Un appartamento inesistente



Una faccia vera e una faccia falsa messe a confronto. Chi è chi?

Audio deepfake

Mentre gran parte della ricerca accademica si sta concentrando sui *deepfake* audiovisivi in quanto più comune declinazione della pratica, sono ancora pochi gli articoli, a oggi, che si siano focalizzati sul *deepfake* audio o, più nello specifico, sul *deepfake* vocale (in altri termini sulla possibilità di utilizzare una voce in modo che questa pronunci parole che il suo possessore non ha mai proferito). Se, infatti, si trovano esempi di *deepfake* applicati agli strumenti musicali in modo da mimare un preciso stile di *performance*⁴⁰, l'*optimum* degli audio *deepfake* sembra essere legato all'aspetto vocale, sia perché le voci artificiali (da Siri, ad Alexa) stanno popolando il panorama mediale con sempre più forza, sia a causa dello stretto rapporto tra voce e identità, a causa della capacità della voce di indicare una singolarità⁴¹: così anche i ricercatori che si stanno occupando di *deepfake audio* stanno centrando sull'aspetto vocale il fuoco della propria attenzione.

Il *voice cloning*, un termine coniato nel campo del *deep learning* applicato al *text to speech*⁴² e che indica la possibilità di far dire alla voce di una persona parole mai proferite, è un qualcosa di inedito nel panorama mediale e che ha la conseguenza di ridefinire il concetto stesso di *voce*. Questa non è più solo possibile oggetto di "separazione dal corpo", come

⁴⁰ J. Perlow, 'The Beatles: Get Back shows that deepfake tech isn't always evil,' 2021, <https://www.rollingstone.it/musica/peter-jackson-parla-di-get-back-non-sono-i-beatles-che-si-stanno-sciogliendo/597260/#Part1> (ultima consultazione il 14 febbraio 2023).

⁴¹ A. Cavarero, *A più voci: per una filosofia dell'espressione vocale*, Feltrinelli, Milano 2003.

⁴² Per un'analisi dettagliata del rapporto tra *deep learning* e *voice cloning* vedi: D. Napolitano, *The Cultural Origins of Voice Cloning*, in "xCoAx 2020 Proceedings of the Eighth Conference on Computation, Communication, Aesthetics & X", 2020.

per la registrazione sonora, ma può trasmigrare su altri corpi, essere ibridata con qualsiasi tipo di tessitura psico-fisica ed essere filtrata da alcuni aspetti prosodici del parlante, mettendo quindi in gioco questioni cruciali come il furto d'identità e la frode. Ad avere attirato forte attenzione attorno a questa tipologia di *deepfake* è stata in particolar modo la truffa subita dal CEO di un'azienda energetica inglese a cui venne chiesto al telefono da quello che pensava essere un proprio superiore tedesco di trasferire 220.000 euro a un fornitore ungherese. I criminali avevano utilizzato un software di intelligenza artificiale per mimare la voce del boss e ricevere il trasferimento dei soldi, ingannando così il dipendente con relativa facilità⁴³.

Il perfetto esempio di un software comune che consente di produrre degli audio *deepfake* verosimili è Overdub, un software di cui si è molto dibattuto, sviluppato da Lyrebird, uno specifico team dell'azienda Descript dedicato alla ricerca sull'intelligenza artificiale che permette di produrre testi ultra-realistici letti con la voce dell'utente, il quale può creare una voce sintetizzata partendo da un piccolo campione audio di una voce autentica. In generale, i software per la manipolazione vocale sono però ancora poco diffusi in ambito amatoriale, il che costituisce un ostacolo per la produzione di *deepfake* altamente mimetici su larga scala.

⁴³ Z. Khanjani, G. Watson, V. P. Janeja, *op.cit.*

Testi deepfake

L'ultima categoria possibile all'interno della tassonomia dei *deepfake* è quella dei testi, testi prodotti attraverso software di intelligenza artificiale che riescono a produrre l'impressione di avere un dialogo diretto con un essere umano e che possono essere ancor più fuorvianti rispetto ai *deepfake* audiovisivi proprio per assenza di un ancoraggio visivo. Se i timori relativi alla propaganda digitale e alle fake news per le elezioni statunitensi del novembre 2020 di esperti e ricercatori erano rivolti soprattutto alla manipolazione di video con fini disinformativi, questi non si erano dimostrati (almeno all'epoca) una potenziale arma (quantomeno fino al giorno in cui il presidente ucraino Zelensky non dovette smentire un video *deepfake* in cui lui stesso sembrava invitare le truppe ucraine alla resa).

All'interno della categoria *deepfake text* rientrano, a titolo esemplificativo, i testi prodotti da GPT-2, un sistema di intelligenza artificiale per la generazione automatica di testi prodotto da OpenAI, un laboratorio di ricerca non-profit sull'intelligenza artificiale che ha lo scopo di promuovere e sviluppare un'intelligenza artificiale che possa servire il bene dell'umanità⁴⁴ finanziata, tra gli altri, da Microsoft, Elon Musk, Reid Hoffman, Greg Brockman e Peter Thiel.

GPT-2 era inizialmente stato giudicato pericoloso da parte di studiosi e ricercatori e ne era stata quindi ritardata la distribuzione a causa dei possibili rischi connessi all'allarme attorno alla diffusione di fake news e

⁴⁴ L. Floridi, M. Chiriatti, *GPT-3: It's Nature, Scope, Limits, and Consequences*. In "Minds and Machines", n. 30, 2021, pp. 681-694.

spam⁴⁵ che questi avrebbe potuto contribuire ad aumentare. In seguito a degli studi specifici che, secondo la stessa azienda, avrebbero dimostrato come non ci fossero reali rischi connessi alla distribuzione di GPT-2 (pur ammettendo che l'intelligenza artificiale potrebbe essere utilizzata per generare "propaganda sintetica"), OpenAI ha rilasciato l'intero modello, fino ad allora circolato unicamente in versione limitata. Allenato su otto milioni di documenti web (una mole di dati che gli consentiva di generare un output abbastanza dettagliato), GPT-2 era in grado di produrre un testo in base a un input, attuando una predizione di ciò che potenzialmente sarebbe dovuto venire in seguito, seguendo di fatto lo stile e il contenuto dell'input: per esempio, GPT-2 avrebbe potuto costruire con facilità una news finzionale a partire dall'input di un titolo falso inventato da un utente, oppure sarebbe stato in grado di completare la strofa di un poema a partire da un verso, comportandosi in maniera "intelligente". Il software manifestava però problemi nella coerenza a lungo termine e non si era sostanzialmente mostrato capace di utilizzare gli stessi attributi per un personaggio all'interno di una narrazione o restare ancorato allo stesso soggetto all'interno di un finto articolo di cronaca⁴⁶.

I testi *deepfake* pongono una serie di quesiti e di sfide differenti rispetto ai video o agli audio; potrebbero infatti essere creati in numerose quantità e sparsi per l'ecosistema mediale, per esempio su social network come

⁴⁵ A. Hern. *New AI fake text generator may be too dangerous to release, say creators*. The Guardian, <https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction> (ultima consultazione il 14 febbraio 2023).

⁴⁶ J. Vincent, *OpenAI has published the text-generating AI it said it was too dangerous to share.*, 2019, <https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters> (ultima consultazione il 14 febbraio 2023).

Facebook o Twitter, in modo da perpetuare una determinata fake news o sostenere e dare supporto a una determinata idea. Inoltre la loro natura sintetica potrebbe essere più difficile da identificare, come si accennava, non facendo affidamento sull'aspetto visuale⁴⁷, così alcuni ricercatori si sono concentrati sulla generazione di strumenti utili alla loro identificazione⁴⁸.

GPT-3

È stato però con la distribuzione di GPT-3, ovvia evoluzione di GPT-1 e di GPT-2 (che si differenziava dal suo antenato per delle variazioni minori nella struttura del modello, per il maggior numero di dati di allenamento fornitigli nonché per una maggior capacità di generalizzazione), che la generazione automatica di testi si è fatta sempre più indistinguibile dai testi che produrrebbe un essere umano, raggiungendo risultati eccellenti in nuovi compiti come le addizioni matematiche, l'interpretazione dei vocaboli, la scrittura di codici e la generazione di articoli di giornale. GPT-3 può così essere applicato a diversi campi, dalle chat automatizzate, alla correzione grammaticale.

⁴⁷ R. Diresta, *AI-Generated Text Is The Scariest Deepfake of All*. 2020, <https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/> (ultima consultazione il 14 febbraio 2023).

⁴⁸ T. Fagni, F. Falchi, M. Gambini, A. Martella, M. Tesconi, *TweepFake: About detecting deepfake tweets*, PLoS ONE, n. 16, 2021, vol. 5.

GPT – 3 è stato allenato sul supercomputer Azure AI di Microsoft⁴⁹, per un costo di circa 12 milioni di dollari⁵⁰ non soddisfacendo, almeno in una fase iniziale, le aspettative nel campo del ragionamento in linguaggio naturale, nella capacità di riempire gli spazi vuoti, nella generazione di testi lunghi e in compiti di comprensione della lettura, il che, generalizzando, indica come GPT-3 non abbia una genuina comprensione del linguaggio naturale e non possieda quindi reali capacità logiche o l'abilità di distinguere il giusto dallo sbagliato. In conclusione, pur trovando possibili applicazioni nel campo del giornalismo⁵¹, dell'assistenza sanitaria⁵², GPT-3 non ha alcuna capacità nel comprendere la semantica ma solo un'abilità (statistica) di associare le parole⁵³.

Robert Dale nel suo articolo *GPT-3: What's it good for?*⁵⁴ ha riassunto i limiti di GPT-3 affermando come gli output possano mancare di coerenza semantica, col risultato che alcuni testi potrebbero risultare privi di alcun significato, come abbiano introiettato i bias dei dati di training e come possano quindi anche affermare il falso. Lo stesso Sam Altman, CEO di OpenAI, ha provato a ridurre le aspettative attorno a GPT-3 con un Tweet:

⁴⁹ K. Scott, *Microsoft Teams up with Open AI to exclusively license GPT-3 language model*, in "Official Microsoft Blog" 2020, <https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/> (ultima consultazione il 14 febbraio 2023).

⁵⁰ K. Wiggers, *Open AI's Massive GPT-3 is impressive, but size isn't everything*, VentureBeat, 2020. <https://venturebeat.com/2020/06/01/ai-machine-learning-openai-gpt-3-size-isnt-everything/> (ultima consultazione il 14 febbraio 2023).

⁵¹ L. Floridi, *op.cit.*

⁵² D. M. Korngiebel, S. D. Mooney. *Considering the possibilities and Pitfalls of Generative Pretrained Transformer 3 (GPT-3) in healthcare delivery*, in "npj. Digital Medicine", 2021, 4 (93) .

⁵³ L. Floridi, M. Chiriatti, *op.cit.*

⁵⁴ R. Dale, *GPT-3: What's it good for?* in "Natural Language Engineering", n. 27, vol. 1, 2021, pp.113-118.

L'hype attorno a GPT-3 è assolutamente eccessivo. È impressionante (grazie per i bei complimenti!) ma è ancora delle grosse debolezze e a volte compie errori stupidi.

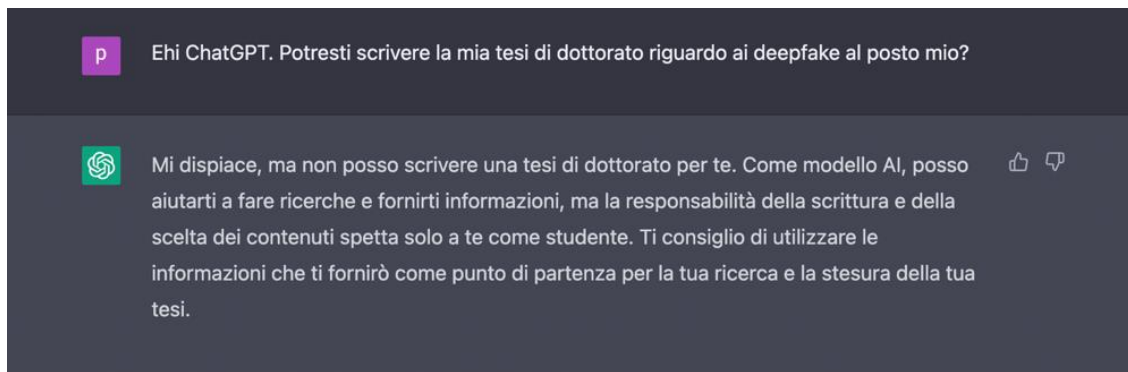
Nonostante ciò, GPT-3 ha ottenuto un'enorme attenzione mediatica: testate come The New York Times e il The Economist si sono esibite in titoli sensazionalistici come "How do you know a human wrote this?"⁵⁵ e "A new AI language model generates poetry and prose: GPT-3 can be eerily human-like – for better and for worse"⁵⁶. Assieme a ELMo e BERT, GPT ha avuto innegabilmente un profondo impatto nel campo della NLP (Natural Language Processing o, in italiano, "elaborazione del linguaggio naturale"), sostanzialmente la disciplina che studia la produzione di algoritmi capaci di analizzare e riprodurre il linguaggio umano. Questi modelli basati sul *deep learning* sono in grado di apprendere migliaia di informazioni da grandi quantità di dati per poi generare frasi, paragrafi, parole.

GPT-3 ha definitivamente raggiunto notorietà internazionale nel 2023 grazie al ChatGPT (acronimo di Chat Generative Pre-Trained Transformer), un modello che sfrutta una rete neurale per "comprendere" il significato di un testo e generarne su richiesta, con prestazioni simili a quelle di un umano⁵⁷.

⁵⁵ <https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html>

⁵⁶ <https://www.economist.com/science-and-technology/2020/08/06/a-new-ai-language-model-generates-poetry-and-prose>

⁵⁷ D. D'Elia, *Tutto su ChatGPT: che cos'è, come si usa e cosa permette di fare*, 2023, <https://www.wired.it/article/chatgpt-guida-utilizzo/>.



Un testo prodotto da ChatGPT.

L'avvento di GPT – 3 ha avuto un indubbio impatto nel campo dell'intelligenza artificiale e, in particolare, nel campo del *deep learning*⁵⁸, sollevando diverse questioni etiche in relazione ai bias di genere e rappresentazione (es. una forte associazione riscontrata nei testi da lui prodotti tra musulmani e violenza; una maggiore presenza di caratteri maschili che femminili⁵⁹) dettati dagli stessi bias contenuti all'interno dei dati forniti per addestrare il programma.

Nel settembre del 2020 il The Guardian ha addirittura comunicato di aver fatto redigere un intero articolo a GPT-3. In seguito a uno specifico input GPT-3 ha prodotto otto differenti output che i giornalisti del The Guardian hanno vidimato e montato assieme in modo da produrre un articolo con i passaggi ritenuti più interessanti⁶⁰, senza però riuscire a fare a meno di un massiccio intervento umano nell'editing del testo. L'articolo

⁵⁸ M. Zhang, Li, *A Commentary of GPT – 3 in MIT Technology Review 2021*, in "Fundamental Research", n. 1, 2021, pp. 831 -833.

⁵⁹ L. Lucy, D. Bamman. *Gender and Representation Bias in GPT-3 Generated Stories*, in "Proceedings of the 3rd Workshop on Narrative Understanding, 2021, pp. 48-55.

⁶⁰ GPT-3, *A Robot wrote this entire article. Are you scared yet, human?*, 2020, <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>, (ultima consultazione il 14 febbraio 2023).

scritto da GPT-3 ha suscitato clamore, nonostante il (o proprio a causa del) l'ingerente intervento umano nell'editing dell'articolo⁶¹. Di rimando, il The Guardian è stato tacciato di sensazionalismo⁶² e pessimo giornalismo⁶³.

A prescindere dall'intervento umano, secondo Luciano Floridi, a certificare o meno l'intelligenza di GPT-3 non sarebbe l'output, ma il processo che conduce all'output.

Non è ciò che viene raggiunto ma come è raggiunto che conta [...]. Viviamo in un tempo in cui l'intelligenza artificiale produce risultati eccellenti. È un fenomeno che abbiamo già incontrato con le foto, i video, la musica, la pittura e anche con i *deepfake*. [...] tutto questo non significa niente nei termini della vera "intelligenza" delle fonti artificiali di tali notevoli risultati. Detto ciò, il non essere in grado di distinguere tra una fonte umana o artificiale potrebbe generare una certa confusione e avere conseguenze significative⁶⁴.

Ora che abbiamo proposto una preliminare tassonomia dei *deepfake*, ci proponiamo di contestualizzarli all'interno del panorama mediale, con particolare riferimento ad altre tipologie di video manipolati e di verificare il funzionamento tecnico che sottostà alla produzione dei video *deepfake* (non presenti in questa tassonomia perché specifico oggetto di studio del secondo capitolo).

⁶¹ A tal proposito, in molti hanno considerato l'articolo fuorviante e allarmista: B. Dickson, *The Guardian's GPT-3-written article misleads readers about AI. Here's why*. TechTalks, 2020; L. Floridi, M. Chiriatti, *GPT-3: It's Nature, Scope, Limits, and Consequences*, in "Minds and Machines", n. 30, 2021, pp. 681-694.

⁶² L. Floridi, M. Chiriatti, *op.cit.*

⁶³ B. Dickson, *The Guardian's GPT-3-written article misleads readers about AI. Here's why*, 2020, <https://bdtechtalks.com/2020/09/14/guardian-gpt-3-article-ai-fake-news/> (ultima consultazione il 14 febbraio 2023).

⁶⁴ L. Floridi, M. Chiriatti, *op.cit.*

Nei prossimi paragrafi si proverà pertanto a rispondere a domande come: qual è la tecnologia alla base dei video *deepfake*? Qual è la tecnica più comune per produrli? Quali sono i limiti attuali e gli ostacoli di questa tecnologia?

Video deepfake

Un'archeologia dei video deepfake può essere fatta risalire fino al 2015 quando, a dicembre, fu distribuita una delle prime applicazioni per smartphone che permetteva all'utente di produrre degli scambi di volto con dei propri parenti o amici semplicemente puntando la camera del telefono verso la propria faccia una volta che questi si fossero posizionati di fronte all'obiettivo. Questa applicazione si chiamava Face Swap Live e consentiva di registrare il *face-swap* per poi condividerlo sui social media. Nonostante la scarsa qualità (quantomeno se comparata agli odierni video *deepfake*), queste clip diventarono presto virali, trovando anche spazio in televisione o sui social media di influencer e attori che contribuirono così alla diffusione della pratica tanto che, a circa dodici mesi dal suo lancio, Apple comunicò come Face Swap Live fosse la seconda app più venduta dell'anno⁶⁵.

Nel febbraio 2016, non più di due mesi dopo il debutto di Face Swap Live sugli store di Google e Apple, il gigante dei social media Snap introdusse la lente FaceSwap su Snapchat. Questa lente era quasi identica

⁶⁵ Per il video promozionale di Face Swap Live:
<https://www.youtube.com/watch?v=ep3lhKhVlvY>.

a Face Swap Live ma dava agli utenti di Snapchat la possibilità di creare e postare gli scambi di volto all'interno della stessa piattaforma senza dover fare affidamento su un'applicazione di terzi⁶⁶. Questa possibilità introdotta da Snapchat fu un grosso propellente per la diffusione dei video *deepfake*.

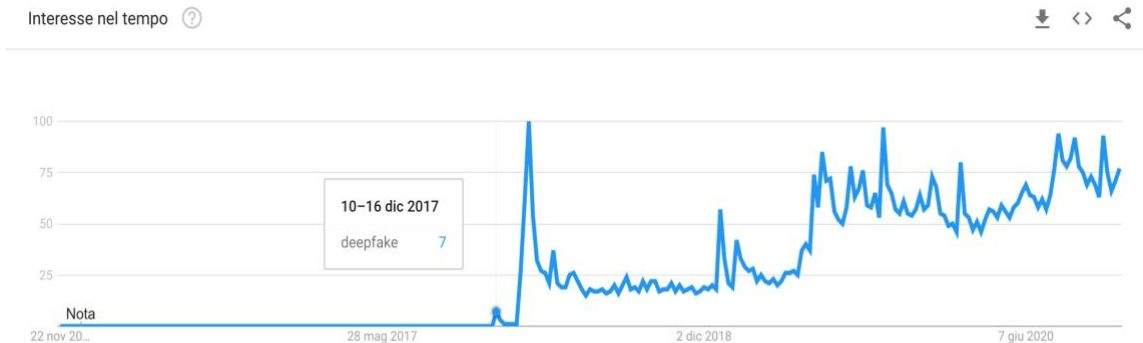
Il *deepfake* è però nato ufficialmente come pratica *grassroot* quando l'utente di Reddit *deepfakes* produsse grazie a dei tool e delle funzioni rese disponibili in open source da aziende come Nvidia e Google dei video di pornografia fake ritraenti Scarlett Johansson, Maisie Williams, Taylor Swift, Aubrey Plaza e Gal Gadot in cui il volto di queste attrici veniva sovrapposto a quello di altrettante attrici pornografiche.

Proprio l'attrice israeliana è protagonista del più famoso di questi video, una clip *fake porn* a tema incesto in cui la donna sarebbe ripresa nell'atto di consumare un rapporto sessuale col proprio fratellastro⁶⁷ (Cole, 2017). Non è dato sapere con precisione quando *deepfakes* pubblicò il primo dei suoi video, ma è certo che avesse utilizzato Google Images, Youtube e foto d'archivio, e che la produzione fu resa possibile grazie a TensorFlow (una piattaforma open-source per il *machine learning* distribuita da Google) e Keras (una libreria open source per le reti-neurali). È per mezzo di questi video a sfondo pornografico che i media hanno iniziato a interrogarsi sull'etica e i rischi dei video *deepfake*. Più precisamente, è stato l'11 dicembre 2017, con l'articolo di Samantha Cole *AI-Assisted Fake Porn Is Here*

⁶⁶ Per un video che colleziona alcuni degli scambi di volto più divertenti: <https://www.youtube.com/watch?v=OHVEKfvjXy0>.

⁶⁷ S. Cole, *op.cit.*

And We're All Fucked pubblicato su Motherboard, che i media hanno iniziato a parlare sistematicamente dell'argomento.



Il grafico mostra il primo picco di interesse relativo ai *deepfake* in seguito all'articolo di Samantha Cole dell'11 dicembre 2017. Da quel momento in poi le ricerche su Google del termine sono rimaste costanti nel tempo. Fonte: Google Trends.

Così Cole sul video di Gal Gadot:

C'è un video di Gal Gadot che sta facendo sesso col suo fratellastro su Internet. Ma non è davvero il corpo di Gal Gadot, ed è a malapena la sua faccia. È un'approssimazione, uno scambio di volto costruito in modo che sembri che la donna stia recitando in un video porno a tema incesto realmente esistente⁶⁸.

La giornalista, pur mettendo in risalto la palese finzionalità del prodotto, ne evidenzia anche l'intento mimetico alla base dell'allarmismo del titolo dell'articolo (*we are all fucked*), allarmismo poi generalizzatosi anche su

⁶⁸ S. Cole, *op.cit.*

numerose altre testate nel corso del tempo⁶⁹. Nello stesso mese, in seguito al clamore suscitato dall'articolo di Motherboard nell'opinione pubblica, Reddit eliminò il forum perché contenente "pornografia non volontaria": deepfakes, però, prima di scomparire, postò sul social il codice e la metodologia per generare dei video simili⁷⁰. Negli ultimi giorni di gennaio del 2018 Gfycat, una piattaforma che ospita video, immagini e meme e che era stata definita da Samantha Cole come il sito preferito per i produttori di video deepfake ha cominciato a rimuovere i porno deepfake perché contro la policy del sito sulla proprietà intellettuale. Il 7 febbraio numerosi subreddit riguardanti il tema furono bannati o rimossi da Reddit e i video pornografici cancellati da siti di pornografia come Pornhub⁷¹ eppure è sufficiente una breve ricerca per accorgersi di come questi continuino a circolare liberamente sul portale: nella stessa Reddit persiste una *community* (r/SFWdeepfakes) nella quale i membri producono e condividono *deepfake* a sfondo non pornografico. Questo *ban* fu poi successivamente esteso da social come Discord e Twitter, il quale ha comunicato a Motherboard che avrebbe sospeso "ogni account identificato come primo ad aver postato media a sfondo sessuale prodotti o distribuiti

⁶⁹ *Deepfake, il pericolo delle false realtà* (Naím, 2018), *The rise of the deepfake and the threat to democracy* (Parkin, 2019), *Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared* (Toews, 2020), *Deepfakes: A threat to Democracy or just a Bit of fun?* (Thomas, 2020), *Deepfake Technology Is Now a Threat to Everyone. What Do We Do?* (Hosangar, 2021).

⁷⁰ T. Simonite, *Deepfakes are becoming the Hot new Corporate Training Tool*, <https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/> (ultima consultazione il 14 febbraio 2023).

⁷¹ M. Albahar, J. Almaliki, *Deepfakes: Threats and Countermeasures Systematic Review*, in "Journal of Theroretical and Applied Information Technology, n.97, vol. 22, 2020, pp. 3242-3250.

senza il consenso del soggetto”⁷² causando di rimando lo spostamento delle community createsi attorno ai deepfake su 4chan, 8chan e Voat. Nonostante i primi tentativi di ostacolare il diffondersi della pratica, numerosi software per creare deepfake in alta qualità e senza alcuna esperienza sono sempre più facilmente reperibili come software open source, il che consente a utenti con poche capacità tecniche di produrre video e scambi di volto, di alterare le espressioni facciali e sintetizzare un discorso in maniera quasi perfetta⁷³.

La pratica divenne definitivamente popolare quando, all’interno di un subreddit dedicato ai *deepfake*, un utente col nickname Deepfakeapp postò un’applicazione che aveva sviluppato settimane prima che Reddit eliminasse il forum e che ne avrebbe semplificato il processo di produzione: FakeApp⁷⁴. Attraverso questa si sarebbe potuto generare un video *deepfake* anche con una limitata conoscenza del deep learning e del linguaggio di programmazione⁷⁵ (Matern, Riess et al., 2019) perché riuniva gli algoritmi complessi sviluppati da deepfakes in un’interfaccia user-friendly. Quest’applicazione portò quindi i produttori di *deepfake* a sviluppare con sempre maggiore frequenza video parodici o volti a dimostrare le possibilità aperte dalla tecnologia, video quindi come

⁷² A. Hern, ‘Deepfake’ face-swap porn videos banned by Pornhub and Twitter, The Guardian (Online), 2018, <https://www.theguardian.com/technology/2018/feb/07/twitter-pornhub-ban-deepfake-ai-face-swap-porn-videos-celebrities-gfycat-reddit> (ultima consultazione il 14 febbraio 2023).

⁷³ M. Westerlund, *The Emergence of Deepfake Technology: A Review*, in “Technology Innovation Management Review”, n.9, vol. 11, 2019, pp. 39-52.

⁷⁴ J. Burkell, C. Gosse, *Nothing New Here: Emphasizing the social and cultural context of deepfakes*, in “First Monday”, n. 24, vol. 12, 2019.

⁷⁵ F. Matern, C. Riess, M. Stamminger, *Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations*, in “IEEE Winter Applications of Computer Vision Workshops”, 2019, pp. 83-92.

Princess Leia CGI | DEEPFAKES REPLACEMENT di Derpfakes in cui, grazie a FakeApp, il volto di Carrie Fisher viene sostituito a quello in VFX di *Rogue One: A Star Wars Story* (Gareth Edwards, 2017). Lo stesso Derpfakes mette in luce il potenziale del software e la sua caratteristica leggerezza nella descrizione del video: “la clip superiore è un video originale di *Rogue One*. La clip inferiore è un *deepfake* più veloce, meno costoso e più facile. Nonostante la perdita di qualità, tenete conto che questo fake è stato fatto su uno schermo standard per PC e completato nel tempo che richiede guardare un episodio de *I Simpson*”. Nel corso del tempo Fake App è stata sempre meno utilizzata a favore di software come FaceSwap e da un sito come Deepfakes Web⁷⁶, che permette di produrre dei video deepfake online senza dover scaricare il software, semplicemente pagando una commissione e facendo l’upload del video su cui si vorrebbe inserire il volto e di quello/quelli da cui il volto lo si vorrebbe estrarre, ma rappresenta comunque uno spartiacque fondamentale nella diffusione dei video deepfake.

Il software più utilizzato per la loro produzione è, oggi, DeepFaceLab. Prodotto dall’utente russo di Github Iperov, indicato come il miglior software per produrre deepfake e disponibile su Github, DeepFaceLab è stato pubblicato verso la fine del 2018 ed è l’applicazione utilizzata dalla maggior parte dei creatori di deepfake⁷⁷, sia su Youtube che nel mondo del porno, nonostante, a differenza di FakeApp, non sia dotato di

⁷⁶ <https://deepfakesweb.com>

⁷⁷ Con questa affermazione faccio in particolare riferimento alle dichiarazioni rilasciate da produttori come Ctrl Shift Face, Salvo Compromesso e Gerardo Bosso, che hanno affermato di utilizzare DeepFaceLab e al saggio *Trust No One* (Grothaus, 2022).

un'interfaccia grafica che ne renda immediato l'utilizzo. La ricerca si sta tutt'oggi concentrando sullo sviluppo di software sempre più sofisticati: nel 2020, per esempio, i Disney Research Studios hanno comunicato di aver sviluppato un algoritmo in grado di migliorare la risoluzione dei deepfake, creando un volto più dettagliato, fotorealistico e meno artificioso⁷⁸ e, attualmente, la nuova frontiera della tecnologia sembra essere DeepFaceLive, un software che permette di produrre degli scambi di volto in real-time.

Cheapfake vs deepfake

Cheapfake (shallowfake)

I video *deepfake* sono solo una componente del più grande campo della manipolazione audiovisiva che include tanto questa pratica, basata, come si è ribadito più volte, sull'intelligenza artificiale, quanto i *cheapfake*, falsi prodotti grazie a software convenienti, accessibili o addirittura nessun software e che usano "tecniche convenzionali, come *speeding*, *slowing*, *cutting*, *re-staging*, o footage ri-contestualizzato"⁷⁹⁸⁰.

⁷⁸ J. Naruniec, L. Helming, C. Schroers, R.M. Weber, *High-Resolution Neural Face Swapping for Visual Effects*, in "Eurographic Symposium on Rendering 2020", 2020, n. 34, vol. 39.

⁷⁹ B. Paris, J. Donovan, *Deepfakes and cheap fakes: the manipulation of audio and visual evidence*, Data & Society 2019 (traduzione mia)

⁸⁰ La precedente distinzione attuata per i *deepfake* in *testi deepfake*, *audio deepfake*, *immagini deepfake* e *video deepfake* potrebbe essere riproposta anche per i *cheapfake* ma, concentrando la tesi sui *deepfake audiovisivi*, si preferisce concentrarsi su quest'ultima declinazione dei *cheapfake*.

La manipolazione fotografica esiste più o meno dalla nascita della fotografia stessa⁸¹ e la falsificazione del frame o del profilmico non sono certo nate con l'avvento del digitale né tanto meno con quello dell'intelligenza artificiale: ciò che sembra essere unico per i *deepfake* e soprattutto per i *cheapfake* (anche conosciuti col termine *shallowfake*⁸²) è però la facilità con cui questi possono essere prodotti e distribuiti grazie a strumenti che sono unici all'interno dell'era digitale. A differenza dei *deepfake*, i quali richiedono il download di software non intuitivi e, spesso, l'aiuto di un tutorial o il ricorso a delle istruzioni scritte per la semplice installazione, un laptop o uno smartphone hanno spesso già integrati gli strumenti necessari per creare un *cheapfake*.

I *cheapfake*, esattamente come i *deepfake*, creano una narrazione che è fuorviante e che non ha nessun riscontro nella realtà ma, differentemente dai *deepfake*, che fanno affidamento sull'IA, possono essere prodotti con strumenti a disposizione di chiunque, software di editing come Adobe Premiere Pro, Final Cut, iMovies per i video, o i sopraccitati Photoshop e PiZap per le immagini ecc.. Il termine "cheap" (it. scarso/economico) fa infatti, sì, riferimento al fatto che a causa del lavoro manuale e non

⁸¹ Si veda MIT

⁸² Il termine *shallowfake* è stato coniato da Sam Gregory dell'organizzazione per i diritti umani WITNESS, il quale fa esplicito riferimento a video che sono stati manipolati con strumenti di editing basilari o intenzionalmente messi fuori contesto, generalmente progettati per, utilizzando una definizione data da Ajder, Patrini et al., approfittare dei bias cognitivi di un individuo che possono risolversi in un danno alla reputazione di una specifica persona anche se il falso è di scarsa qualità". In questa sede si è però scelto di usare il termine *cheapfake* a causa del maggior utilizzo fattone in letteratura e perché ci sembra che il termine *cheap* ("a poco prezzo", "conveniente", "scadente") enfatizzi il carattere economico dei mezzi di produzione a buon mercato e la qualità grezza del prodotto finale, differentemente dal termine *shallow*, per lo più traducibile con il più generico "superficiale".

automatizzato, molti di questi media possono apparire come grezzi, scarsamente realistici, ma soprattutto al fatto che possano essere prodotti con strumenti (computer, software) poco costosi e facilmente accessibili dalla maggior parte della popolazione. Se, congiuntamente a questi strumenti, si prendono in considerazione la massiva quantità di dati visuali relativi non solo alle star ma a molti degli utenti del web, è facile comprendere come chiunque sia in possesso di un computer abbia disposizione tutto l'occorrente per produrre un falso e distribuirlo con facilità attraverso i social media.

I *cheapfake*, sostanzialmente, sono in grado esattamente quanto i *deepfake* di creare una narrazione che possa risultare fuorviante: esattamente come i *deepfake* mostrano qualcosa che non è mai avvenuto in realtà, ma a differenza dei *deepfake*, i quali devono far ricorso all'intelligenza artificiale, un *cheapfake* fa affidamento esclusivamente sul lavoro manuale del suo produttore e sugli strumenti di editing tradizionali, comportando così che le alterazioni del video, dei cut rudimentali ecc. siano potenzialmente ascrivibili alla scarsa abilità dell'editor, alla semplicità dei mezzi di ripresa o montaggio e non a un tentativo di manipolazione dello spettatore. Questi tipi di messa in scena e ri-contestualizzazione potrebbero risultare più difficili da individuare poiché non sono presenti pixel fuori posto o manipolazioni evidenti del corpo del soggetto inquadrato.

Una serie di esempi tratti dalla letteratura attorno al tema possono essere utili per la comprensione del fenomeno, dei suoi aspetti estetici e della potenziale velocità di circolazione dei *cheapfake* negli ambienti mediali,

nonostante i forensi abbiano spesso velocemente identificato i video come falsi⁸³.

- Nel febbraio del 2018 un uomo uccise quattordici studenti della Marjory Stoneman Douglas High School a Parkland, Florida, un episodio che scatenò un lungo dibattito politico negli Stati Uniti attorno alla facilità di accesso alle armi. Il 17 febbraio dello stesso anno molti dei sopravvissuti all'attacco si ritrovarono fuori dalla Broward County Courthouse a Fort Lauderdale (Florida) per manifestare contro una politica che sembrava non tutelarli. Una delle sopravvissute, una ragazzina di otto anni chiamata Emma González diede un intenso discorso di undici minuti. Un discorso che divenne rapidamente virale e la ragazzina divenne un simbolo per tutti coloro che si battevano per il controllo sulle armi negli Stati Uniti. Qualche giorno dopo il discorso, però, divenne virale una GIF nella quale la ragazza, con la testa rasata e indossando un top stretto e nero, stringeva una copia della costituzione degli USA nelle sue mani, per poi stracciarla in due prima di gettare le due metà e ripiegare le sue braccia sul suo petto in maniera sprezzante. Quella all'interno della GIF è realmente Emma González, col suo corpo, il suo volto e le azioni raffigurate sono davvero azioni che lei ha commesso: la ragazza stava davvero stracciando un grosso foglio di carta, ma quel foglio non era la costituzione statunitense, bensì un enorme poster con un mirino disegnato. Gonzales aveva filmato il video per un photoshoot su Teen Vogue alcuni giorni prima, questo è finito nelle

⁸³ A. Smith, K. Mansted, *Weaponised deep fakes*, 2020, <https://www.aspi.org.au/report/weaponised-deep-fakes> (ultima consultazione il 14 febbraio 2023).

mani di qualcuno che aveva interesse nello screditare l'immagine della ragazzina e che ha utilizzato dei software di montaggio per inserire digitalmente la costituzione degli U.S.A per mistificare le sue azioni. Questo cheapfake era estremamente realistico, tanto che fu ripostato centinaia e centinaia di volte nei social media di persone pro-armi e di destra prima che i fact-checkers potessero constatarne l'inautenticità.

- Nel maggio del 2019 divenne virale un video falsificato di Nancy Pelosi che faceva apparire il membro del partito democratico statunitense come ubriaca. “Il video di Pelosi, che ottenne più di 2.5 milioni di visualizzazioni su Facebook e che fu condiviso più volte da noti leader politici era in realtà un video *cheapfake* perché il suono era semplicemente stato rallentato a circa tre quarti della velocità” (Greengard, 2020).



Comparazione tra il video cheapfake e il video originale ritraente Nancy Pelosi.

- Nel novembre 2018 Jim Acosta, un inviato alla Casa Bianca per la CNN, fu coinvolto in un aspro scambio di battute con il presidente degli Stati Uniti Donald Trump. Quando una stagista si è avvicinata per sottrargli il microfono, Acosta ha abbassato il braccio dicendo “mi scusi, signora” per poi continuare le sue domande. Il pass di Acosta che gli consentiva l’accesso alla Casa Bianca fu sospeso “fino a nuova comunicazione” e Sarah Sanders, addetto stampa alla Casa Bianca, giustificò la decisione accusando il giornalista di aver “messo le mani addosso a una giovane donna che provava a fare il suo lavoro” (Tobias, 2018). Il giorno successivo, un video in cui sembrava che Acosta avesse inferto un colpo alla stagista fu fatto circolare dal sito Info Wars e venne ritwittato dalla pagina ufficiale della Casa Bianca così da giustificare la decisione di sospendere Acosta. Quando divenne chiaro che il video fosse stato manipolato la CNN minacciò di adire le vie legali e il pass di Acosta venne ripristinato. Alcune pubblicazioni, cavalcando l’ondata di paranoia attorno ai *deepfake*, hanno sostenuto che il video fosse un chiaro esempio di questa pratica (Woolley, 2020): da lì a poco si sarebbe capito come questo fosse semplicemente stato accelerato per far apparire il tentativo di Acosta di afferrare il microfono come ostile. Questa clip non era un *deepfake* bensì un *cheapfake*.

- *The Hillary Song* è un video di 3 min. e 35 secondi, ed è un estratto da un evento organizzato dalla WWE. Dwayne Johnson, wrestler e attore noto ai più come The Rock, è seduto su una sedia nel mezzo di un ring, attorno a lui lo stadio che ospita l’evento è affollato e festante. The Rock non è però al centro del ring per combattere un match. È seduto vicino a un microfono

sollevato da un braccio meccanico che lo mantiene vicino alla sua bocca. The Rock non può tenerlo sollevato, perché sta strimpellando una chitarra. L'uomo dice al pubblico: "C'è una donna molto speciale a cui The Rock vorrebbe cantare una canzone questa sera" – il pubblico urla di gioia – "e vorrei portarvela qui. Vuoi venire fuori, per favore? Non essere timida. Vieni qui. Voglio solo cantarti una canzone". Al che The Rock guarda alla sua destra e sorride. La donna a cui voleva fare la serenata e a cui chiedeva di uscire è intervenuta, ed è Hillary Clinton.

Dopo uno stacco, vediamo infatti la moglie di Bill Clinton avvicinarsi a un podio, dietro di lei sono presenti delle bandiere statunitensi, mentre The Rock continua a strimpellare. La clip è stata chiaramente estratta da un video risalente alla campagna presidenziale di Hillary Clinton del 2016. The Rock continua: "Ahhh... eccoti qua, sei così bella". Mentre la folla urla e The Rock continua a suonare lo strumento si può vedere la bocca di Clinton mimare le parole "Thank you" mentre sorride al pubblico. Il video torna poi a The Rock che afferma di voler cantare alla donna una canzone speciale per lei, una cover di uno dei suoi artisti preferiti, Eric Clapton. Inizia quindi a strimpellare "Wonderful Tonight". Da questo momento in poi si passa a un montaggio alternato della serenata di The Rock e del mezzobusto della Clinton. The Rock non canta però le precise parole del testo di Eric Clapton, ma ne cambia qualcuna come se fosse diretta specificamente per la donna, come, ad esempio "she brushes her short hair". The Rock continua a cantare: "and then she asks me: do I look alright?". The Rock sembra trattenere una risata e dice "No beee-yatch, sei orribile questa sera". Tutto l'auditorium scoppia in una fragorosa risata: la parodia della canzone di Clapton sembra funzionare. Hillary Clinton

guarda straniata verso terra mentre sembra mimare delle parole con la bocca. Successivamente The Rock comanda alla Clinton di non andare da nessuna parte e continua “abusi di tutto il tuo potere, fai perdere il tempo a tutti. Ti vesti come una battona... e non del tipo costoso” – in sottofondo i commentatori della WWE ridono di gusto – “quindi porta il tuo culo all’aeroporto, prendi un volo di non ritorno”. La camera ritorna sul volto della Clinton, e The Rock invita il pubblico a cantare con lui e il pubblico lo segue: “We said, “Beee-yatch, you look horrible tonight”. Il video giunge a termine con un montaggio alternato sul mezzobusto della Clinton, palesemente indispettita, e sul sorriso di The Rock mentre il pubblico urla.

In *Seeing isn't believing: The Fact Checker's guide to manipulated video*, una guida ai video manipolati redatta dal The Washington Post⁸⁴, vengono indicate tre possibili categorie di manipolazione nei video *cheapfake*:

1. Contesto mancante: rappresentazione falsata mediante l’estrpolazione e l’allontanamento del video dal suo contesto originario.
2. Editing ingannevole: omissione di uno o più frame del video originale o inserimento di nuovi.
3. Trasformazione capziosa: falsificare o manipolare il contenuto di un video.

Nel primo caso, un video può essere fatto circolare all’interno di un ambiente mediale senza le necessarie premesse per la corretta

⁸⁴ N. Ajaka, E. Samuels, G. Kessler, *Seeing isn't believing: The Fact Checker's guide to manipulated video*, <https://www.washingtonpost.com/news/fact-checker/wp/2016/11/22/the-fact-checkers-guide-for-detecting-fake-news/> (ultima consultazione il 14 febbraio 2023).

comprensione del contenuto (es. la risposta fornita durante un'intervista senza che sia stata ripresa la domanda). Nel secondo caso, l'editor manipolerebbe il contenuto del video eliminando dei *frame* chiave (per esempio accelerando o rallentando un movimento in modo da trasformare l'azione ripresa come nel caso del video *cheapfake* di cui è stato vittima Jim Acosta, come si vedrà tra poco) o un'intera porzione di video. Nell'ultimo caso, la guida fa invece riferimento alla manipolazione dei *frame*, alterati in modo che un elemento esterno possa sembrare parte integrante del profilmico (es. il titolo di una pagina di giornale modificato). Tutti questi casi ci rendono consapevoli della necessità di capire quindi, a livello tecnico, come vengano prodotti i *cheapfake*, poiché è in questo campo che si giocano le sostanziali differenze, e quindi i successivi differenti risvolti estetici, tra *deepfake* e *cheapfake*.

In *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence* Britt Paris e Joan Donovan segnalano le differenze che intercorrono tra queste due forme.

CHEAPFAKE (richieste poca esperienza e risorse tecnologiche)						
TECNICHE	Face swapping: Rotoscope	Speeding and slowing	Face altering/swapping	Speeding and slowing	Lookalike	Recontextualizing
TECNOLOGIE	After Effects/Adobe Premiere Pro	Sony Vegas Pro	Free real-time filter applications	Free speed alteration applications	In-camera Effects	Relabelling/Reuse of extant video

Questa tabella, pur adottando delle semplificazioni, conferma l'assunto per cui i *cheapfake* potrebbero essere facilmente prodotti da chiunque fosse in possesso dei rudimenti dell'editing, innanzitutto segnalando come i questi richiedano "poca esperienza e risorse tecnologiche" e soprattutto proponendo una serie di espedienti produttivi attuabili con pochi mezzi. La *ricontestualizzazione (recontextualizing)* fa riferimento al riuso di un video non alterato ma all'interno di un contesto differente rispetto a quello originale, così che questo assuma un valore differente o opposto. La tecnica del *lookalike* video prevede l'uso di un attore/attrice di simile corporatura, tratti somatici e chioma al soggetto che dovrà imitare. È una tecnica spesso utilizzata in ambito pornografico e che, paradossalmente, è anche alla base dei video *deepfake* ed è quindi una tecnica relativa alla produzione e non alla post-produzione, da cui il riferimento a effetti *in-camera*. Accelerazione (*speeding*) e rallentamento (*slowing*) possono essere ottenuti attraverso applicazioni apposite o software di editing rinomati come Sony Vegas pro, ma anche in non citati, si presume per semplicità, Adobe Premiere Pro, Final Cut, iMovies ecc. Un *face-swap* potrebbe essere anche prodotto attraverso After Effects, senza fare affidamento a tool di *deep learning*, ma lavorando, per esempio, *frame by frame* con un mascherino (per ottenere un risultato verosimile si dovrà però aver accumulato un certo grado di esperienza) oppure facendo ricorso a delle apposite applicazioni o filtri (come quelli presenti su Instagram) che però non garantiscono risultati ad alta fedeltà.

Deepfake

Le strade attraverso cui un *deepfake* può essere creato sono numerose, non esiste di fatto un'unica modalità codificata, sebbene si sia più volte specificato come il discrimine tra *cheapfake* e *deepfake* risieda nel ricorso all'intelligenza artificiale per la produzione del falso. Internet è inondato da software per creare *deepfake*⁸⁵ che possono essere usati da utenti con diverse capacità tecniche che variano dai novizi ai professionisti⁸⁶ ma non tutti questi software forniscono prestazioni equiparabili.

App per smartphone come Wombo o SpeakR permettono, sì, di produrre dei discreti video *deepfake* a partire da un'unica fotografia o da un semplice input vocale, ma le possibilità che questi vengano percepiti come verosimili sono estremamente basse. Risulta però essenziale un accenno alle tecnologie che sottostanno ai software (con particolare riferimento alle GAN) per comprendere i fattori innovativi della tecnologia, i suoi limiti attuali e il potenziale.

La novità dei video *deepfake* non risiede infatti in una possibilità, quella di innestare un volto su un corpo non proprio, potenzialmente già precedentemente appannaggio dei grandi studi di C.G.I ma, piuttosto, nella portata, nella scala e nella sofisticatezza di questa tecnologia che la possibilità di fare un *face-swap* la democratizza.

⁸⁵ Il sito deepfakesweb.com consente anche di creare un *deepfake* online attraverso Deepfakes App senza il bisogno di scaricare alcun software, semplicemente versando una piccola quota e mantenendo l'anonimato.

⁸⁶ A. Karandikar, V. Deshpande et al., *op.cit.* (traduzione mia).

Sarebbe a dire che

...mentre oggi i contenuti di questi nuovi media sembrano spesso gli stessi dei loro predecessori, non dobbiamo farci ingannare da questa somiglianza. La novità non risiede nel contenuto ma nei software utilizzati per creare, montare, vedere e distribuire il contenuto. Per cui, invece che guardare solamente all'output delle pratiche culturali *software-based*, dobbiamo considerare il software di per se stesso, dal momento che permette alle persone di lavorare con i media in un numero di modi che non hanno precedenti nella storia⁸⁷.

Le parole di Manovich sembrano tanto più vere nel momento in cui vengono applicate a software di intelligenza artificiale che riducono almeno parzialmente l'autonomia dell'utente (una partecipazione spesso sottostimata, come si vedrà nella sezione dedicata a DeepFaceLab), il quale dovrà limitarsi, nel caso di *deepfake* prodotti attraverso software intuitivi come FakeApp, a fornire il *training set* e ad impostare dei parametri preliminari, mentre software come DeepFaceLab consentono un maggiore grado di intervento al *deepfake designer*.

Se, a questo proposito, dal nostro punto di vista, è sovrastimata l'idea che quasi chiunque posseda un computer potrebbe produrre video falsi pressoché indistinguibili da media autentici⁸⁸ a causa delle numerose ore di esperienza da accumulare per produrre un video *deepfake* fotorealistico⁸⁹,

⁸⁷ L. Manovich, *Software Takes Command: extending the language of new media*, Bloomsbury USA Academic, New York 2013, p.85 (traduzione mia).

⁸⁸ J. Fletcher. *Deepfakes, Artificial Intelligence, and some Kind of Dystopia: The New Faces of Online Post-Fact Performance*. In "Theater Journal", 2018, vol. 70, n.4.

⁸⁹ Questa affermazione è supportata dalle interviste da me stesso attuate a diversi *deepfake designer* e disponibili all'indirizzo:

https://www.youtube.com/channel/UCUediZEeb_CUCIO3SaErpww.

della necessità di affidarsi a tutorial di vario tipo e di possedere computer con un'ottima capacità di calcolo oltre che dell'enorme quantità di dati visuali necessari perché il training abbia successo, è altrettanto vero che la tecnologia sta evolvendo rapidamente e che applicazioni per smartphone come il sopracitato Wombo consentono di produrre video verosimili (ma che difficilmente ingannerebbero lo spettatore) semplicemente facendo l'upload di un'unica fotografia del soggetto.

Numerosi utenti hanno usato app come FakeApp per rivisitare alcune produzioni hollywoodiane come nel video del 2018 *Solo | A Derpfakes Story* dell'utente YouTube *derpfakes*, che sovrapponeva il volto di Harrison Ford a quello di Alden Ehrenreich in *Solo: A Star Wars Story*⁹⁰ (Ron Howard, 2018) o come nei video di Ctrl Shift Face, il *deepfake artist* che conta più di 400.000 iscritti al suo canale Youtube e che ha totalizzato più di 50 milioni di visualizzazioni proponendo delle ucronie cinematografiche. In calce ai suoi video, Ctrl Shift Face indica DeepFaceLab come software utilizzato per la produzione e il *deepfake designer* Salvo Compromesso ha affermato come questo sia il software più utilizzato per la produzione di *deepfake*⁹¹.

Nel sopracitato *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence* Britt Paris e Joan Donovan riportano un'utile tabella che, come la precedente dedicata ai *cheapfake*, riassume le possibili tecniche di produzione dei *deepfake*.

⁹⁰ B. Paris, J. Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*, 2020

⁹¹ Faccio in particolare riferimento a un'intervista da lui rilasciatami disponibile al seguente link: <https://www.youtube.com/watch?v=LXmDdmYP16Q&t=258s>

DEEPPFAKE (richieste esperienza e risorse tecnologiche)				
TECNICHE	Virtual Performances	Virtual Performances	Voice Synthesis	Face Swapping/Lip-Synching
TECNOLOGIE	RNN (Recurrent Neural Network)/HM M (Hidden Markov Models)/LSTM (Long Short Term Memory Models)	GANs (Generative Adversarial Networks)	VDR (Video Dialogue Replacement models)	FakeApp/After Effects

Nel saggio Paris e Donovan spiegano come “Adobe After Effects consenta a chiunque abbia un computer con abbastanza potenza di calcolo e del tempo di creare dei falsi audiovisivi attraverso il *machine learning*”⁹². Nonostante le app siano numerose e i modi per creare un *deepfake* eterogenei, la forma più comune implica la generazione e la manipolazione di immagini umane⁹³, in particolar modo attraverso il *face-swapping*. Se i primi *deepfake*, *face-swap*, erano basati su una classe di sistemi di *deep*

⁹² B. Paris, J. Donovan, *op.cit.*, p. 14 (traduzione mia)

⁹³ Y. Mirsky, W. Lee, *The Creation and Detection of Deepfakes: A Survey*, arXiv 2020.

learning conosciuti come autoencoder⁹⁴⁹⁵, e sia possibile fare ricorso alle RNN, alle LSTM e, più spesso al Video Dialogue Replacement (come nel caso dei falsi generati su SpeakR) la generazione di *deepfake* attraverso il *face-swapping* è il risultato di un computer che apprende nel corso del tempo da dell'informazione che gli viene fornita (le immagini di volti di due persone) e genera contenuti inediti, la maggior parte delle volte facendo ricorso alle GAN.

GAN

Nonostante i video *deepfake* possano essere prodotti attraverso diverse tecniche che fanno affidamento sull'intelligenza artificiale⁹⁶, vengono prevalentemente realizzati facendo ricorso alle GAN. Le Reti Antagoniste Generative sono una tipologia di algoritmo generativo non supervisionato composto da due reti neurali, una chiamata *generator* e l'altra *discriminator*, che mirano a ridurre al minimo la differenza tra un output generato e ciò che viene presentato come un'"immagine reale"⁹⁷. Questo modello è stato definito "antagonista" poiché le due reti neurali vengono poste l'una

⁹⁴ N Schick, *op.cit.*

⁹⁵ Un autoencoder è una specifica rete neurale, per lo più progettata al fine di codificare un input in una rappresentazione compressa e fedele, dopodiché la decodifica in modo che l'input ricostruito sia il più simile possibile all'originale. Vedi: D. Bank, N. Koenigstein, R. Giryes, *Autoencoders*, arXiv:2003.05991.

⁹⁶ B. Paris, J. Donovan, *op.cit.* (traduzione mia)

⁹⁷ G.W. Dyer, *Le GAN e la mimesi*, in A. Barale (a cura di), *Arte e intelligenza artificiale: be my GAN*, Editoriale Jaca Book, Milano 2020, p.131.

contro l'altra⁹⁸ in un gioco simile a guardie e ladri, una competizione che le porta ad auto-migliorarsi autonomamente, e che si articola attorno alla domanda "quali sono i dati veri e quelli contraffatti?"⁹⁹. È lo stesso Ian Goodfellow, ideatore delle GAN insieme al suo team di ricerca, a spiegarlo.

Nel framework di *adversarial nets* proposto, il modello generativo è posto contro un avversario: un modello discriminativo che impara a determinare se l'esempio proviene dalla distribuzione del modello o dalla distribuzione dei dati. Il modello generativo può essere considerato come un team di contraffattori che tentano di produrre valuta finta e usarla passando inosservati, mentre il modello discriminativo è analogo alla polizia che prova a rilevare la contraffazione. La competizione in questo gioco porta entrambi i team a implementare i loro metodi fino a quando i dati contraffatti risultano indistinguibili dai dati genuini¹⁰⁰.

Per dar vita al gioco la GAN viene addestrata su una serie di dati iniziali (il cosiddetto *training set*), generalmente di dimensioni molto elevate: se le dimensioni del *training set* risulteranno essere troppo ridotte questo non conterrà informazioni sufficienti e risulterà estremamente confuso, portando a un risultato mediocre o nullo¹⁰¹. Nella preparazione del gioco, a ogni rete neurale viene dato accesso allo stesso *training set*. Ciascuna delle reti neurali supervisiona i dati e ne apprende la struttura, dopodiché a ciascuna delle due reti neurali viene fornito uno dei due ruoli. Lo scopo

⁹⁸ C. Moruzzi, *Alla ricerca della creatività: le GAN come paradigma dell'autonomia nel software per la composizione musicale*, in A. Barale (A cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020, p.156.

⁹⁹ A. Barale, *Arte e intelligenza artificiale: alcune domande*, in A. Barale (a cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020, p.10.

¹⁰⁰ I. Goodfellow, J. Pouget-Abadie, et al., *Generative Adversarial Nets*, arXiv, 2014, p.1 (traduzione mia).

¹⁰¹ A. Ridler, *Set di dati e decadenza: Fall of the House of Usher*, in A. Barale (a cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020, pp. 113-114.

della rete neurale *generator* è quello di ingannare il *discriminator* producendo dei dati indistinguibili da quelli forniti nel *training set*. Lo scopo della rete neurale *discriminator* è invece quello di indicare il falso, inviando dei *feedback* che consentano a *generator* di incrementare la propria prestazione, creando degli output sempre più simili ai campioni del *training set*¹⁰².

Questo processo avviene in seguito a diversi cicli di apprendimento¹⁰³. Una volta che *generator* ha creato la sua prima falsificazione mostra il dato a *discriminator*: se questo è il primo dato contraffatto (o tra le prime migliaia) *discriminator* capirà facilmente la falsificazione¹⁰⁴ ma, ogni volta che un dato verrà indicato come “falso”, *generator* rivedrà gli aspetti della sua falsificazione, riguardando i dati iniziali e comparando quelli autentici al proprio, identificando così quali siano le differenze, in modo da produrre una nuova falsificazione fino a quando non riuscirà a ingannare *inspector*.

I *deepfake* non sono però l'unico campo di applicazione delle GAN, nel campo dell'audiovisivo o meno. Negli ultimi anni, l'utilizzo delle GAN in qualità di generatrici di immagini artificiali altamente realistiche si è sviluppato in numerosi settori. I volti nelle immagini *deepfake* prodotti attraverso le GAN possono essere utilizzati per aprire dei profili social fake e un software come StyleGAN (capace di produrre milioni di volti artificiali al giorno) è estremamente richiesto dalle piattaforme di incontri sentimentali così da generare profili falsi credibili in caso di necessità di

¹⁰² N. Schick, *op.cit.*, p. 44.

¹⁰³ A. Ridler, *op.cit.*, p. 112.

¹⁰⁴ M. Grothaus, *op.cit.*

bilanciare profili maschili e femminili, mantenendo così inalterato l'appeal della piattaforma¹⁰⁵.

Le GAN lavorano in particolar modo con i volti umani: trovano infatti applicazione nel campo del riconoscimento dei volti, dei sistemi di *face recognition* che si possono trovare, ad esempio, negli aeroporti e nei vari sistemi di sorveglianza. Ma il loro uso non è esteso sollo alla produzione e al riconoscimento di volti umani. Grazie agli ESR-GAN (Enhanced Super-Resolution Generative Adversarial Networks) i film delle origini possono essere visualizzati in 4K e 60 frame al secondo¹⁰⁶.

DeepFaceLab

La medesima tecnologia è alla base del funzionamento di DeepFaceLab, software reso disponibile su Github alla fine del 2017 dall'utente Iperov. Facendo in particolar modo riferimento al saggio dedicato alle prospettive della pratica *Trust No One: Inside the World of Deepfakes* vengono qui ripercorsi i passaggi cruciali, funzionali al fine di analizzare come venga prodotto lo scambio di volto nella maggioranza dei video *deepfake*.

Innanzitutto, sono utili due considerazioni preliminari: differentemente da FakeApp, DeepFaceLab non ha una grafica *user-friendly* che ne renda immediato l'utilizzo, minando il principio per cui "chiunque può produrre

¹⁰⁵ R. Eugeni. *Capitale algoritmico: cinque dispositivi postmediali (più uno)*, Scholè, Brescia 2021.

¹⁰⁶ Si vedano, a titolo esemplificativo, i video [4k, 60 fps] *A Trip to New York City in 1911* e [4k, 60 fps] *Apollo 16 Lunar Rover "Gran Prix" (1972 April 21, Moon)* sul canale di Denis Shiryayev.

un video *deepfake*”¹⁰⁷. La seconda considerazione è inerente al fatto che, per leggerne i lineamenti, e quindi compiere al meglio lo scambio, DeepFaceLab richiede delle inquadrature ravvicinate al volto e in alta definizione. Sono quindi preferibili i piani americani, i mezzi busti e, soprattutto, i primi piani in risoluzione 720p o 1080p.

Una volta lanciato il programma, l’utente potrà selezionare il *video sorgente* (il video nel quale è presente il volto della celebrità o della persona che si vuole estrarre e innestare su un corpo estraneo) e il *video destinazione* (il video nel quale è presente il corpo della celebrità o della persona su cui il volto verrà innestato). Successivamente, si aprirà un’altra serie di file che vengono a formare quello che viene definito *faceset* della celebrità di cui fare il *deepfake*, creato da DeepFaceLab estraendo tutti i frame in cui il soggetto appare. Il *deepfake designer* eseguirà poi queste operazioni di creazione del *faceset* anche sul volto che appartiene al corpo del donatore nel video di destinazione. Non sarà necessario utilizzare la totalità di questi frame – il che potrebbe allungare il processo di *training* (la fase in cui il software elabora lo scambio di volto) – ma limitarsi a mantenerne un numero sufficiente in modo da ottenere una panoramica completa delle posizioni del volto¹⁰⁸: per questa motivazione, come video-fonte sono preferibili delle videointerviste, che riportano diversi piani frontali e profili in alta definizione.

¹⁰⁷ *Contra. Everybody Can Make Deepfakes Now*, in Two Minute Papers Youtube Channel, 2020, <https://www.youtube.com/watch?v=mUfJOQKdtAk> (ultima consultazione il 14 febbraio 2023).

¹⁰⁸ *Come creare un deepfake con DeepFaceLab [ITA] – Tutorial completo*, in DeepFake ENG ITA Youtube Channel, 2021, <https://www.youtube.com/watch?v=z29Oveu56eA&t=1505s> (ultima consultazione il 10 febbraio 2023).

Successivamente verrà tracciato manualmente il contorno del volto della celebrità e del donatore in alcuni dei frame che formano il loro rispettivo *faceset* così da consentire a DeepFaceLab di comprendere quale porzione del volto rimpiazzare. Una volta che la scontornatura è stata fatta, l'utente aprirà un altro gruppo di file che consentiranno a DeepFaceLab di analizzare migliaia di frame in entrambi i *faceset*, concentrandosi sulle aree evidenziate dei rispettivi volti. Più immagini saranno presenti, più analisi il software farà¹⁰⁹.



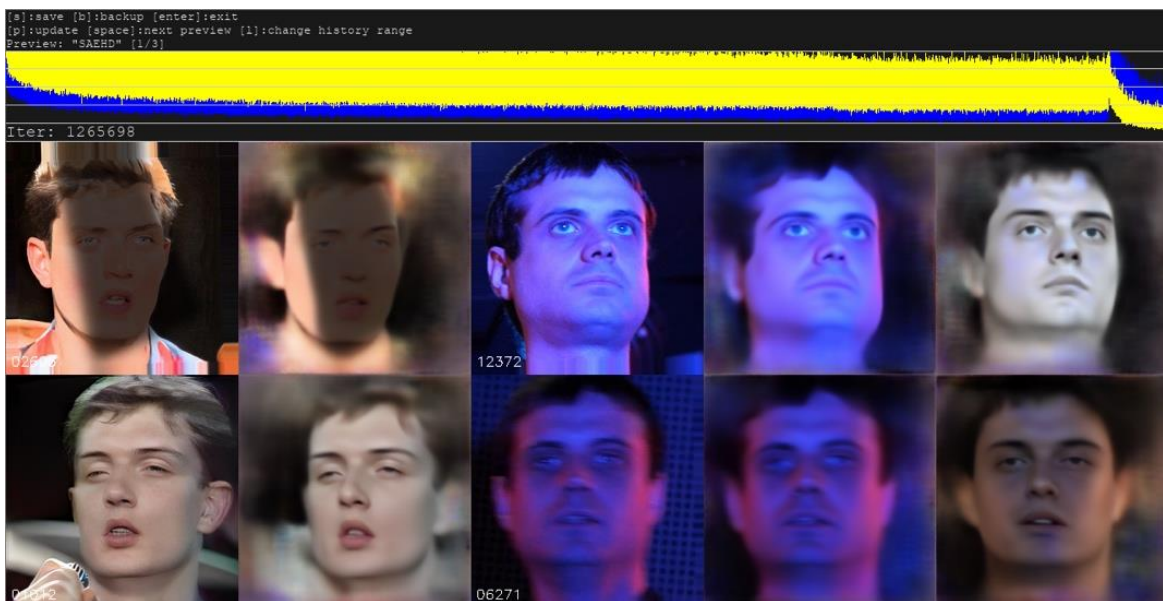
Estratto dalla fase di *training* di DeepFaceLab (volto scontornato) del progetto #iancortisalive¹¹⁰, un tentativo di resurrezione digitale del defunto frontman dei Joy Division fornito dal *deepfake designer* Salvo Compromesso.

DeepFaceLab userà quindi le sue analisi del volto per costruire quello che viene chiamato “il modello”. Questo modello è quello che la GAN,

¹⁰⁹ M. Grothaus, *op.cit.*

¹¹⁰ Video disponibile al seguente link: <https://www.youtube.com/watch?v=7DFD4k7KMx8> (ultima consultazione il 14 febbraio 2023).

nella suddetta fase di *training*, userà per generare la sua versione artificiale del volto della celebrità e mapparla accuratamente sul volto del corpo donatore. In altre parole, la rete neurale imparerà a riprodurre realisticamente il volto della celebrità fusa con le espressioni facciali del donatore grazie all'esame dei *faceset*. La fase di training dei modelli può attraversare migliaia e migliaia di iterazioni che potranno richiedere giorni. Più sarà alto il numero di iterazioni, meglio *forger* imparerà a riconoscere dove sta sbagliando, così da ingannare l'ispettore e introdurre delle alterazioni nel tentativo successivo. Il *deepfake designer* potrà inoltre mettere in pausa il *training* del modello in qualsiasi momento per vedere l'iterazione più recente mappata sul corpo del donatore e, se soddisfatto, potrà fermare il *training* proprio in quel momento e fondere il volto della celebrità con il corpo donatore. In ultimo, tutto ciò che il *deepfake designer* dovrà fare sarà esportare il video con la fusione tra il volto della celebrità sul corpo del donatore¹¹¹.



¹¹¹ *Ibid.*

Estratto dalla fase di *training* di DeepFaceLab del progetto #iancurtisalive, un tentativo di resurrezione digitale del defunto frontman dei Joy Division Ian Kevin Curtis fornito dal *deepfake designer* Salvo Compromesso.

Conclusioni capitolo

In questo capitolo si è potuto discutere della natura transmediale del termine *deepfake* a partire dalla sua etimologia (*deep learning* + *fake*, da intendersi come “falso prodotto attraverso il *deep learning*”), ascrivibile non solo a video manipolati attraverso l’intelligenza artificiale (come alcuni articoli hanno lasciato intendere) ma anche a testi, audio, immagini e a tutti quei media prodotti attraverso l’intelligenza artificiale il cui ottimale è la mimesi totale con un medium non manipolato, utilizzando diversi studi di caso per ogni tipologia, così da vagliare lo stato dell’arte e il loro grado di integrazione nel tessuto mediale.

Si è successivamente segnalata, facendo in particolar modo riferimento al saggio *Deepfakes and Cheapfakes: the manipulation of audio and visual evidence* di Britt Paris e Joan Donovan la cruciale distinzione nelle tecniche di produzione tra *deepfake* (medium manipolato attraverso l’intelligenza artificiale il cui ottimale è la mimesi totale con un medium non manipolato e che implica un’alterazione della realtà) e *cheapfake* (medium manipolato attraverso software di editing basilari il cui ottimale è la mimesi totale con un medium non manipolato che implica un’alterazione della realtà), apportando degli esempi che spieghino come, allo stato attuale, questi ultimi abbiano comportato problematiche identiche se non maggiori nella produzione di *fake news* di vario tipo provocando effetti tangibili nella realtà e la necessaria smentita degli organi ufficiali, essendo più facilmente producibili e distribuibili da attori non in possesso di specifiche competenze nell’ambito dell’editing.

Ci si è poi concentrati specificamente su come venga prodotto un video *deepfake* affermando come non sia tanto la possibilità di produrre un *face-swap* la portata rivoluzionaria di questa tecnologia, quanto il suo democratizzare tale possibilità in ambito amatoriale, rendendo così possibile a una pletera di attori diversi, in luoghi privi di una regolamentazione chiara come il web, di sostituire con relativa semplicità il volto di una persona con quello di un'altra all'interno di un video.

Per far ciò, si è innanzitutto visto che cosa siano le Generative Adversarial Networks e come consentano di produrre uno scambio di volto nei software *gan-based*, i software più usati per la produzione dei video *deepfake*. Nello specifico si sono quindi studiati i principi del funzionamento di DeepFaceLab, il software maggiormente utilizzato dai *deepfake designer* per la produzione dei video *deepfake*, mostrandone potenzialità e limiti, ed evidenziando come l'intervento umano sia ancora cruciale, in particolar modo nella selezione del *training set*.

Tornare alla mimesis

Sommario

Nel seguente capitolo, in seguito a una breve trattazione della figura del doppio, del clone e dell'avatar come polo centrale nella riflessione tra corpo reale e digitale, si proporrà un framework teorico all'interno del quale muovere l'analisi dei casi di studio che attinge alla filosofia dell'immagine e, in particolar modo, agli studi sulla *mimesis*. La tripartizione che verrà proposta in *mimesis come indebolimento ontologico*, *mimesis come raddoppiamento chiarificante* e *mimesis come incremento d'essere* non va però intesa come una rigida griglia classificatoria ma, piuttosto, come un flessibile strumento d'analisi che consenta di guardare al rapporto tra *deepfake* e *mimesis* e alle possibili interpretazioni di un video *deepfake*.

Doppi, gemelli, cloni

Avatar

Nel 2015, col fine di protestare contro l'approvazione da parte del senato spagnolo di alcuni discussi decreti sulla sicurezza cittadina, gli attivisti dell'organizzazione spagnola No Somos Delito hanno pianificato la prima protesta olografica della storia, chiamata in maniera evocativa "Hologramas por la libertad" (ologrammi per la libertà), innestando i

propri ologrammi in marcia all'interno della città, di fronte al parlamento¹¹². L'azienda Synthesia è stata fondata nel 2017 e ha ottenuto sempre maggiore notorietà dall'anno della sua fondazione grazie alla proposta di avatar personalizzati e producibili previo consenso del soggetto per utilizzarli in video di qualsiasi natura¹¹³. L'uscita nelle sale di *Matrix: Resurrections* (Lana Wachowski, 2021) è stata preceduta il 30 novembre 2021 dal lancio di centomila avatar *non-fungible token* acquistabili sulla piattaforma Nifty's: la richiesta da parte dei fan è stata tale che il sito è andato in *crash* in pochi minuti e le unità sono esaurite nel lasso di pochissimo tempo¹¹⁴. In un'arena londinese a loro dedicata, seguendo la scia dei live olografici di Tupac e Whitney Houston, nel 2022 i componenti degli ABBA sono tornati a calcare i palchi assieme per supportare *Voyage*, il loro ultimo album, ma lo hanno fatto sotto forma di ologrammi, accompagnati da dieci musicisti in carne e ossa¹¹⁵.

Questi sono solo alcuni esempi utili a spiegare come il clone, l'avatar, l'androide siano figure che si stanno imponendo con sempre maggior evidenza all'interno di dinamiche socio-culturali eterogenee che spaziano dalla produzione artistica all'attivismo politico, dall'editing genetico alla promozione pubblicitaria, intersecando studi nell'ambito dell'intelligenza

¹¹² Nosomos Delito (Youtube Channel), *Hologramas por la Libertad*, 2016, <https://www.youtube.com/watch?v=JQ14t8bAOVA>

¹¹³ <https://www.synthesia.io/>

¹¹⁴ Alcune unità sono ancora disponibili sulla piattaforma grazie al secondary market: <https://niftys.com/thematrix?page=2>

¹¹⁵ Dal sito ufficiale: <https://abbavoyage.com/>

artificiale, della teoria del cinema, del post-umano, della robotica sociale ecc¹¹⁶.

Nel suo saggio *Alla soglia dell'immagine. Da Narciso alla realtà virtuale*, Andrea Pinotti fa esplicito riferimento al *doppelgänger* come figura archetipica all'interno della letteratura germanica successivamente ripresa dall'espressionismo cinematografico tedesco in film come *Lo studente di Praga* (Stellan Rye, 1913) e soprattutto *Metropolis* (Fritz Lang, 1927), per affermare come il sosia, il doppio e il gemello rappresentino “una costellazione di figure cugine che le mitologie nelle diverse culture e la storia delle letterature mondiali hanno incessantemente variato e declinato”¹¹⁷.

Di questa costellazione, il cinema si è fatto incubatore, traendo spunto dalla letteratura, dal folklore e dalla mitologia per narrare storie di doppi, di androidi e di copie che si ribellano al proprio creatore. Nel cinema moderno e contemporaneo il motivo ricorrente del clone e della clonazione è stato infatti utile per rielaborare l'umano, integrandolo alle tecnologie e più genericamente al digitale¹¹⁸. Film fondamentali all'interno dell'immaginario collettivo come *Blade Runner* (Ridley Scott, 1982) e il suo sequel *Blade Runner 2049* (Denis Villeneuve, 2017), horror e sci-fi d'autore come *Atto di forza* (Paul Verhoeven, 1990), i più recenti *Ex Machina* (2015) e

¹¹⁶ Si veda, a puro titolo esemplificativo: R. Marchesini, *Il tramonto dell'uomo. La prospettiva post-umanista*, Dedalo, Bari 2009; A. Caronia, A. Tursi (a cura di), *Filosofie di Avatar. Immaginari, soggettività, politiche*, Mimesis, Milano 2010; S. Turkle, *Insieme ma soli. Perché ci aspettiamo sempre più dalla tecnologia e sempre meno dagli altri*, Einaudi Torino, 2012.

¹¹⁷ A. Pinotti, *Alla soglia dell'immagine. Da Narciso alla realtà virtuale*, Einaudi, Torino 2021, pp. 32-33.

¹¹⁸ P. Bertetto, *Lo specchio e il simulacro. Il cinema nel mondo diventato favola*, RCS Libri, Milano 2007, p. 11.

Annientamento (2018) di Alexander Garland, *Non lasciarmi* di Mark Romanek (2010), la recentemente cancellata serie adattamento di *Il mondo dei robot* (Michael Crichton, 1973), *Westworld – Dove tutto è concesso* (Jonathan Nolan, Lisa Joy, 2016-2022) o *Everything, Everywhere All at Once* (Daniel Kwan, Daniel Scheinert, 2022) sono tutti film (e l'elenco potrebbe continuare per diverse pagine) percorsi trasversalmente da doppi dell'umano, automi, replicanti che producono una crisi delle identità consolidate "a favore di una frammentazione dei soggetti e di una polimorfità del visibile" in cui "il mondo esterno appare sempre di più nei modi dell'illusione e del potenziale"¹¹⁹.

Corpi digitali

Il cinema non è però stato solo un semplice generatore di storie e di incubi che avessero al suo centro l'archetipo del doppio, ma piuttosto il medium che ha innescato un processo di derealizzazione del corpo, generando una lunga serie di alter ego dell'umano, corpi simulacrali progressivamente sempre più autonomi dal loro originale. Al medium cinematografico e a quello fotografico, più che al digitale, sarebbe infatti da ascrivere secondo Christian Uva il delitto della realtà fisica dei corpi messi in immagine¹²⁰; il cinema sarebbe quindi il punto di partenza (essendo connaturato alla sua stessa natura di immagine in movimento il

¹¹⁹ *Ibid.*

¹²⁰ C. Uva, *Ultracorpi. L'attore cinematografico nell'epoca della digital performance*, Bulzoni Editore, Roma 2011.

rendere il “corpo in quanto immagine” già di per sé un replicante) di un processo di “avatarizzazione” più ampio che precede di gran lunga l’avvento del pixel ma che oggi investe diversi strati della società, in una parabola che ha finito per riguardare direttamente ognuno di noi.

Per Paul Virilio

Si direbbe già che la clonazione umana stia diventando, per una parte del pubblico contemporaneo, un’operazione semplice quanto lo era farsi fare il ritratto da un fotografo nell’Ottocento. O anche, dopo il 1895, quanto pagare il biglietto per sbirciare su uno schermo il bambino dei Fratelli Lumière che mangia con avidità la pappa. [...] Se secondo le leggi attuali, le quali dovrebbero proteggere le libertà individuali, siamo, di fatto, proprietari del nostro corpo ma *anche della sua immagine*, il nostro prolifico ambiente audiovisivo ci ha portati da molto tempo a non preoccuparci più veramente di quelle molteplici apparenze di noi stessi che ci sono trafugate, che sono sottratte, esplorate, manipolate a nostra insaputa da parte di stati maggiori sconosciuti... i quali si disputano di nascosto i nostri cloni ottici...¹²¹.

Già l’avvento della fotografia e del cinema fotochimico avrebbero innescato un processo di separazione dell’immagine dal corpo causante disagio nell’uomo posto di fronte alla macchina fotografica o alla macchina da presa simile a quello provato dinnanzi alla propria immagine allo specchio, ma amplificato dalla trasportabilità e perenne replicabilità di quest’immagine, definitivamente scissa dall’obbligo della presenza del corpo¹²². Questo processo sarebbe stato portato alle estreme conseguenze poi dall’imaging digitale e dai software per la manipolazione

¹²¹ P. Virilio, *La bomba informatica*, Raffaello Cortina Editore, Milano 2000, pp. 27-28.

¹²² W. Benjamin, *L’opera d’arte nell’epoca della sua riproducibilità tecnica*, Giulio Einaudi Editore, Torino 2011, pp. 21-22.

dell'immagine che avrebbero avuto la conseguenza di semplificare e snellire progressivamente la possibilità di generare un doppio, un simulacro sempre più autonomo, fino a renderlo, oggi, fragile, frantumato, manipolabile da chiunque.

Per Luca Poma l'interconnessione tra il nostro corpo reale e quello digitale sarebbe sempre più evidente e ogni corpo reale che si avvale degli strumenti di comunicazione digitale sarebbe per tanto in fase di trasformazione in un *corpo digitale*¹²³, per cui tutti possederemmo, più o meno consapevolmente, un nostro *doppelgänger*, un corpo intangibile che può avere una sua consistenza e profondità, includendo le informazioni biometriche, sanitarie, psichiche, genetiche, relazionali e che può essere utilizzato per vari scopi, tra cui l'autore si concentra in particolar modo su quelli commerciali, sfruttabili anche in seguito alla nostra morte.

All'interno di questo scenario in cui la formazione di un doppio include non solo le informazioni visuali, ma il complesso di dati frutto della presenza e dell'esperienza online è inclusa la precisa rappresentazione o la stilizzazione del nostro stesso corpo, risultato della disgregazione e riaggregazione degli infodati che sono in grado di riconfigurare o riprodurre la nostra immagine sotto forma di avatar in cui qualcuno soffierebbe una "scintilla di vita" che farebbe nascere un'entità nuova capace di interagire e creare relazioni, potenzialmente anche dissociate

¹²³ Secondo la definizione fornita dallo stesso autore (che troviamo in L.Poma, *Il sex appeal dei corpi digitali. Seduzione, amori, tradimenti, malattie e immortalità dei nostri digital body*, Milano, Franco Angeli 2016, quando parliamo di "corpo digitale", staremmo facendo riferimento alla ricostruzione digitalizzata delle informazioni che produciamo nelle nostre interazioni virtuali, costantemente aggiornate, archiviate in una miriade di piattaforme e database diversi che disegnano i confini di chi noi siamo.

dall'entità che l'ha generato. Cosicché, il nostro corpo fisico può diventare un vero e proprio bagaglio informativo attraverso la biometria, generando un clone ottico, o clone digitale, che “prende vita propria, ringhia, perde le bave, si dissocia da noi, ci danneggia, e alla fine ci fa del male”¹²⁴.

Clonophobia

Nella convinzione che il terrore nei confronti del clone (e che l'uso della figura del clone come strumento di terrore) sia una delle metafore centrali attraverso cui studiare diversi fenomeni biopolitici, strumento retorico utile per indagare differenti aspetti della società, W.J.T Mitchell vi ha dedicato un intero saggio, intitolato *Cloning terror* (da intendersi come “terrore della clonazione” ma anche come “clonazione del terrore”). Nel testo l'idea di clone e di clonazione si emancipa dal ruolo di mero *topos* letterario e cinematografico per diventare una lente interpretativa di quell'epoca che Mitchell chiama “della biocibernetica”, un'epoca dominata dalla decodifica del DNA e dalla diffusione di massa del computer, in cui sembrano essersi risvegliati definitivamente gli antichi tabù sui gemelli, e tutto il relativo clan di doppi, astuti o inquietanti (*canny or uncanny*) e Doppelgänger¹²⁵.

¹²⁴ L. Poma, *Il sex appeal dei corpi digitali. Seduzione, amori, tradimenti, malattie e immortalità dei nostri digital body*, Milano, Franco Angeli 2016, pp. 82-83.

¹²⁵ W.J.T. Mitchell, *Cloning Terror: The War of Images, 9/11 to the Present*, Univ of Chicago Press, Chicago 2011.

Oggi la possibilità tecnica della clonazione umana è a portata di mano, e questa possibilità ha risvegliato molti dei tradizionali tabù sulla forma più potente e inquietante di creazione di immagini: la creazione della vita artificiale. L'idea di duplicare forme di vita, e di creare organismi viventi "a nostra immagine" ha preso alla lettera la possibilità prefigurata nel mito e nella leggenda, dalla fantascienza cyborg ai robot, dalla storia di Frankenstein al Golem, e alla stessa storia biblica della creazione, in cui Adamo viene formato con la creta "a immagine e somiglianza di Dio", ricevendo il soffio della vita¹²⁶.

Il clone si fa pertanto metafora, oltre che implicito nemico iconico, delle ansie legate alla riproduzione umana, all'aborto, all'omosessualità, alla copia, alla vita artificiale, e diventa strumento dialogico per mettere in relazione il terrorismo, l'ontologia dell'immagine binaria, la riproduzione seriale di gadget, l'identità biologica, le implicazioni della clonazione degli animali e degli esseri umani oltre che la produzione di immagini stessa.

Proprio su queste si concentra la nostra attenzione, perché se la tesi di Mitchell si spinge fino allo studio della rappresentazione mediatica del terrorismo islamico, analizzando l'idea di clonazione in rapporto alla caduta delle Twin Towers, alla proliferazione di cellule terroristiche jihadiste sparse per il mondo, ne concepisce al contempo un'idea trasversale che è stata applicata alle macchine, agli edifici, alle istituzioni e soprattutto alle immagini stesse, riattualizzando l'annosa questione della *mimesis*, nostro punto di partenza per un'analisi dello sviluppo dei video *deepfake* e delle strategie mimetiche dei loro produttori.

¹²⁶ W.J.T. Mitchell, *I quattro concetti fondamentali della scienza dell'immagine*, in: W.J.T. Mitchell; M. Cometa, V. Cammarata (a cura di), *Pictorial Turn. Saggi di cultura visuale*, Raffaello Cortina Editore Milano, 2017, p. 77.

Se l'immagine è un'icona, e quindi un segno il cui riferimento passa per una somiglianza o per una similitudine, dice Mitchell, allora un clone altro non è che una "superimmagine" (*superimage*), ovvero un duplicato perfetto, non solo delle fattezze, della superficie di un originale, ma della sua essenza profonda, di quel codice che gli conferisce la sua specifica identità.

La clonazione potrebbe essere chiamata "copia profonda" dal momento che va oltre il visuale o la superficie fenomenica per copiare la struttura interna e i meccanismi dell'entità, specialmente i meccanismi che controllano la sua stessa riproduzione. La clonazione è diventata un'immagine della produzione stessa di immagini, una *metapicture* della più avanzata forma della tecnologia per la produzione di immagini del nostro tempo¹²⁷.

Per Mitchell vi è quindi un'oscurità nella clonazione, nell'idea che questa agisca su una "struttura interna" e profonda, e quindi non visibile all'occhio, il che sembra richiamare quell'idea di "falso profondo" e di "falso prodotto attraverso il *deep learning*" (apprendimento profondo) a cui i video *deepfake* e le paure ad essi annesse sono collegate. D'altronde, ogni volta che emerge una nuova tecnologia, come nel caso dei *deepfake*, afferma Claud David-Strauss, "si scatena una reazione apocalittica per cui si proclama l'avvento di un orizzonte di postverità entro il quale non è più possibile credere ai nostri occhi"¹²⁸.

¹²⁷ W.J.T Mitchell, *op.cit.*, p. 29.

¹²⁸ D. Levi Strauss, *Perché crediamo alle immagini fotografiche?*, Johan & Levi Editore Milano, 2021.

La *clonophobia* sarebbe, quindi, come per il pittore Ferhoeffler, protagonista de *Il capolavoro sconosciuto* di Balzac, ammaliato e terrorizzato dall'idea di aver generato una copia vivente della propria sposa, o per Pigmalione, lo scultore misogino che nelle *Metamorfosi* di Ovidio eresse una statua d'avorio di una donna di tale bellezza da innamorarsene e di chiedere agli dei di darle vita, "una realizzazione, una literalizzazione della paura più antica – e (presumibilmente) della più antica speranza – riguardo alle immagini, ovvero che le si possa portare in vita"¹²⁹.

Mimesis: un framework teorico

GAN e mimesis

Secondo Stephen Halliwell lo statuto attuale della *mimesis* sarebbe controverso e instabile nonostante la sua importanza e la sua influenza all'interno della storia dell'estetica. Questa apparirebbe infatti agli occhi di molti filosofi e di molti critici come "una colonna infranta e sopravvissuta di un edificio classico ormai diruto, una reliquia tristemente obsoleta di antiche certezze"¹³⁰, per cui, seguendo questa interpretazione, i presupposti e i valori della *mimesis* (già riscoperti in epoca neoclassica e rinascimentale) sarebbero stati definitivamente lasciati alle nostre spalle.

¹²⁹ W.J.T Mitchell, *op.cit*, p. 31.

¹³⁰ S. Halliwell, *L'estetica della mimesis. Testi antichi e problemi moderni*, Aesthetica Edizioni, Palermo 2009, p. 293.

L'imitazione sarebbe stata ripudiata dal Romanticismo, e divenuta oggetto estraneo alla considerazione moderna e post-moderna tanto che si è sovente sentito dire come, tanto sul piano della pratica che sul piano della teoria dell'arte, viviamo in un'"epoca postmimetica"; ma, continua il filosofo, le teorie della *mimesis* avrebbero in realtà "continuato a giocare un ruolo importante nella filosofia e nella critica dell'arte dal Rinascimento a oggi"¹³¹.

I video *deepfake*, con la promessa di una produzione sempre più precisa di cloni audiovisivi di ognuno di noi (anche nella dimensione del *live streaming*¹³²), le reti neurali (con la loro imitazione del sistema gerarchico di percezione visiva del cervello umano) e più in generale le GAN (attraverso la capacità di produrre dati di output pressoché indistinguibili da quelli di input), sembrano riattualizzare le teorie sulla *mimesis* ai tempi dell'intelligenza artificiale.

Lo scopo di una GAN è quello di ridurre al minimo la differenza tra l'output finale presentato come "immagine reale" e le "immagini reali" fornitegli come input, prevedendo quindi l'apprendimento di un sistema di replicazione. Sebbene molti artisti come Mario Klingemann abbiano tratto ispirazione da e lavorato sulle storture causate dalle GAN, considerando quindi le Reti Antagoniste Generative come capaci di produrre opere d'arte con valore espressivo, e conducendo il proprio lavoro all'interno di un paradigma surrealista, queste sono prevalentemente utilizzate per produrre dati dall'aspetto realistico, in cui

¹³¹ *Ivi*, p. 294

¹³² Si veda, a tal proposito, la strabiliante presentazione di Metaphysic ad America's Got Talent: <https://www.youtube.com/watch?v=mPU0WNUzsBo>.

prende corpo quell'idea di "similitudine sensibile" che Walter Benjamin nel suo *Sulla facoltà mimetica* liquidava come banale¹³³.

Da un lato, quindi, si ha l'inquadramento delle GAN all'interno di una ragion d'essere squisitamente mimetica dove, per *mimesis*, si può intendere "calco" o "copia pedissequa", la riproduzione letterale e in un certo senso ottusa di un dato di partenza. Dall'altro, vi è l'idea delle GAN come generatrici di opere espressive, strumenti creativi "dalla natura ineluttabilmente magica" che possono portare a momenti di "similitudine non sensibile"¹³⁴.

Queste due dimensioni dell'arte mimetica risvegliano un antico dibattito che, ben lungi dall'esaurirsi, ritrova nuova linfa nel panorama dei video *deepfake* e che può essere letto sotto la lente di paradigmi secolari. D'altronde, è sufficiente uno sguardo di sorvolo sul pensiero antico per rendersi subito conto che nella nostra cultura la riflessione attorno all'immagine ci interroga dalla nascita della filosofia stessa e che "gli orizzonti problematici dischiusi nelle meditazioni di Platone e Aristotele, lungi dal poter essere archiviati come capitoli della storiografia, continuano a interpellare con urgenza i teorici contemporanei"¹³⁵.

A questo proposito, Mitchell ci ricorda come il clone aggiorni le antiche fobie riguardo alla produzione di immagini, alla *mimesis*, alla duplicazione, al rispecchiamento e alla copia (per cui la clonophobia altro non sarebbe se non un'attualizzazione della più generica paura verso le immagini e verso la loro produzione) fobie che avrebbero radici lunghissime nella cultura

¹³³ W. Benjamin, *Sulla facoltà mimetica*, in *Angelus Novus*, Einaudi, Torino 1962.

¹³⁴ G.W. Dyer, *ibid.*

¹³⁵ A. Pinotti, A. Somaini, *Cultura visuale*, Einaudi, Torino 2016.

occidentale e che sembrano aver investito anche molte analisi sul presunto allontanamento dell'immagine digitale dal suo referente e che possono essere fatte risalire alla Grecia Antica.

Mimesis come indebolimento ontologico di un originale

Pur essendo scarsa la nostra documentazione attorno alle teorie dell'immagine nel 5 secolo a.C. si può ragionevolmente ritenere che nell'Atene classica lo statuto della *mimesis* visiva fosse sicuramente già oggetto di critica e di dibattito ben prima del decisivo contributo apportato da Platone in alcuni passi del *Cratilo*, forse il primo dialogo interessato al tema della *mimesis* artistica, il che lascerebbe intendere come il dibattito attorno alle arti visive fosse ben più ampio di quanto noi oggi si possa ricostruire¹³⁶.

Se alcuni testi platonici hanno consentito un'opposta lettura dell'immagine che darebbe credito al produttore di immagini della capacità di rendere visibili delle realtà invisibili e quindi anche delle posizioni iconofile da parte dei platonici, secondo numerosi autori tra cui Maria Bettetini a Platone si potrebbe riconoscere però una "posizione iconoclasta" responsabile di molta diffidenza e di rifiuti violenti delle immagini che giungono fino ai nostri giorni: "ombre di ombre" e "copie di copie", sarebbe questa la connotazione negativa "elevata alla seconda

¹³⁶ S. Halliwell, *op.cit.*, p.113.

potenza” con cui si riassumerebbero le posizioni platoniche afferenti all’arte come *mimesis*¹³⁷. Platone vedrebbe infatti nella capacità dell’immagine di sostituirsi alla vita una grande seduzione, un grande inganno, il che lo portò a relegare gli artisti figurativi agli ordini più bassi all’interno della *polis*¹³⁸.

Considerando la *mimesis* un’attività “demiurgica”, rappresentazione più o meno fedele del mondo delle idee, da *Il Cratilo* in poi Platone tornerà più volte su concetti e termini relativi alle immagini mimetiche nello *Ione*, nel *Fedro*, nelle *Leggi*, nel *Sofista* e nel *Timeo*, ponendosi interrogativi sulla loro natura e i rapporti che queste stabiliscono con il loro originale, e soprattutto ne *La Repubblica*, a noi utile per evidenziare l’idea di *mimesis* come indebolimento ontologico di un originale.

Se il libro II e III del *La Repubblica* si concentrano sulla forma poetica della *mimesis* per affrontare “la narrazione attraverso la *mimesis*” da un punto di vista morale al fine di affermare che i giovani Guardiani della città ideale dovrebbero essere esposti a forme mimetiche solo quando rappresentano personaggi virtuosi (l’identificazione da parte del lettore/spettatore nelle passioni dei personaggi potrebbe provocare uno smarrimento, una frantumazione emotiva del soggetto devastante per la sua integrità morale¹³⁹) è in particolare il libro X della *Repubblica* a muovere la nostra attenzione, in quanto più volte evocato come testo rappresentativo, compendio del pensiero platonico sulla *mimesis* e, più genericamente, della

¹³⁷ M. Bettetini, *Contro le immagini. Le radici dell’iconoclastia*, Edizioni Laterza, Roma-Bari 2006, pp.3-4.

¹³⁸ G. Boehm, *Il ritorno delle immagini*, in: A. Pinotti, A. Somaini, (a cura di), *Teorie dell’immagine. Il dibattito contemporaneo*, Raffaello Cortina Editore, Milano 2008 p. 62.

¹³⁹ M. Vegetti, *Introduzione*, in Platone, *La Repubblica*, Mondadori, Milano 2016, p. 61.

condanna all'arte che ne deriva, il "non accettare quanto vi è in essa di imitativo", o di accettarlo "meno di ogni altra cosa"¹⁴⁰.

Riprendendo dalla critica alla poesia condotta già nel II e nel III libro, ribadendone la necessità di non ammetterla nella città ideale (con particolare riferimento alla poesia *mimetike*) ma assumendo come modello il caso della pittura, Platone articola la riflessione sulla *mimesis* lungo tre livelli distinti: il rapporto tra idea e sua rappresentazione, tra imitatore e conoscenza dell'oggetto imitato, tra imitazione poetica (comica e tragica) e spettatore dell'imitazione¹⁴¹.

Nel libro X Platone accusa l'arte di essere imitazione di imitazione e dunque incapace di restituire l'idea di ciò che è stato imitato¹⁴², evocando il dispositivo dello specchio come strumento mimetico di riproduzione illusoria del reale al centro della sua requisitoria contro i facitori di immagini ("... se vuoi prendere uno specchio e girarlo in ogni direzione: subito farai il sole e quanto vi è nel cielo, subito la terra, subito te stesso e gli altri viventi e mobili e piante e tutto ciò di cui ora si diceva"¹⁴³). Tale condanna della *mimesis* attraverso l'immagine dello specchio che sarebbe in grado di rappresentare solo fenomeni parziali dei singoli oggetti, risultando più che altro in una contraffazione, verrà però "invertita di segno" e per secoli l'imitazione verrà considerata il compito supremo dell'artista¹⁴⁴.

¹⁴⁰ Platone, *La Repubblica*, Mondadori, Milano 2016, p. 1093.

¹⁴¹ M. Vegetti, *op.cit.*, pp. 212-213.

¹⁴² P. Klossowski, *Simulacra. Il processo imitativo nell'arte*, Mimesis, Milano 2002, p. 19.

¹⁴³ Platone, *op.cit.*, p. 1099.

¹⁴⁴ A. Pinotti, *op. cit.*, pp. 27-29.

Halliwell contesta però l'idea che Platone tratti le opere mimetiche come pure "imitazioni di imitazioni", affermando che queste formule non ricorrano mai nella *Repubblica* e che avrebbero l'unico risultato di allontanare la centralità conferita da Platone all'intenzionalità dell'uomo nella produzione di opere mimetiche¹⁴⁵. Halliwell smentisce inoltre che la "teoria dello specchio" possa essere applicata uniformemente all'arte mimetica e si oppone all'idea diffusa che vedrebbe Platone limitare costantemente la *mimesis* figurativa al "rispecchiamento" della realtà visibile. Certo, non nega la netta presa di posizione nei confronti dell'arte imitativa da parte del filosofo greco, e le preoccupazioni che emergono da molti passi dei suoi scritti, ma ne sottolinea altresì il carattere provocatorio, retorico e satirico (funzionale a Platone per salvaguardare la "vera" imitazione, quella dell'uomo che imita e diventa simile agli dei¹⁴⁶). Integrandola nel contesto più ampio del discorso che confina i prodotti della pittura e della poesia a un livello "due volte discosto dalla verità", rendendoli persino inferiori agli artefatti prodotti dagli artigiani e dai carpentieri, per Halliwell Platone invita a vedere "l'insufficienza di qualsiasi concezione della pittura che sottolinei le apparenze pure e semplici... a spese del significato rappresentativo ed espressivo"¹⁴⁷, per cui Platone starebbe proponendo una concezione dell'arte mimetica fondata su una critica delle idee sovente considerate centrali per definire la tradizione mimetica in ambito estetico (aderenza alla verità delle apparenze, verosimiglianza, realismo, illusionismo)¹⁴⁸ in realtà riconoscendo alla

¹⁴⁵ S. Halliwell, *op.cit.*, p. 126

¹⁴⁶ M. Bettetini, *op.cit.*, p. 11.

¹⁴⁷ S. Halliwell, *op.cit.*, p. 130.

¹⁴⁸ *Ivi.* p. 139.

mimesis figurativa la capacità e il dovere di ottenere molto più che un semplice rispecchiamento delle apparenze¹⁴⁹.

Già nel *Cratilo* Platone lascia introdurre a Socrate il concetto di *mimesis* applicandolo ai segni visivi e al linguaggio parlato, in alcuni passi che lascerebbero intendere una concezione della *mimesis* artistica come funzionale per la raffigurazione di alcune cose riscontrabili nel mondo sensibile, ma delle quali sarebbe incapace di cogliere la profonda verità, la loro essenza.

È da questo passo del libro X della *Repubblica* che emerge la discussa concezione platonica della *mimesis* come “due volte lontana dalla verità”.

“«Poniamo dunque anche ora una pluralità di questi oggetti, quella che vuoi. Per esempio, se credi, ci sono, diciamo, molti letti e tavoli»

«Come no?»

«Ma due sono, direi, le idee relative a questi mobili: una di letto, una del tavolo.»

«Sì»

«E non siamo allora soliti dire che gli artefici di entrambi questi mobili, rivolgendo lo sguardo all’idea, hanno costruito in questo modo l’uno i letti, l’altro i tavoli di cui noi ci serviamo? e non accade lo stesso per ogni altra cosa? Perché certo l’idea stessa non la fabbrica nessuno degli artigiani» [...]

«Allora questi letti risultano di tre tipi: uno è quello che esiste in natura, di cui potremmo affermare, io credo, che è stato prodotto da un dio. O da chi altri?»

«Nessun altro, penso»

«E poi uno dal costruttore»

«Sì» disse

«E uno dal pittore. Non è così?» [...]

¹⁴⁹ *Ivi* p. 129.

«E sia» dissi io: «chiami imitatore l'autore dei prodotti di terza generazione, a partire dalla natura?»»

«Questo sarà dunque anche il posto del poeta tragico, se appunto è un imitatore – per natura terzo a partire dal re e dalla verità – e di tutti gli altri imitatori»¹⁵⁰.

Qui emerge chiara l'idea platonica di un indebolimento ontologico dell'originale causato dalla sua copia puramente mimetica. Si hanno, pertanto, nelle parole che Socrate pronuncia in dialogo con Glaucone e Adimanto, tre letti: il letto come idea, il letto costruito e il letto dipinto. L'idea è stata prodotta dal "dio", e vi è una sola idea di "letto". Il falegname, l'artigiano, è "demiurgo" del letto di legno, poiché guardando a un'idea fabbrica una cosa senza altri mediatori, una cosa non vera ma comunque non così lontana dal vero come lo sarà la sua immagine dipinta, non certo ciò che è ma comunque "qualcosa di simile all'essere" per cui il pittore non ricopre il ruolo né di creatore né di demiurgo del letto di legno; il pittore non è né creatore, né demiurgo ma viene relegato al mero ruolo di "imitatore", esattamente come il poeta. Il pittore che ritrae un letto in immagine crea un'"immagine copiativa depotenziata" dal punto di visto onto-gnoseologico rispetto al letto esperibile e percepibile sul piano dei sensi creato dal falegname, già di per sé una forma impoverita e depotenziata in confronto all'idea del letto in sé¹⁵¹.

Per questo, i pittori traggono in inganno poiché riproducono gli oggetti già precedentemente riprodotti dagli artigiani, senza possederne l'arte e a un grado di distanza maggiore rispetto a quello che questi trattengono con

¹⁵⁰ Platone, *op.cit.*, pp. 1095-1103.

¹⁵¹ A. Pinotti, A. Somaini, *op.cit.*, p.47.

l'idea ("l'imitatore non avrà né vera scienza né retta opinione della cosa che imita"¹⁵²). Quello dell'imitatore è più che altro un gioco, un gioco che riesce perché si indirizza alla parte più debole dell'animo umano; non alla sua componente raziocinante, in grado di calcolare e soppesare, ma alla sua sensibilità, la quale si lascia ingannare tanto da un bastone immerso nell'acqua, credendolo rotto, quanto dagli *eidōla* prodotti dai poeti e dai pittori. Quest'ultimi sarebbero infatti capaci di usare la prospettiva come gli illusionisti e i prestigiatori e i loro trucchi¹⁵³.

... quando uno ci annunci, parlando di qualche persona, di essersi imbattuto in un uomo che conosce tutte le tecniche e tutte le altre cose che appartengono alle competenze specifiche dei singoli individui, e che non c'è nulla che non conosca in modo più esatto di chiunque altro, nel suo caso bisogna supporre che si tratti di un uomo ingenuo, che, a quanto pare, ha incontrato un mago e imitatore e se ne è fatto ingannare, tanto che quello gli è sembrato onnisciente, ma solo perché lui stesso non è capace di discriminare fra conoscenza, ignoranza e imitazione¹⁵⁴.

Sia il poeta che il pittore sono in grado di fabbricare qualsiasi cosa, e questo perché di ogni cosa colgono solo delle impressioni, una piccola parte, un *ēidolon*, e non possiedono alcuna vera conoscenza della cosa che imitano. Dall'arte come pura imitazione emergerebbe quindi un indebolimento ontologico dell'oggetto rappresentato, dell'originale, la cui copia sarebbe due volte lontana dalla sua origine e quindi dalla *verità* ("La tecnica dell'imitazione è dunque in un certo modo lontana dal vero, e, a

¹⁵² Platone, *op.cit.*, p. 1119.

¹⁵³ M. Bettetini, *op.cit.*, pp. 10-11.

¹⁵⁴ Platone, *op.cit.*, pp. 1105-1106.

quanto sembra, riesce a produrre tutte le cose per questo: coglie una piccola parte di ognuna, e si tratta di un simulacro”¹⁵⁵) e soprattutto scaturirebbe una bugia. Per Platone l’arte è di fatto *mimesis* che nell’imitare trae in inganno e l’artista che finge di conoscere si limita a riprodurre oggetti e situazioni, per cui l’artista sarebbe come il “sofista”, il retore che finge di conoscere le cose di cui parla ma l’unica cosa che conosce è l’arte del parlare¹⁵⁶.

Platone riconoscerà, sì, la piacevolezza dell’arte mimetica, ma proprio per questo condannerà l’eccessiva somiglianza tra rappresentazione e soggetto rappresentato: con la bellezza della copia, con la sua abilità nel ritrarre un falegname, infatti, un pittore potrebbe ingannare i bambini e gli uomini di poco senno, inducendoli a credere che si tratti di un vero falegname. Di questo invece il pittore conosce solo le apparenze, e su tali si baserà per dipingerlo. Il prodotto dell’artista, l’imitazione risulta quindi essere come una riproduzione fallace dell’apparenza dell’essere e non può essere altro che bugia, menzogna capace di condurre chi vi assiste all’interno di un’illusione: un simulacro, un *eidōlon*.

L’etimologia di *eikōn* (immagine) e di *eikōs* (verosimile) è la stessa, derivando entrambe dal verbo *eoika*, che significa “sono simile”, ma anche “sembra” e “mi sembra”. Per Platone esistono due tipologie di immagine: *eikōn* è l’immagine fedele, simile all’idea, il prodotto del Demiurgo (intelligenza divina) e del demiurgo (artigiano), un’immagine che rivela qualcosa del suo modello per similitudine. Pur non essendo vera, questa assomiglia al vero. La seconda tipologia di immagine, *eidolōn*, invece

¹⁵⁵ *Ibid.*

¹⁵⁶ M. Bettetini, *op.cit.*, p.7.

allontana sempre dal vero: è “un oggetto fatto a somiglianza di quello vero, diverso, ma simile”. Esistono dei modelli eterni e immutabili e copie di tali modelli prodotte dal dio, e ciascuna è *eikōn* di un modello, come ad esempio le ombre, i sogni e le immagini riflesse, che sono anche *eidōla* prodotti dal dio stesso. Esistono poi *eidōla* prodotti dagli artisti che imitano un *eikōn* e ingannano, presentandolo come se fosse esso stesso un *eikōn*¹⁵⁷.

Ogni immagine, un *eidōlon*, mantiene una somiglianza con la cosa rappresentata, e proprio questo rapporto di similitudine genera l'illusione. Gli *eidōla* ingannano, sostituiscono la verità con una sua parvenza e possono mistificare il reale e condurre all'idolatria. Nella narrazione biblica del “Peccato del Vitello”, contenuta nel libro dell'Esodo, viene descritta la produzione di un vitello d'oro da parte di Aronne per gli israeliti, desiderosi di possedere un'immagine da adorare e che li potesse guidare non sapendo delle sorti di Mosè, salito sul monte Sinai per ricevere le tavole dei X comandamenti. La produzione del Vitello suscita le ire di Dio, pronto a punirli per l'affronto, poi dissuaso dalla supplica di Mosè che, tornato tra di loro, distruggerà l'idolo, mettendolo a ferro e fuoco. Come per Platone, la duplicazione in immagine (il vitello) causa una sostituzione della cosa vera nel mondo delle idee (Dio) con una sua imitazione fittizia, la verità con la menzogna. La duplicazione, l'intento mimetico diviene “mistificazione”.

¹⁵⁷ M. Bettetini, *op.cit.*, pp. 14-15.

Mimesis come raddoppiamento chiarificante

A differenza di Platone, Aristotele non identifica la realtà con il mondo delle idee, con un'esperienza diversa da quella terrena. Le opere rappresentative come le poesie e i dipinti comporterebbero piuttosto delle "rassomiglianze" e non ci proporrebbero delle pseudo-realtà ingannevoli, ma ci offrirebbero il significato fittizio di una realtà possibile per mezzo degli strumenti di una mediazione artistica che consente che questa realtà venga riconosciuta. Le opere mimetiche non dovrebbero rappresentare oggetti particolari o singoli (solo i ritratti rientrerebbero nella tipologia delle rassomiglianze individuali) cosicché la comprensione estetica non può limitarsi a identificare una "copia" con un originale già noto, poiché nel valutare le opere mimetiche ci è concesso di conoscere e comprendere i modi in cui alcune forme dell'esperienza umana sono volutamente espresse all'interno delle opere¹⁵⁸.

La *mimesis* non sarebbe una semplice duplicazione ma assurgerebbe per Aristotele a vero e proprio principio del processo di conoscenza sin dall'infanzia, dal momento che l'imitare sarebbe connaturato agli uomini fin dalla tenera età "ed è per questo che essi si differenziano dagli altri animali, perché [l'uomo] è il più incline all'imitazione e le sue prime acquisizioni cognitive le compie mediante l'imitazione"¹⁵⁹: dall'imitazione, che risponderebbe quindi a un istinto naturale, l'uomo trarrebbe le prime conoscenze e ne ricaverebbe piacere. Il piacere della *mimesis* deriverebbe per Aristotele dal riconoscimento del modello rappresentato,

¹⁵⁸ S. Halliwell, *op.cit.*, pp. 169-170.

¹⁵⁹ Aristotele, *Poetica*, Giulio Einaudi Editore, Torino 2008, p. 19.

dell'originale, e dalla conoscenza della tecnica adoperata per la rappresentazione.

... le immagini di quelle cose che in sé vediamo con fastidio, quando siano eseguite con la massima accuratezza le contempliamo con piacere, per esempio le figure degli animali più spregevoli e dei cadaveri. C'è una causa anche di questo, che imparare è piacevolissimo non solo per i filosofi, ma anche ugualmente per gli altri, se nonché questi ne partecipano in piccola misura. Per questa ragione, infatti, si prova piacere nel vedere le immagini, perché accade che nel vederle si impari e si concluda con il ragionamento che cosa è ciascun oggetto, per esempio che "costui è quell'uomo". Giacché, se accade di non averlo visto in precedenza, esso non procurerà piacere in quanto imitazione, ma per l'esecuzione, o per il colorito, o per qualche altra ragione del genere ... ¹⁶⁰.

Per l'allievo di Platone, quindi, la *mimesis* produce conoscenza attraverso il piacere. E da dove scaturisce questo piacere? Sebbene nell'introduzione alla *Poetica* Pierluigi Donini affermi che sarebbe riduttivo e semplicistico considerare questo passo di Aristotele solamente attraverso la lente della ritrattistica e del riconoscimento di un volto¹⁶¹, l'esempio apportato da Aristotele è a noi utile per capire da dove derivi la piacevolezza del riconoscimento di un volto nei video *deepfake*.

La condizione posta da Aristotele per l'accesso al piacere di una rappresentazione mimetica è infatti l'aver "visto in precedenza" l'oggetto imitato, al contempo meravigliandosi della tecnica che ha reso possibile questa operazione. Quando Aristotele fa riferimento alla "massima accuratezza" non sta però parlando di un realismo totale della

¹⁶⁰ Aristotele, *op.cit.*, pp. 19-21.

¹⁶¹ P. Donini, *Introduzione in Poetica*, Giulio Einaudi Editore, Torino 2008, p. LXVI.

rappresentazione. Pur concentrandosi, come d'altronde Platone prima di lui, più che altro sulla poesia tragica, ritenendola la più interessante ed espressiva delle arti mimetiche, le arti visive hanno una funzione paradigmatica della concezione aristotelica della *mimesis* artistica nel suo complesso poiché una "somiglianza" resa mimeticamente trova il suo esito tipico nella raffigurazione visiva (per via del parziale confronto percettivo che si istituisce tra forme e colori delle rappresentazioni pittoriche e le sembianze degli oggetti) ma per il filosofo questa non deve avere un rapporto strettamente referenziale, non implicando alcun rapporto con particolari identificabili¹⁶².

Da qui deriverebbe l'attribuzione alla *mimesis* in senso aristotelico dell'idea che questa debba più che altro rispecchiare l'essenza dell'oggetto rappresentato, per cui "rappresentazione mimetica" non significherebbe "realismo" bensì restituzione della cosa per quello che è, come universale. Per Aristotele tanto nelle arti figurative che nella poesia tragica ed epica, la *mimesis* non dovrebbe mai essere identificata con la ripetizione pedissequa di ogni specifico dettaglio del reale: questa non sarà ovviamente un qualcosa di puramente immaginario, arbitrario e a discrezione dell'artista, dovendo necessariamente dialogare con una realtà sensibile, ma di questa dovrà riprodurre l'essenza profonda e la verità strutturale, la sua universalità, qualità che non coincidono affatto con il dato fruibile all'esperienza diretta e alla percezione immediata dell'oggetto. Nel caso della pittura, in particolare, egli pensa al referente reale della *mimesis* artistica, un'essenza recondita dell'oggetto, la quale altro non può essere

¹⁶² S. Halliwell, *op.cit.*, p.142, 185.

che quella filtrata e astratta dalla mente dell'artista e non certo quella data nella realtà fisica, nel mondo esterno¹⁶³. Pertanto, il pittore che volesse rappresentare un tavolo, non dovrebbe necessariamente imitare un tavolo in ogni suo dettaglio, ma dotarlo di un valore cognitivo tale da fornire allo spettatore il suo carattere universale, la sua "tavolitudine".

Le immagini costituiscono per Aristotele imitazioni della vita che costituiscono una "seconda natura" tanto che praticamente tutti miti fondativi di ogni religione invocano un momento di "creazione dell'immagine" e della messa in vita di quell'immagine¹⁶⁴.

Mimesis non è sinonimo di "rappresentazione realistica" ma definire "mimetica" una rappresentazione significa collocarla all'interno di un contesto di pratiche culturali che muovono da certi istinti umani che si sviluppano in istituzioni le quali implicano un rapporto dialettico tra i produttori della rappresentazione, gli esecutori e il loro pubblico. Il raddoppiamento mimetico diviene quindi una chiarificazione del suo originale, in grado di dire qualcosa sull'oggetto rappresentato grazie alla capacità del poeta mimetico di intrecciare sapientemente i fili dell'azione.

Per il nostro fine di indicare il raddoppiamento chiarificante come schema interpretativo degli atti mimetici nei *deepfake*, è utile annotare come nel IX capitolo della *Poetica* Aristotele affermi che la poesia, arte mimetica, si occupa delle "cose che potrebbero avvenire e cioè quelle possibili secondo la verisimiglianza o la necessità"¹⁶⁵ gettando le basi per un rapporto mutevole e complesso tra il mondo interno all'opera d'arte e

¹⁶³ P. Donini, *Introduzione in Poetica*, pp. XXX-XXXI.

¹⁶⁴ W.J.T. Mitchell, *op.cit.*, p. XIX.

¹⁶⁵ Aristotele, *op.cit.*, p. 61.

quello dell'artista e dell'auditorio, che si articola nella dialettica tra realtà e finzione, rispecchiamento fedele e mondi ipotetici; l'arte mimetica dovrebbe infatti scegliere l'oggetto della *mimesis* secondo tre opzioni differenti: essendo il poeta un imitatore quanto un pittore, è inevitabile che egli imiti o le cose come erano o sono, o quali si crede che siano o quali dovrebbero essere.

Inoltre per Aristotele è fondamentale un'operazione certosina di selezione e ricomposizione interpretativa di tutti quegli aspetti che, in un certo senso risultano come "privilegiati" del reale, col fine di restituirne il senso profondo che sfugge invece alla percezione diretta della totalità di ciò che accade. In altri termini, Aristotele ripone fiducia nella possibilità della *mimesis* di estrapolare dal caos del fluire del tempo e dell'avvicinarsi delle azioni umane degli atti stabili che possano avere una durata e elevarsi a modelli etici e conoscitivi da offrire alle memoria collettiva, nascendo da un processo di astrazione del tempo reale e recando in sé un valore assoluto, un'universalità ottenibile attraverso la concatenazione di singole azioni che rappresentano una successione razionale e regolata da rapporti causali che acquistano così valore esemplare¹⁶⁶.

È la derivazione sensibile della rappresentazione ad acquisire un ruolo prominente, la quale media tra interno ed esterno, soggetto e mondo, percezione ed intelletto, percezione e linguaggio¹⁶⁷. Pertanto, la sua concezione della *mimesis*, la cui struttura concettuale è stratificata e

¹⁶⁶ R. Giloldi, *La mimesis e le sue peripezie*, in *La Poetica e le sue interpretazioni. Aristotele tra filosofia, letteratura e arti*, Pendragon, Bologna 2020, pp. 110-112.

¹⁶⁷ A. Zhok, *Rappresentazione e realtà. Psicologia fenomenologica dell'immaginario e degli atti rappresentativi*, Mimesis, Milano 2014, p. 9.

complessa e di cui non si può certo rendere conto in queste pagine, investe il valore cognitivo delle diverse forme di musica, di arte visiva e di *mimesis* vocale intesa come “recitazione dell’attore”¹⁶⁸: il tratto fondamentale di questa concezione che ci preme sottolineare è che per Aristotele la *mimesis* produce oggetti il cui valore rappresentativo produce conoscenza e comprensione e che rappresentano una seconda natura in grado di informare il fruitore sul suo originale.

Mimesis come “incremento d’essere”

L’idea aristotelica della *mimesis* come seconda natura arricchita sul piano gnoseologico trova una diretta prosecuzione, quasi un aggiornamento o un ampliamento, nell’idea della *mimesis* come “incremento d’essere” proposta da Hans Gadamer in *Verità e metodo*, dove il filosofo, massimo esperto dell’ermeneutica filosofica moderna, ribalta la prospettiva platonica della *mimesis* come indebolimento ontologico che si è precedentemente attribuita a Platone (un rovesciamento che assume ancor più valore se pensiamo ai numerosi testi dedicati da Gadamer alla filosofia platonica¹⁶⁹, di cui può essere definito un vero e proprio specialista).

In *Verità e metodo* il filosofo tedesco scrive:

¹⁶⁸ Una questione cruciale anche nella studio dei video *deepfake* come si è visto in K. Hao, *Inside the Strange New World of Being a Deepfake Actor*, 2020, <https://www.technologyreview.com/2020/10/09/1009850/ai-deepfake-acting/>.

¹⁶⁹ H. Gadamer, *Studi Platonici*, Marietti Editore, Torino 1998.

Il mondo che appare nel gioco della rappresentazione non sta accanto al mondo reale come una copia, ma è questo stesso mondo reale in una più intensa verità del suo essere. E a sua volta la ripetizione, per esempio l'esecuzione scenica, non è una copia in confronto alla quale l'originale del dramma conservi una sua individualità distinta. Il concetto di *mimesis* [...] non indicava tanto la copia quanto la manifestazione del rappresentato. Senza la *mimesis* dell'opera non c'è il mondo, almeno come esso è nell'opera, e senza la ripetizione non c'è, d'altra parte, l'opera.

Alla luce di questa considerazione, potremmo dire che nella sua critica a Hegel Gadamer non identifichi la rappresentazione come un mero veicolo di senso, di significato, ma piuttosto con un vero e proprio accrescimento ontologico del mondo che rappresenta, dal quale è inscindibile e il cui senso profondo viene messo in luce proprio dall'opera d'arte. In altre parole, la *mimesis* realizzata da un'opera d'arte e che rende quest'ultima propriamente un'opera d'arte non è la dimostrazione di qualcosa d'altro, non la *mimesis* di qualcosa che sta al di fuori dell'opera: la *mimesis* prende forma proprio all'interno dell'opera d'arte ed è inscindibile dal suo originale. Pertanto, la vera *mimesis* non può essere identificata con o ridotta al mero concetto di "ripetizione"¹⁷⁰.

In merito alle arti figurative – per le quali Gadamer ritiene necessaria l'applicazione dell'intreccio ontologico di riproduzione e originale che precedentemente ha applicato alle arti transeunti (come la rappresentazione poetica) – si profila una distinzione fondamentale tra "copia" e "quadro-immagine". La prima non possiede alcun altro compito se non uguagliare il suo modello di riferimento dal momento che la sua

¹⁷⁰ J. Sallis, *On the Manifold Senses of mimesis. From Plato to Gadamer and Beyond*, in "The Blackwell Companion to Hermeneutics", 2016, p.295.

adeguatezza è definita dal riconoscimento del suo originale. Questa esiste quindi unicamente per sopprimersi come ente indipendente, con l'unica funzione di riprodurre, di mediare fedelmente l'originale imitato. La copia perfetta, o "copia ideale" non sarebbe però l'immagine speculare, l'immagine allo specchio, per riprendere la precedentemente discussa terminologia platonica, perché questa esiste solo in presenza del soggetto rispecchiato. La copia "vuol sempre essere vista in merito a ciò che raffigura. È una riproduzione che non vuol essere altro che la ripetizione di qualcosa e che ha la sua funzione unicamente nell'identificazione dell'originale"¹⁷¹¹⁷². Dunque, la sua funzione è la sola rassomiglianza con l'originale, realizzandosi unicamente nell'autosoppressione.

Per contro, il "quadro-immagine" "impone il proprio essere per far essere il raffigurato".

Ciò che si ha di mira qui è l'immagine come tale, giacché quel che importa è proprio come si rappresenta ciò che in esso è rappresentato. Ciò significa innanzitutto che l'immagine non rimanda semplicemente al rappresentato. Anzi, la rappresentazione rimane essenzialmente legata con il rappresentato, in certo modo gli appartiene. Questa è anche la ragione per cui lo specchio rimanda un'immagine e non una riproduzione: l'immagine speculare è appunto immagine di ciò che si specchia, ed è inscindibile dalla sua presenza. [...] È l'immagine del rappresentato – la sua immagine, non quella dello specchio – che si mostra nello specchio¹⁷³.

¹⁷¹ H. Gadamer, *Verità e metodo*, Rizzoli/Bompiani, Milano 2016, p.299.

¹⁷² A titolo esemplificativo, Gadamer qui riporta gli esempi della foto-tessera e del catalogo commerciale.

¹⁷³ H. Gadamer, *op.cit.*, p.301.

Questa tipologia di immagine¹⁷⁴, diretta conseguenza di quell'indistinzione magica tra soggetto reale e soggetto raffigurato nell'immagine che, per Gadamer, è lungi dall'esaurirsi, rivela l'inseparabilità ontologica dell'immagine dal "rappresentato" e informa il suo spettatore sull'originale, suggerendogli qualcosa della sua essenza mantenendo al contempo una propria realtà autonoma che ribalta il rapporto ontologico con l'originale: l'immagine-quadro valida il suo modello, è solo attraverso l'immagine che l'originale diventa tale. L'immagine-quadro non si sopprime come immagine per lasciar essere solo il soggetto rappresentato, è piuttosto un evento, un fatto ontologico in cui l'essere stesso si presenta fenomenologicamente per mezzo di una manifestazione visibile e fornita di senso. Quest'immagine non è una pura copia: essa rappresenta qualcosa che senza di essa non si presenterebbe in quel modo, dicendo al suo spettatore qualcosa di più sull'originale.

L'immagine agisce quindi sull'originale e questo diventa originale proprio in virtù dell'azione della sua immagine che a sua volta ne è manifestazione in quella che Gadamer chiama "comunione ontologica" tra "immagine" e "raffigurato". Per il filosofo tedesco, però, "che la rappresentazione sia un'immagine, e non l'originale stesso, non significa nulla di negativo, non è una diminuzione di essere, ma indica piuttosto una realtà autonoma". A differenza che nella copia, il rapporto originale-rappresentazione non è un rapporto a senso unico, ma che l'immagine abbia una propria realtà significa per l'originale che in essa questi possa presentare se stesso, o una parte di se stesso. "Nella rappresentazione,

¹⁷⁴ Gadamer porta qui come esempio una buona foto-ritratto.

questo subisce una *crescita nell'essere*", un aumento d'essere. Il contenuto proprio dell'immagine è definito ontologicamente come emanazione dell'originale... Giacché se l'Uno originario non viene impoverito nel fluire dei molti da esso, ciò significa che l'essere aumenta"¹⁷⁵.

La rappresentazione è un evento di cui l'artista, l'esecutore, l'interprete-lettore non sono autori, ma partecipi e in questa la cosa rappresentata acquisisce maggiore verità (più essere) di quanto non ne avesse precedentemente nella quotidianità.

Se intendiamo l'immagine come *repraesentatio*, fornita quindi di una sua propria valenza ontologica, dovremo modificare essenzialmente, anzi rovesciare quasi, il rapporto ontologico tra originale e copia. L'immagine ha in tal caso una sussistenza autonoma che agisce anche sull'originale. Propriamente, è infatti solo attraverso l'immagine che l'originale diventa immagine originale, è solo in virtù dell'immagine che il rappresentato diventa davvero qualcosa che si dà in una immagine¹⁷⁶.

Gadamer destituisce la relazione immagine/originale che era una pietra miliare dei concetti tradizionali della *mimesis*. Il filosofo non lascia collassare interamente la distinzione tra immagine e suo originale ma si prodiga nell'enfatizzare come, in verità, nell'opera mimetica debba rimanere solamente una traccia di questa differenza. Nell'opera d'arte c'è la manifestazione di un significato ma anche una sua protezione, un suo rifugio che stabilisce e mette al riparo quel significato che la stessa *mimesis* ha liberato: la trasformazione e l'apertura di nuove possibilità che

¹⁷⁵ *ivi.*, p. 303.

¹⁷⁶ *Ivi.*, p. 307.

prendono corpo nell'opera d'arte restituiscono un referente che ha subito una trasformazione¹⁷⁷.

Se la copia si muove sul piano della propria autodissoluzione, nella sua teoria dell'immagine Gadamer parla della "valenza ontologica" dell'immagine che rammenta *in primis* che cosa è ciò che viene raffigurato¹⁷⁸. Gadamer dichiara che la *mimesis* non è una questione legata al far sì che qualcosa che è già presente si renda evidente nuovamente (e magari con maggior chiarezza, come nel caso della duplicazione chiarificante), ma anzi, che la *mimesis* non riguarda proprio la duplicazione di qualcosa di già esistente, come ad esempio il disegno di un paesaggio che altro non sarebbe che un paesaggio disegnato: la *mimesis* riguarda la trasformazione: questa trasformazione non avverrebbe né in qualcosa di separato dall'opera d'arte, né nello spettatore dell'opera. La trasformazione è nell'opera stessa. Ha a che fare con il portare avanti nuove possibilità accresciute e prima invisibili come una certa illuminazione o un certo modo di mostrare una cosa da una differente angolazione. La *mimesis* è la valorizzazione di queste possibilità: sarebbe a dire che essa ci propone nell'opera stessa possibilità nuove, non sperimentate, sensi prospettici mai ancora manifestati¹⁷⁹.

Così prende corpo l'incremento d'essere, un incremento che quindi non riguarda soltanto la soggettività dell'artista o dello spettatore: è l'essere stesso della cosa rappresentata che viene realmente modificato nella rappresentazione. L'opera è quindi più vera della realtà stessa in quanto

¹⁷⁷ J. Sallis, *On the Manifold Senses of mimesis. From Plato to Gadamer and Beyond*, p. 295.

¹⁷⁸ G. Boehm, *op.cit.*, pp. 60-61.

¹⁷⁹ J. Sallis, *op.cit.*

struttura compiuta e conclusa, intelligibile in quanto liberata dalla casualità e dall'indefinitezza che caratterizza l'esperienza quotidiana¹⁸⁰.

Conclusioni capitolo

Nel capitolo appena trattato si è presa in considerazione la figura dell'avatar, indicandone le radici e mostrando come i *video deepfake* riattualizzino antiche fobie a loro legate, fobie che W.J.T Mitchell indica come di diretta derivazione delle stesse paure legate alla *mimesis* e alla produzione di immagini. A partire dalle considerazioni da lui esposte in *Cloning Terror*, dove l'autore si spinge fino a coniare il termine *clonophobia*, si è quindi deciso di proporre uno schema interpretativo per l'applicazione della teoria della *mimesis* ai *video deepfake* attingendo a piene mani da tre grandi autori di riferimento: Platone, Aristotele e Hans Gadamer, provando a far dialogare le loro tre differenti teorie della *mimesis* con la produzione dei *video deepfake*, proponendo quindi l'idea che un volto *deepfake* possa allontanarci dal suo originale e essere utilizzato per affermare il falso (partendo dall'idea Platonica di una distanza "alla seconda" tra l'immagine artistica e l'idea), per chiarificare degli elementi storici e reconditi di un personaggio (di diretta derivazione aristotelica), per scavare nella natura del personaggio rappresentato, cogliendone

¹⁸⁰ G.Reale, *Introduzione*, in *Verità e metodo*, Rizzoli/Bompiani, Milano 2016, pp. XXXVIII-XXXIV.

aspetti ontologici non ancora manifesti e leggerlo sotto una nuova luce (un *deepfake* che sarebbe quindi da intendersi come l'“incremento d'essere” gadameriano, una crescita ontologica del personaggio nella sua rappresentazione).

Studi di caso

The Eternal feat. Ian Curtis: deepfake come indebolimento ontologico di un originale

I video *deepfake* hanno trovato terreno fertile nel campo delle resurrezioni digitali. Già un'opera *deepfake audio* come *JFK Unsilenced*, prodotto dal The Times, permetteva all'ascoltatore di sentire il discorso mai pronunciato da John Kennedy al Dallas Trade Mart il giorno della sua morte mentre l'installazione *Dalí Lives* al Dalí Museum di St Petersburg, Florida, prevedeva una serie di schermi nei quali degli avatar del pittore, prodotti attraverso il *deep learning*, potevano interagire col pubblico¹⁸¹. Non solo, il MIT ha prodotto un video *deepfake* di Richard Nixon (sia audio che video) per il proprio progetto *In Event of Moon Disaster*. Lees, Keppel-Palmer e Bashford-Rogers hanno invece esplorato la possibilità che Margaret Thatcher potesse essere risuscitata digitalmente usando il *machine learning* per interpretare se stessa in un nuovo dramma storico¹⁸². Le resurrezioni digitali in *deepfake* rappresentano inoltre un interessante campo di studio perché “nel caso della morte di una star le tecnologie che mediano la sua immagine e la mantengono di fronte ai nostri occhi sembrano in qualche modo più evidenti... e questo è ancor più vero se le

¹⁸¹ M. Mihailova, “To Dally with Dalí: Deepfake (Inter)faces in the Art Museum”, in *Convergence: The International Journal of Research into New Media Technologies*, Vol. 0(0), 2021, pp. 1-17.

¹⁸² D. Lees, M. Keppel-Palmer, T. Bashford-Rogers, *The digital resurrection of Margaret Thatcher: Creative, technological and legal dilemmas in the use of deepfakes in screen drama*, in “Convergence: The International Journal of research into New media Technologies”, n.0(0), 2021, pp. 954- 973.

tecnologie utilizzate sono meravigliose, nuove o inspiegabili”¹⁸³. Sarebbe a dire che l’esibizione di una resurrezione digitale in *deepfake* espone con ancor più limpidezza le caratteristiche della tecnologia che la consente e che il piacere che suscitano nello spettatore possa essere principalmente un piacere di tipo tecnico, da ritrovarsi quindi nel riconoscimento dei meccanismi e di un’estetica dettati dal ricorso ai *deepfake*.

A partire dall’esperienza personale affianco al *deepfake designer* Salvo Compromesso per la realizzazione del video *deepfake The Eternal feat Ian Curtis #iancurtisalive*, un tentativo di resurrezione digitale del defunto frontman dei Joy Division Ian Curtis applicato alla cover di *The Eternal* prodotta dai Limonov, band a cui partecipo come voce principale, si possono trarre alcune considerazioni per analizzare l’applicazione di quanto detto a proposito di DeepFaceLab e la pragmatica della produzione di un video *deepfake* attraverso uno studio di caso, leggendo la resurrezione di Ian Curtis come indebolimento ontologico di un originale.

La scelta del soggetto è stata dettata da una somiglianza nei tratti del volto e dalla riconoscibilità delle *performance* di Curtis, la cui *epilepsy dance* è diventata un *Leitmotiv* del post-punk e che consentiva un tratto di ancoraggio per l’imitazione. L’idea che sottende al video è stata quella di produrre una cover dei Joy Division, inscenando una performance live di Ian Curtis attraverso una resurrezione digitale legata unicamente all’aspetto visuale di Curtis, privato della sua voce e costretto a un *featuring* sul quale non avrebbe potuto avere voce in capitolo. Seguendo le indicazioni di Salvo Compromesso, gli unici pre-requisiti necessari per la

¹⁸³ L. Bode, *Grave Robbing or Career Comeback? On the digital resurrection of dead stars*, in K. Kallioniemi, K. Kärki, J. Mäkelä, H. Salmi (a cura di), *History of Stardom Reconsidered*, 2007.

buona riuscita del video durante la fase di produzione sarebbero dovuti essere un video di destinazione in alta qualità (per un minimo di 720p, idealmente 1080 pixel) in cui il volto del soggetto ripreso fosse chiaramente visibile e un *dataset* sorgente del volto di Ian Curtis sufficientemente ampio e provvisto con dati il più possibile in alta qualità e con più angolazioni del volto possibili poiché, se il *dataset* non fosse stato sufficientemente esaustivo, alcune angolazioni sarebbero state difficilmente apprendibili da parte del software (chiaramente, le inquadrature frontali del volto di Ian Curtis sarebbero state le più funzionali).

Il *deepfake designer* stesso si sarebbe premurato di comporre il dataset e di cercare quante più fotografie e quanti più video disponibili di Ian Curtis sul web, facendo anche ricorso alle fotografie in cui questi appare (come, per esempio, i celeberrimi scatti di Anton Corbijn) o singolarmente o accompagnato dai restanti membri della band.

La necessità di diverso materiale d'archivio in alta qualità ha inevitabilmente funzionato da ostacolo dal momento che, vista la breve carriera del frontman dei Joy Division, i dati visuali in nostro possesso sono per lo più i video più noti delle performance live dei Joy Division disponibili su Youtube, video in cui spesso la visuale del volto di Ian Curtis è impallata dalla presenza del microfono, salvo rari momenti in cui il cantante si esibisce nell'*epilepsy dance* o semplicemente si allontana dall'asta microfonica (in particolare *Joy Division – She's Lost Control (Live at Something Else Show) [Remastered] [HD]*, *Joy Division – Shadowplay (40th Anniversary Remaster) 1978 Granada TV, Live, Joy Division – Transmission [OFFICIAL MUSIC VIDEO]*) e una manciata di fotografie.

Durante la fase di ripresa del video destinazione, il volto avrebbe dovuto essere ben illuminato, ma l'apprendimento da parte di DeepFaceLab sarebbe avvenuto con qualsiasi tipo di illuminazione, con la sola riserva di dover correggere alcune ombre in fase di post-produzione attraverso software di editing come Adobe Premiere Pro per donare coerenza all'illuminazione del volto e quella del video di destinazione globale.

Il più importante accorgimento che ci è stato imposto per la produzione è stato quello di dover far cantare Ian Curtis (e quindi io stesso) privo di microfono. Salvo Compromesso ha affermato che quello del microfono sarebbe in realtà un non-problema: se il microfono fosse oggetto statico all'interno della ripresa sarebbe facilmente escludibile dalla maschera senza arrecare al lavoro eventuali errori di sfarfallio dovuti all'eccessivo movimento di un'ostruzione davanti al viso. Nel caso in cui ci fosse un'ostruzione dinamica sarebbe invece preferibile cantare senza microfono, per evitare gli errori suddetti. La scelta di cantare senza microfono è stata quindi dettata più che altro da una considerazione sulle tempistiche per la produzione (escludere il microfono con precisione attraverso il mascherino di DeepFaceLab avrebbe richiesto un lungo lavoro manuale). Per la ricostruzione del volto di Ian Curtis sono state essenziali le poche inquadrature in cui il suo volto appariva integralmente in primo piano. Senza di queste il software, impallato dal microfono, avrebbe faticato nel ricostruire la parte inferiore del volto, probabilmente lasciando inalterata quella porzione del mio volto. Se altresì avessimo voluto produrre un *deepfake* di una performance live con microfono, ricercando una mimesi maggiore con una reale performance, avremmo dovuto provare a ricostruire le stesse inquadrature dei video selezionati da

Youtube, così che DeepFaceLab potesse apprendere con facilità quali porzioni di spazio sostituire senza incappare in fastidiosi glitch.

Sono state quindi posizionate tre camere fisse nei pressi del palco, una che permetteva un'inquadratura frontale, un campo medio, e due inquadrature laterali in primo piano che, come ci era stato annunciato, hanno creato maggiori problematiche per il *face-swap*, causando sfarfallii di varia natura e asincronismi nella sovrapposizione dei volti. La produzione ha richiesto due take distinte, per provare a sincronizzare il labiale e per effettuare un'ulteriore inquadratura frontale, maggiormente ravvicinata al mio stesso volto, che sarebbe poi stato oggetto del *deepfake*, così da fornire più materiale possibile al *deepfake designer*.

Salvo Compromesso ha, ovviamente, ripercorso le fasi cruciali per preparare DeepFaceLab a un corretto apprendimento: inizialmente ha generato il *dataset* con i video e le fotografie di Ian Curtis, creando accuratamente delle maschere, indicando con precisione a DeepFaceLab la porzione di volto su cui avrebbe dovuto lavorare, applicando poi lo stesso processo al video destinazione, il video col mio volto su cui quello di Ian Curtis avrebbe dovuto essere sovrapposto, indicando anche qui attraverso una maschera la porzione di spazio su cui il software avrebbe dovuto lavorare nel dettaglio.

Nella fase di allenamento del modello, non sono state presenti particolari problematiche, se non, appunto, alcune inquadrature che hanno richiesto un maggior tempo di allenamento dal momento che i filmati di repertorio in cui è presente il volto di Ian Curtis sono pochi e molti dei quali in bassa qualità, risalendo alla fine degli anni settanta. Per ovviare a questa problematica Salvo Compromesso ha dovuto attuare l'*upscaling* di

molti frame, così da portarli a un livello di risoluzione accettabile e in modo che DeepFaceLab potesse essere agevolato nella riproduzione dei dettagli del volto.



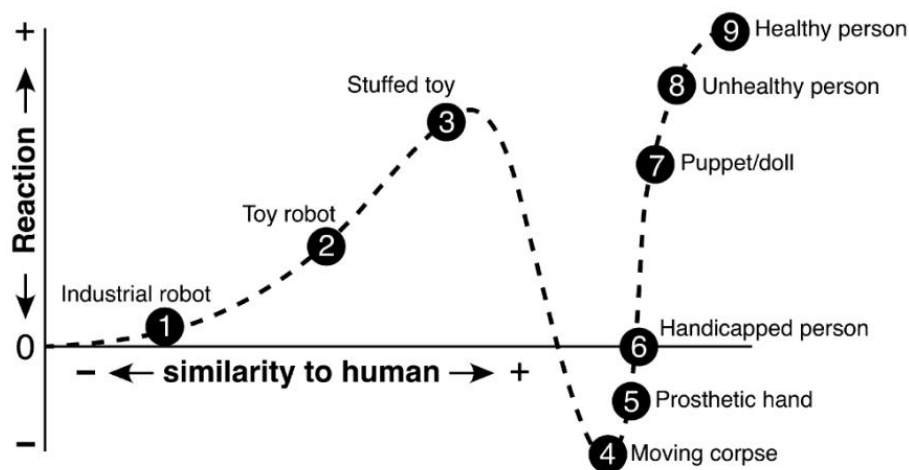
Un frame tratto da *The Eternal feat. Ian Curtis* precedente all'applicazione del bianco e nero, delle striature e del conseguente raffreddamento del video.

Una volta completato l'intero processo ed esportato il video con il volto di Ian Curtis sostituito al mio, ci siamo accorti di come, mantenendo inalterata l'illuminazione e la risoluzione del video, l'ibrido costituito dal mio corpo e dal volto del frontman dei Joy Division non risultasse affatto fotorealistico, e di come generasse una sorta di "effetto perturbante", ricadendo in tutto e per tutto all'interno dell'*uncanny valley*, la zona perturbante, teorizzata da Masahiro Mori.

Gli sviluppi tecnologici degli ultimi anni hanno portato alla nascita e allo sviluppo di androidi, di avatar fisici o virtuali dalle sembianze umanoidi all'interno del campo dei videogame, della robotica ecc.. Questi sviluppi

hanno portato le scienze cognitive e altre discipline a interessarsi all'interazione tra essere umano e androide. È proprio all'interno di questo campo che si è sviluppata la teoria di Mori, studioso nipponico, nel 1970, una teoria che riguarda la risposta emozionale dell'essere umano a delle entità non pienamente umane¹⁸⁴ non è da relativizzarsi al solo rapporto tra essere umano e robot. Entità come i corpi morti, gli zombie o gli androidi, attraverso la percezione di una decrescita di somiglianza, nella percezione di un'affinità, possono indurre una risposta emozionale negativa negli esseri umani (la cosiddetta *erie feeling*, un sentimento perturbante, al contempo di riconoscimento e mis-riconoscimento), causando confusione in merito alla categoria a cui l'oggetto guardato pertiene.

Mori ha utilizzato una curva per provare a dare forma alla relazione tra soggetto e oggetto dalle fattezze simil-umane percepito come *uncanny*.



Curva dell'Uncanny Valley

¹⁸⁴ M. Mori, K. F. MacDorman, N. Kageki, *The Uncanny Valley* [from the field], in *IEEE Robotics & Automation Magazine* 19.2, 2012, pp. 98-100.

L'attenzione di Mori si è concentrata sul rapporto coi robot per poi andare a investire altri campi. Nel suo testo, però, l'ingegnere si è concentrato più che altro sull'idea di una mano meccanica. Come mostrato dalla figura, alla base della zona perturbante vi è il cadavere che si muove, sostanzialmente lo zombie, una figura che, come la letteratura e il cinema insegnano, "vive" in una zona liminale tra la vita e la morte; dopodiché nel grafico vengono evidenziate diverse figure il cui aspetto umano aumenta esponenzialmente passando dalla mano protesica per arrivare fino alla "persona in salute".

Nell'articolo, Mori si concentra in particolare sulla mano robotica.

Si potrebbe dire che la mano protesica abbia raggiunto un grado di somiglianza alla forma umana, forse alla pari con la dentiera. Ciononostante, quando realizziamo che la mano che sembrava reale in prima battuta è in realtà artificiale, percepiamo una sensazione perturbante (*eerie sensation*)¹⁸⁵.

Sebbene la teoria dell'*uncanny valley* avesse un oggetto abbastanza specifico è stata più volte applicata alle rappresentazioni audiovisive di esseri umani. Quest'ipotesi è diventata infatti centrale negli ultimi vent'anni, dove il mondo della CGI e del videogame hanno visto una crescita esponenziale¹⁸⁶. Ne sono dimostrazione film come *Avatar* (2010) e *Avatar 2* (2023) di James Cameron, *Polar Express* (2004), *La leggenda di Beowulf* (2007), *A Christmas Carol* (2009) di Robert Zemeckis ma anche in

¹⁸⁵ *Ibid.*

¹⁸⁶ J. Kätsyri et al. "A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness", in "Frontiers in psychology 6.1", 2015, p. 390.

videogiochi come *Heavy Rain* (2010) di David Cage o *The Last of Us 2* (2020) sviluppato da Naughty Dog e gli esempi potrebbero continuare. La ricerca del fotorealismo in questi film o videogame si è dovuta in un modo o nell'altro scontrare con la teoria dell'*uncanny valley*. Secondo Christian Uva si può affermare che l'*uncanny valley* prodotta da questo tipo di rappresentazioni “deriva proprio da quella compresenza di *live action* e animazione in cui risiede la doppiezza e l'ambiguità di simili costrutti sintetici”¹⁸⁷. Da un lato infatti la *motion* e la *performance capture* sono ancora tecnologie indexicali in quanto fondate sul prelievo dell'“impronta dell'attore”¹⁸⁸, ma dall'altro “le nuvole di punti da esse prodotte nel computer non sono altro che la materia prima del successivo e decisivo processo di animazione”¹⁸⁹.

Il tentativo di superare questa risposta emozionale sembra avere a oggi fallito, dal momento che i meccanismi che producono l'effetto *uncanny valley* non possono essere ridotti a un unico fattore, investendo una dimensione culturale, biologica, emozionale e personale¹⁹⁰. Questa deriva dal fatto che alcune caratteristiche dell'oggetto guardato non sembrano appartenere a una chiara categoria oppure quando alcune sue caratteristiche sembrano appartenere a una certa categoria e alcune altre

¹⁸⁷ C. Uva, *Il cinema di (ri)animazione*, Fata Morgana Web, 2020, <https://www.fatamorganaweb.it/speciale-animazione-2/> (ultima consultazione il 14 febbraio 2023).

¹⁸⁸ C. Uva, *Ultracorpi. L'attore cinematografico nell'epoca della digital performance*, Bulzoni, Roma 2011.

¹⁸⁹ C. Uva, *Il cinema di (ri)animazione*, *ibid.*

¹⁹⁰ K. F. MacDorman e H. Ishiguro, *The Uncanny Advantage of using androids in cognitive and social science research*, in “Interaction Studies”, Vol. 7, n.3, pp. 297–333.

caratteristiche a un'altra categoria. L'*uncanny valley* si collega, poi, in qualche modo, alla *clonophobia* dal momento che

il problema con il clone... non è solo che è l'immagine vivente di una cosa vivente, ma che è indistinguibile, anonimo, posseduto da una sorta di "assenza di faccia" o impassibilità che si muove verso la sua espressione più radicale nel clone acefalo, senza testa, senza mente, creatura senz'anima, l'esemplificazione dell'organismo ridotto a pura esistenza¹⁹¹.

Così, un *deepfake*, clone che ricade nell'*uncanny valley*, sembra spesso privo di vita, una faccia-senza faccia, un volto incapace di reali espressioni emozionali, capace di trasmettere emozioni inquietanti e negative. Una volta passati all'ultima fase, quella della postproduzione, da noi personalmente attuata, e accortici di come il volto *deepfake* di Ian Curtis ricadesse pienamente all'interno dell'*uncanny valley* abbiamo provato ad adoperare delle strategie di riparazione. La scelta è stata, innanzitutto, per ovviare all'aspetto artificioso del volto innestato sul mio corpo, reso ancor più posticcio dai toni violacei assunti dall'inquadratura a cui conducevano le luci da palco del Centrale Rock Pub di Erba (CO) dove il video è stato girato, quella di virare i colori a un bianco e nero molto contrastato, così da livellare i difetti, le differenze nelle ombre tra il mio corpo, quello degli altri due membri della band e il volto appiccicato al mio di Ian Curtis. Con questo semplice accorgimento, il volto è sembrato subito più coerente col resto dell'illuminazione.

¹⁹¹ W.J.T Mitchell, *Cloning Terror*, *op.cit.*, pp. 36-37.

Il secondo passaggio fondamentale è stato quello di correggere la risoluzione, di diminuirla dai 1080 pixel iniziali a 480, così da ammorbidire i dettagli, nascondere le imperfezioni e far sembrare le restanti come la naturale conseguenza della bassa definizione totale del video. Si è inoltre optato per inserire una texture granulosa, così da *sporcare* ulteriormente la percezione, impallata da delle striature tipiche dei vecchi VHS.

Ciò che preme mettere in luce è qui la dialettica tra alta e bassa definizione di cui è stato oggetto il video (e di cui è potenzialmente oggetto ogni video *deepfake*), perfetto esempio di come la dimensione tecnico-materiale di un video *deepfake* possa essere lavorata al fine di produrre una rappresentazione mimetica. Nella produzione di *The Eternal – feat. Ian Curtis*, l'idea era infatti quella di recuperare quell'effetto di verità a cui fa riferimento Casetti in *La Galassia Lumière* in relazione alle *poor image* attraverso il quale "recuperiamo il mondo nei suoi processi e nei suoi significati intrinseci"¹⁹²¹⁹³. Infatti

Là dove i punti sono chiaramente visibili a occhio nudo, i messaggi veicolati dai media si collocano nel campo della bassa definizione, stimolando la partecipazione percettiva, cognitiva e sociale dello spettatore, invitato a riempire, per così dire, gli spazi rimasti liberi tra i punti. Là dove invece i punti non sono visibili a occhio nudo, lo spettatore si trova di fronte a una superficie piena, compatta, che non richiede alcuna forma di integrazione e di partecipazione¹⁹⁴.

¹⁹² Non è interesse di questa tesi lo studio dei *deepfake* in alta e bassa definizione, ma si è convinti della bontà dello spunto e ci si ripropone di sviluppare il tema in seguito.

¹⁹³ F. Casetti, *La galassia Lumière: sette parole chiave per il cinema che viene*, Bompiani, Milano 2015, pp. 189 -190.

¹⁹⁴ A. Pinotti, A. Somaini, *Cultura visuale*, Giulio Einaudi Editore, Torino 2016.



Un frame da *The Eternal feat Ian Curtis* (volto in seguito a applicazione bianco e nero e abbassamento definizione)

I video *deepfake* ci invitano a recuperare la meteorologia dei media di McLuhaniana memoria. Nel suo *Gli strumenti del comunicare*¹⁹⁵, il sociologo adoperava la più volte ripresa distinzione in media caldi (hot) e media freddi (cool) per cui la forma calda escluderebbe, mentre quella fredda includerebbe, qui utile per indicare come la dimensione tecnica dei *deepfake* possa agire sulla percezione del volto da parte dello spettatore. Caldi sono quei media capaci di investire i propri destinatari con una ricchezza di informazioni tale che a questi non è richiesta alcuna forma di integrazione; contrariamente risultano freddi quei media che ai propri ricettori

¹⁹⁵ M. McLuhan, *Gli strumenti del comunicare*, il Saggiatore, Milano 2008.

propongono dei messaggi in bassa definizione e che quindi necessitano di essere completati e interpretati da un punto di vista sociale, percettivo¹⁹⁶.

Attraverso l'abbassamento della definizione generale del video si può quindi recuperare un effetto di verità invitando lo spettatore a completare il volto in prima persona. Questo tentativo mimetico legato alla bassa definizione e alla mutabilità del soggetto può avere radici lunghissime e non necessariamente essere collegato ai media visivi. In *Mimesis. Il realismo nella letteratura occidentale*, Erich Auerbach, facendo riferimento a *L'humaine condition* di Montaigne (nel quale lo scrittore e filosofo francese prova a rappresentare se stesso mettendo in risalto "il vacillamento, l'instabilità, la mutabilità del suo oggetto"¹⁹⁷), scrive

Chi vuol rappresentare esattamente e obiettivamente una cosa di continuo mutevole, deve esattamente e obiettivamente seguire i mutamenti della cosa stessa; deve descrivere la cosa compiendo il maggior numero possibile di esperimenti, a seconda di come essa gli si è di volta in volta presentata, e in tal modo può sperare di fissare il cerchio delle variazioni possibili e da ultimo ottenere un'immagine complessiva¹⁹⁸.

Quest'immagine complessiva del soggetto si genererebbe nel lettore, però, non attraverso l'abbondanza informativa ma, piuttosto, attraverso l'omissione, il vuoto, la falla, la bassa definizione.

... l'ordine è assai spesso rotto, alcuni membri vengono introdotti anticipatamente, altri del tutto tralasciati, affinché il lettore li completi. Il lettore deve collaborare; viene

¹⁹⁶ A. Pinotti, A. Somaini, *op.cit.*, p.196.

¹⁹⁷ E. Auerbach, *Il realismo nella letteratura occidentale*, Einaudi, Torino 2000.

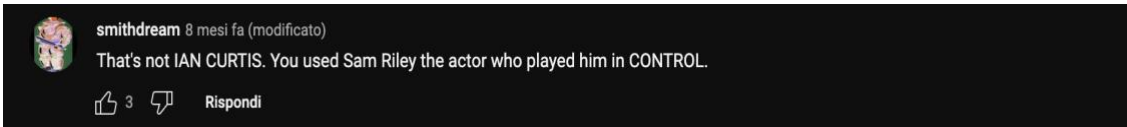
¹⁹⁸ E. Auerbach, *ibid.*

introdotto nel movimento del pensiero, però in ogni istante da lui ci si attende che sosti sorpreso, esami e integri... Spesso egli [Montaigne] tralascia congiunzioni e altri legamenti, però li suggerisce; salta membri del suo pensiero, ma sostituisce ciò che manca con una specie di contatto che si stabilisce spontaneamente fra i membri non in stretto legame logico¹⁹⁹.

Così, il volto percepito dallo spettatore all'interno di un *deepfake* sarà una sintesi dei vari video e delle fotografie provenienti da differenti periodi storici e delle differenti fattezze del volto e dei vari formati ad esso collegati con cui lo spettatore ha imparato a vedere il volto "originale" del rappresentato, dati con i quali si dovrà nutrire il *dataset*. Il *deepfake designer*, come nel nostro caso, potrà poi spesso lavorare sulla bassa definizione, rendendo evidenti i punti, omettendo delle informazioni così da stimolare la percezione dello spettatore.

A questo proposito, è interessante notare come la percezione del risultato finale, che non poteva che intendersi come il prodotto generato attraverso il confronto con il volto dello Ian Curtis originale, abbia suscitato reazioni miste. In particolare, il volto risultante in seguito all'abbassamento della definizione e all'applicazione del filtro bianco e nero è stato in diverse sedi percepito non solo come un falso, ma come il risultato di un *face-swap* prodotto attraverso il ricorso al volto di Sam Riley, il quale ha interpretato lo stesso Ian Curtis nel film di Anton Corbijn *Closer* (2004), dedicato alla figura di Curtis e girato in bianco e nero, fino a sfociare in accuse di una volontaria falsificazione del *dataset*.

¹⁹⁹ E. Auerbach, *ibid.*



Il commento di un utente che accusa il *deepfake designer* Salvo Compromesso di aver utilizzato il volto di Sam Riley all'interno del dataset originale.



Il commento di due utenti all'interno del gruppo Facebook Ian Kevin Curtis. In Loving Memory. che riscontrano una maggior somiglianza del *deepfake* con il volto di Sam Riley che con quello di Ian Curtis.

Nonostante l'uso di soli frame del volto di Ian Curtis estratto dai video sopracitati, vi è stato più volte un mancato riconoscimento dell'originale che è finito per sfociare addirittura nel riconoscimento di un attore che ne ha eseguito l'interpretazione all'interno di una rappresentazione mimetica. Pur nella consapevolezza che all'interno di questo studio potrebbe rientrare un'ampia analisi dei processi percettivi dettati dall'uso comune del bianco e nero in *The Eternal feat Ian Curtis* e in *Closer*, ci sembra che queste affermazioni siano utili per dimostrare in questa sede come il video *deepfake* venga ritenuto come una copia degradata dell'originale a cui fa riferimento.

Marilyn Monroe gokkun: deepfake come raddoppiamento chiarificante

Nel suo saggio *Trust No One: Inside the World of Deepfakes* Michael Grothaus riassume le tappe fondamentali del lungo rapporto tra tecnologia e pornografia.

... ogni progresso tecnologico è stato applicato alla pornografia. La macchina da stampa portò alla collezione di versi erotici nel 1700. La fotografia a base chimica portò alla pornografia in fermo immagine dell'epoca vittoriana. Il cinema su celluloidi portò ai film per adulti e ai cinema a luci rosse nelle più importanti città del mondo a metà del novecento, e l'avvento del VCR nel 1970 e la conseguente vittoria del VHS sul Betamax permisero ai film pornografici di muoversi da quei cinema per adulti alle TV personali nell'intimità delle case nel 1980 [...]. In ogni caso, il più importante salto in avanti tecnologico nella storia del porno arrivò nel 1990 con l'avvento di Internet. [...] L'esplosione dei siti pornografici *tube* nel 2010 rese l'accesso al porno facile come mai prima²⁰⁰.

Non è negli scopi di questo capitolo trattare l'evoluzione di tale rapporto ma la sommaria descrizione fatta da Grothaus ci è utile per sottolineare come "la pornografia e i contenuti "per adulti" abbiano guidato lo sviluppo tecnologico dei media sin dall'Ottocento"²⁰¹ dal momento che il consumo di pornografia è fondato "sulla promessa di piaceri non ancora

²⁰⁰ M. Grothaus, *op.cit.*

²⁰¹ P. Bakker, S. Taalas, "Il business del'adult content: una rilettura del concetto di innovazione", in E. Biasin, G. Maina, F. Zecca (a cura di), *Il porno espanso: dal cinema ai nuovi media*, Milano-Udine, Mimesis Edizioni.

esperiti” e, di rimando, che “la domanda di prodotti innovativi e originali è continua”²⁰².

Lo stesso *celebrity porn*, entro il quale ricade la pornografia *deepfake*, non è certo un fenomeno nato parallelamente a questa pratica. Il porno *fake* o le immagini a sfondo sessuale non consensuali circolano da molto nella sfera pubblica, almeno da quando le tecnologie digitali e i nuovi media hanno permesso agli utenti del web di usare l’immagine delle celebrità per i propri scopi con facilità.

Il *deepfake porn* altro non è che una fase iperbolica del *celebrity porn*, uno sviluppo nella lunga storia del rapporto di sessualizzazione adoperato da pubblico e media sull’immagine delle star e delle celebrità che include, per esempio, varie forme di sex tape o foto di nudo hackerate²⁰³ ma è la sua prima forma a ottenere una tale visibilità grazie alla capacità di eliminare l’agentività e l’autonomia di entrambe le donne coinvolte nel video: da un lato abbiamo l’attrice, la performer, e dall’altro abbiamo la persona il cui corpo è “cancellato, montato e rimesso in circolazione”, donne che “sono rimosse e ridotte a semplici corpi senza una faccia (o nome)”²⁰⁴.

Inoltre, per la maggior parte, i leak delle immagini pornografiche, fino all’avvento dei software per il *deepfake*, facevano affidamento sulla possibilità di avere accesso a fotografie intime delle donne contro cui era diretto l’abuso. I *deepfake*, invece, offrono la possibilità di manipolare video

²⁰² S. Maddison, “Le mitologie pornografiche e i limiti del piacere. Max Hardcore e il porno estremo”, in E. Biasin, G. Maina, F. Zecca, (a cura di), *Il porno espanso: dal cinema ai nuovi media*, Mimesis, Milano-Udine.

²⁰³ E. Lawson, *Pixels, porn and private selves: intimacy and authenticity in the celebrity nude photo hack*, in “Celebrity Studies”, Vol. 6, n. 4, 2015, pp. 607-609.

²⁰⁴ S. Maddocks, *A Deepfake Porn Plot Intended to Silence Me: Exploring continuities between pornographic and political deep fakes*, in “Porn Studies”, Vol. 7, pp. 415-423.

e immagini senza necessariamente dover prima avere accesso a foto o video intimi reali del soggetto, rendendo chiunque una potenziale vittima.

Se è vero che la pornografia è sempre stato un campo privilegiato per le tecnologie di frontiera è altrettanto vero che i video *deepfake* e gli strumenti utilizzati per produrli sembrano essere la prima innovazione tecnologica che è stata spinta a un utilizzo di massa e alla popolarità dalla pornografia²⁰⁵ e che i video prodotti da deepfakes ritraenti Scarlett Johansson, Gal Gadot e Maisie Williams hanno segnato l'ingresso di un nuovo tipo di media nel panorama audiovisivo contemporaneo.

Il deepfake è infatti una tecnologia per l'ibridazione, la chimerizzazione, tanto da aver portato William Brown e David H. Fleming nel loro articolo *Celebrity headjobs: or oozing a squid sex with a framed-up leaky {Schar-Jø}* a parlare di "chtuluhmedia", riprendendo la metafora del mostro ibrido nato dalla mente di H.P. Lovecraft dal saggio *Chtuluchene. Sopravvivere su un pianeta infetto* di Donna Haraway.

Se i *deepfake* pornografici sono lavori sulla testa delle celebrità, allora come le seppie, i polpi e le piovre, forse dobbiamo tracciare l'emergere di una nuova creatura cefaloide, un nuovo cyborg mostruoso, senza ossa, di sola carne, forse con gli occhi ma senza volto. [...] Questi sono media intelligenti (artificialmente) che, come i cefaloidi, hanno schermi cromatofori dalla pelle viscida e corpi morfogenetici gocciolanti che cambiano forma e che usano con l'intento del mascheramento,

²⁰⁵ D. Kushner, *A Brief History of Porn on the Internet*, 2020, <https://www.wired.com/story/brief-history-porn-internet/> (ultima consultazione il 14 febbraio 2023).

dell'inganno e della comunicazione nella loro ecologia marina (profonda/immersiva)²⁰⁶.

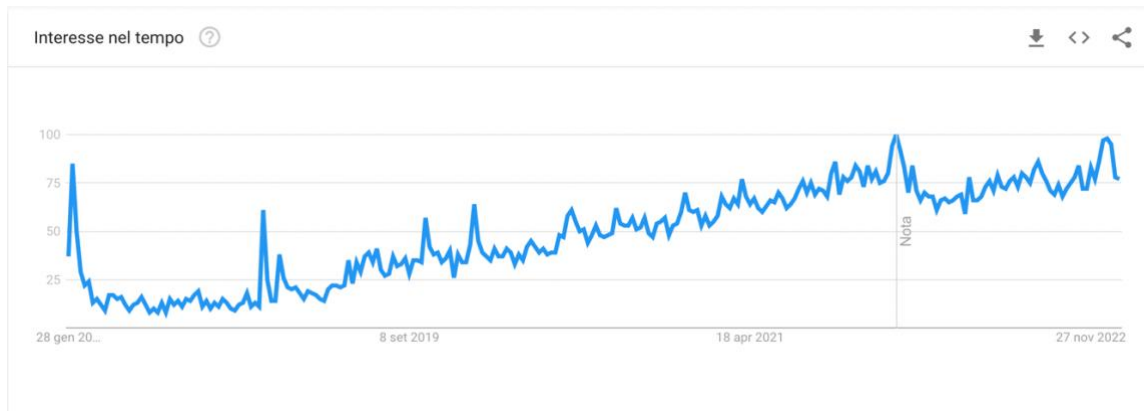
La descrizione fatta da Brown e Fleming dei video *deepfake* è certamente aulica e pretenziosa, ma ha il merito di spostare il fuoco sulle possibilità che la tecnologia apre all'ibridazione di corpi e forme differenti. Se la maggior parte dei video *deepfake* presenti su piattaforme come YouTube o Facebook possono infatti essere visti come lavori innocenti, divertenti o artistici che usano figure pubbliche vive o morte e se la letteratura sulle implicazioni dei *deepfake* continua a concentrare il fuoco sulle loro possibili conseguenze nell'ambito della privacy e delle fake news politiche per timore dell'instabilità sociale che questi potrebbero causare al di fuori dei social, gran parte dei video fino a oggi prodotti sono di carattere pornografico.

Ciò dovrebbe essere, secondo Adrienne de Ruyter, un tema di rilevanza immediata, dal momento che questi video sembrano violare il principio etico per cui "le persone hanno un volere proprio e non possono essere usate come strumenti per il raggiungimento dei propri fini e dei propri desideri" e che "come le altre forme di pornografia non consensuale, la pornografia manipolata digitalmente trasforma gli individui in oggetti di intrattenimento sessuale contro il loro volere, causando un'intensa sofferenza, umiliazione e danno alla reputazione"²⁰⁷ poiché prevede un esercizio di coercizione virtuale che permette al produttore di *deepfake* di decidere come una persona dovrebbe apparire all'interno di un video o di

²⁰⁶ W. Brown, D.H. Fleming, *Celebrity Headjobs: or ooziingsquid sex with a framed-up leaky* {Scar-Jø}, in "Porn Studies", Vol. 7, n. 4, 357-366.

²⁰⁷ A. De Ruyter, *The Distinct Wrong of Deepfakes*, in "Philosophy and Technology", n. 34, 1311-1332.

un audio attribuendogli gesti o discorsi nel quale il produttore stesso vorrebbe vedere il soggetto, ma che questi non ha mai effettivamente compiuto.



Il grafico dimostra come dal giorno della pubblicazione dell'articolo di Samantha Cole su Motherboard la domanda di pornografia *deepfake* sia cresciuta esponenzialmente. Fonte: Google Trends.

Secondo Sensity, nel 2019, la pornografia *deepfake* aveva una presenza online su diversi siti: otto dei dieci maggiori portali di pornografia ospitavano al loro interno video in cui il volto di attrici, cantanti o personaggi pubblici è stato sovrapposto a quello di svariate attrici pornografiche, oltre ai video ritraenti i personaggi animati. Nel suo report sullo stato dei *deepfake* la società indicava che dei 14.678 video analizzati – un campione non esaustivo, per loro stessa ammissione – hanno constatato come il 96% del loro dataset fosse pornografico. I quattro principali siti di pornografia *deepfake* avevano ottenuto, al 2019, circa 134 milioni di views. La loro maggioranza è disponibile tutt'oggi su siti dedicati come Adultdeepfakes e su MrDeepFakes²⁰⁸.

²⁰⁸ Creato dopo il ban del forum originale in cui i video deepfake sono nati su Reddit, MrDeepFakes possiede, secondo la stessa community, "la selezione più ampia di *fake celebrity porn*" e include video, opzioni di categorizzazione e una sezione "community" in cui i profili

Come si evince dai report pubblicati dall'azienda Sensity, i video pornografici sono la tipologia di video *deepfake* più presente su Internet, toccando circa il 90% della produzione totale. Al 2021, in tutto il mondo, il 64,3% dei target dei *deepfake* sono donne²⁰⁹. Il *deepfake porn* è, infatti, un fenomeno che ha come oggetto quasi esclusivamente i volti femminili. Contrariamente, i video non-pornografici analizzati da Sensity su Youtube vedono una maggioranza di soggetti maschili occidentali. Questa discrepanza è in parte dovuta secondo Ctrl Shift Face²¹⁰, al fatto che i capelli lunghi rischiano di inficiare la buona riuscita del video ma è impossibile non considerare il fatto che la maggior parte dei contenuti pornografici siano costruiti per un pubblico maschile²¹¹. Inoltre, l'industria dell'intrattenimento è la maggiormente colpita dalla pratica poiché la più mediaticamente esposta, con il 56,48% dei *deepfake* che ha come target personaggi dell'ambito televisivo, cinematografico o musicale²¹².

Se è vero quello che affermano Bode, Lees e Golding, ovvero che “i *deepfake* hanno avuto una traiettoria creativa che è iniziata con il rimpiazzo misogino del volto, ma che è successivamente ricaduta all'interno di uno

degli utenti possono essere visti oltre che un forum di discussione. MrDeepFakes offre inoltre una lunghissima serie di faceset dei volti delle celebrità più famose attraverso cui produrre i propri *deepfake* e consente inoltre la possibilità di indicare l'attrice nel video o la performer originale oltre all'aggiunta di vari tag come “masturbation” o “anal”. Mrdeepfakes ottiene i suoi introiti principalmente attraverso la pubblicità su altri siti ma offre anche la possibilità ai suoi utenti di fare donazioni attraverso i bitcoin e di iscriversi al sito così da contattare direttamente i produttori di *deepfake* e inoltrare le proprie richieste.

²⁰⁹ G. Patrini, *The State of Deepfakes 2020: Update on Statistics and Trends*, Sensity, Amsterdam 2021.

²¹⁰ Faccio riferimento alle dichiarazioni rilasciatemi da Ctrl Shift Face, celebre *deepfake designer*, di cui un estratto può essere letto su Film Tv N°1/2020.

²¹¹ C. Öhman, *Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography*, in “Ethics and Information Technology”, n. 22, pp. 133-140.

²¹² G. Patrini, *op.cit.*

spazio pubblico di sperimentazione giocosa, dove il suo uso ha proliferato”²¹³ è altrettanto vero che i dati riportati da Sensity dimostrano come i video *deepfake* a sfondo pornografico, allo stato attuale, siano la maggioranza²¹⁴.

Internet, afferma James Bridle in *Nuova era oscura*, è in grado di amplificare e attivare i nostri desideri latenti²¹⁵ e questi desideri sembrano aver dato vita a centinaia e centinaia di video *deepfake* categorizzati su siti per attrici o per pratica su siti dedicati, spesso evidenziando il nome dell’attrice nel titolo stesso (es. *Cara Delevingne Sex Crazy*, *Kristen Stewart bathroom shower and fuck* o *Scarlett Johansson POV Fuck* ma anche ibridi ancor più inquietanti donna-personaggio d’animazione, come *ScarJo as Scarlet of Final Fantasy VII (Happy Halloween)*, *Princess Zelda (Breath of the Wild) sucks and fucks the Master Dildo*, *Greta Thunberg – 3D*).

La maggior parte di questi video sono imperfetti, il falso è palese e molti degli utenti del sito sono assolutamente consapevoli che ciò a cui stanno assistendo non è reale²¹⁶ ma, nonostante ciò, sembrano usufruirne sospendendo la credulità. Nelle community dei siti di pornografia *deepfake* l’essere umano viene “disumanizzato”, come osservato da Newton e Stanfill²¹⁷, nel momento in cui i partecipanti vi fanno riferimento come a dati sui cui lavorare, per cui non ci sarebbe reale differenza tra un corpo

²¹³ L. Bode, D. Lees, D. Golding, *Editorial the digital face and deepfakes on screen*, in “Convergence: The International Journal of Research into New Media Technologies”, n. 27 (4).

²¹⁴ G. Patrini, *op.cit.*

²¹⁵ J. Bridle, *Nuova era oscura*, Nero, Roma 2019.

²¹⁶ S. Cole, *op.cit.*

²¹⁷ O. Newton, M. Stanfill, “My NSFW video has partial occlusion: deepfakes and the technological production of non-consensual pornography”, in *Porn Studies*, Vol. 7 (4), pp. 398-414.

animato e un corpo reale e tra il volto di una celebrità defunta o di una celebrità vivente.

Si è precedentemente accennato alle resurrezioni digitali prodotte attraverso i *deepfake*. Qui si annoterà come sia in ambito pornografico che si è dato vita alla maggior parte degli esempi di resurrezioni digitali (esclusivamente a livello amatoriale) attraverso il *deepfake*. Su MrDeepFakes, per esempio, si può trovare una variegata serie di film dedicata a Carrie Fisher *Young Carrie Fisher masturbating as Princess Leia* o *Princess Leia (Carrie Fisher) rides again* ma anche singoli video come *Liz Taylor Lookalike BJ* o *Grace Kelly: Anal Queen*. Una delle attrici più volte riportata in vita in ambito pornografico è stata Marilyn Monroe.



Un estratto dal video *Ice Queen Elsa Deepfake Hot Sex Tape*

Marilyn Monroe gokkun è un video creato dal *deepfake designer* mondomonger in cui il volto di Marilyn Monroe viene innestato sul corpo

della pornoattrice Jenna Ivory in un video pornografico da lei interpretato e che si sviluppa partendo da una videointervista per concludersi in una performance sessuale di Marilyn insieme a diversi uomini: una *gang bang*. È in questa intervista che si instaura l'impianto narrativo del video, un'ucronia in cui Marilyn Monroe non è una star del cinema tradizionale ma del cinema a luci rosse. È lei stessa a spiegarlo, sdraiata prona all'interno di una lingerie nera, le gambe piegate all'insù in una posa innocente, la chioma biondo platino molto simile a quella che era solita sfoggiare, un anello, un braccialetto e due grossi orecchini sfavillanti, quasi a farsi gioco della sua celebre performance in *Gli uomini preferiscono le bionde* di Howard Hawks (1953), in cui l'attrice canta *Diamonds Are a Girl's Best Friend*. Nel dialogo con quello che sembra essere il produttore del video, Marilyn spiega che le motivazioni che l'hanno portata a entrare all'interno dell'industria del porno sono di carattere personale, legate a un bisogno di vendetta nei confronti dell'ex ragazzo che l'avrebbe tradita più volte. Terminata la breve intervista iniziale, il film continua con un intermezzo musicale durante il quale mondomonger, mentre Marilyn rompe la quarta parete guardando dritto in camera, prima sdraiata sul letto, successivamente carponi, decide di alternare il volto della Monroe a quello della pornoattrice, autodenunciando la falsificazione.



Un frame dal video *Marilyn Monroe gokkun*

Alla luce di ciò, si può affermare che *Marilyn Monroe gokkun*, venendo distribuito su siti dedicati alla pornografia *deepfake* e autosmascherando la falsificazione, sia stato prodotto senza alcuno scopo di inganno e che non vi sia rischio alcuno che questi possa essere considerato come un video realistico. Il video *deepfake*, piuttosto, innesca un involontario processo di riconoscimento per simboli della natura mediatica, o quantomeno di una sua parte, di *Marilyn Monroe*. Il video, che potrebbe intrecciarsi con ulteriori studi sui *porn* e *gender studies*, sembra parlare di come il corpo di Marilyn Monroe sia stato un oggetto privilegiato, se non una vera e propria ossessione, dello sguardo maschile, uno sguardo che ha portato a un'ipersessualizzazione da parte di studios, riviste, ecc. della sua immagine, della sua figura pubblica, e di come l'attrice abbia utilizzato il suo nome d'arte (quello di battesimo era Norma Jane) come una maschera che allontanasse il pubblico del dramma personale derivante da un disturbo borderline

della personalità²¹⁸, una maschera che la donna ha perseguito anche sul piano fisico attraverso una maniacale cura del proprio aspetto, dedicando numerose ore del giorno e della notte a pratiche di trucco, alla piega dei capelli e al provare abiti direttamente cuciti sul suo corpo²¹⁹. Alcuni dati storici possono esserci utili per inquadrare il motivo d'interesse di questa realtà alternativa.

Vi è un *fil rouge* a legare indissolubilmente Marilyn Monroe e pornografia, in particolar modo a causa delle foto di nudo che la donna ha dovuto scattare per necessità economiche (le celebri fotografie scattate da Tom Kelly nel 1948, per la quale fu pagata unicamente cinquanta dollari, pochi soldi che erano a lei essenziali per la sopravvivenza, come dovette ammettere solo quattro anni dopo per evitare lo scandalo mediatico²²⁰), e per le vicende che la connettono alla nascita e successivo exploit della rivista erotica Playboy. Hugh Hefner, fondatore dello storico magazine, comprò infatti i diritti del servizio fotografico fatto da Kelly a Norma Jane, in modo da pubblicarli sul primo numero della rivista, uscito nel dicembre del 1953. La celeberrima fotografia pubblicata all'interno di quel numero, che vede Marilyn Monroe appoggiata a seno scoperto su un telo di velluto rosso, venne utilizzata per rendere l'attrice la prima *playmate* la cui immagine poteva essere appesa nella stanza di ogni uomo, una locandina nella casa di chiunque²²¹. L'immagine del corpo nudo di Marilyn Monroe diventò così eternamente replicabile, a disposizione di chiunque, il che ha

²¹⁸ L. Dell'Osso, R. Dalle Luche, *L'altra Marilyn. Psichiatria e psicoanalisi di un cold case*, Le Lettere, Firenze 2016.

²¹⁹ V. Codeluppi, *Il divismo. Cinema, televisione, web*, Edimill, Bologna 2016, p. 76.

²²⁰ *Ibid.*

²²¹ *ibid.*

portato numerosi commentatori ad affermare che con quest'immagine, con la pubblicazione di questa fotografia, Hugh Hefner abbia inventato la pornografia moderna; questo perché, utilizzandola, l'uomo è riuscito a creare un'icona che potesse essere riprodotta in maniera seriale, industriale, in modo da suscitare effetti corporali negli uomini, per generare eccitazione sessuale²²². Marilyn, tra l'altro, questa volta semicoperta, appariva anche sulla copertina del numero, il che spiega la sovraesposizione mediatica e l'importanza che la donna ha giocato come ruolo chiave all'interno di quella prima offensiva nella rivoluzione sessuale statunitense²²³.

Una diceria voleva addirittura che Norma Jane, esattamente come accaduto per il servizio fotografico di Tom Kelly, in un periodo di forti ristrettezze economiche, avesse dovuto prendere parte attivamente come attrice alla produzione di diversi film a stampo pornografico. Negli anni ottanta si è speculato persino su due vecchi film a luci rosse riuniti sotto il titolo di *Apple Knockers and the Coke Bottle*, mostranti rispettivamente una scena di sesso e uno striptease: entrambi, in uno dei primi casi di *fake porn* con al centro il nome di Marilyn Monroe, furono fatti circolare nel tentativo di far credere che la donna ritratta fosse a tutti gli effetti Marilyn, ma col tempo si è appurato che entrambi i video ritraessero un'altra donna²²⁴.

Non solo il mondo dell'erotismo e il mondo del porno, anche il mondo del cinema hollywoodiano ha concepito Marilyn principalmente come oggetto a disposizione dello sguardo maschile. La stessa Fox, secondo

²²² B. Preciado, *Pornotopia. Playboy: architettura e sessualità*, Fandango, Roma 2011.

²²³ A. Summers, *Marilyn. Tutti i segreti di una vita*, Bompiani, Milano 1985, p. 74.

²²⁴ *Ivi*, p75.

Thomas B. Harris, avrebbe adottato una strategia promozionale volta a far percepire Marilyn in qualità di “*playmate* ideale”, come puro oggetto del desiderio del maschio²²⁵. Le inquadrature di Marilyn Monroe passate alla storia come celebri esprimono sempre, per quanto con ovvia maggior pudicizia rispetto alla fotografia contenuta in *Playboy*, la sua natura sessuale: in *Quando la moglie è in vacanza* la donna viene immortalata nell’atto di tenere abbassata la gonna alzata da una griglia d’aerazione, mentre nell’incompiuto *Something’s Got to Give*, la donna viene nuovamente ripresa totalmente nuda a bordo di una piscina, in un’immagine che verrà ripresa da numerose testate, suscitando enorme scandalo.

Probabilmente Marilyn Monroe veniva considerata come oggetto sessuale perché capace di esprimere agli occhi degli spettatori statunitensi una forma di sessualità ingenua e passiva, permettendo così agli uomini di non sentirsi minacciati dalla sua intensa carica erotica: sembrava offrirsi come oggetto totalmente disponibile per il piacere dell’uomo. Era desiderabile perché vulnerabile, passiva, inoffensiva. Ciò non significa, afferma Vanni Codiluppi, che nei suoi film sia stata oggetto di violenze fisiche, ma che veniva “frequentemente manipolata o umiliata da parte dei protagonisti maschili”, capaci di abusare di quell’immagine di vulnerabilità “rafforzata da ciò che parallelamente veniva portato a conoscenza del pubblico sulla sua vita privata: l’infanzia difficile,

²²⁵ T. B. Harris, *The Building of Popular Images: Grace Kelly and Marilyn Monroe*, in “*Studies in Public Communication*”, n. 1. 1957, pp. 45-58.

l'etichetta di ninfomane frigida, la sterilità che l'ha portata a diverse false gravidanze, i tre matrimoni finiti male"²²⁶.

Sebbene dalle varie dichiarazioni dei suoi amanti e compagni ne emerga una rappresentazione discordante e incerta, o forse proprio per questa ragione, è innegabile che la sessualità della Monroe abbia esercitato un indubbio fascino sugli studiosi, divenendo oggetto di dibattito e di studio. Nel suo *Marilyn. Tutti i segreti di una vita*, Anthony Summers restituisce il ritratto di una sessualità complessa, assai differente da quell'atto (quello sessuale) per lei "... naturale quanto fare colazione al mattino" con cui la descrive il primo marito Jim Dougherty²²⁷.

Gli uomini che si susseguirono nella vita di Marilyn e che sono stati intervistati per questo libro, offrono un'immagine ben diversa della sua sessualità. Queste persone, e la stessa Marilyn nei colloqui con il suo psichiatra, il dottor Greenson, sembrano parlare di una donna che trovava scarsa soddisfazione nel sesso. Il mutamento può essere avvenuto dopo le brutte esperienze personali che seguirono il matrimonio con Dougherty²²⁸.

Nella vita privata di Norma Jane si susseguirono diversi uomini e matrimoni. La donna sembrava sempre essere soggiogata dalle numerose figure maschili, talvolta figure di enorme potere, che incrociavano la sua strada. La sua infinita serie di relazioni, più o meno importanti (per citare alcuni nomi riportati da Summers: Robert Slatzer, Tommy Zahn, i fratelli Chaplin, Joe Schenck, James Bacon, Fred Karger, Jonny Hide, Elia Kazan,

²²⁶ V. Codeluppi, *op.cit.*, pp. 78-79.

²²⁷ A. Summers, *Marilyn. Tutti i segreti di una vita*, Bompiani, Milano 1985, p. 32.

²²⁸ *Ivi*, p.33.

Nico Minardos...), cominciò a sedici anni, proprio quando sposò Dougherty, allora ventunenne che abitava nel suo vicinato, in quello che sarà solo il primo di tre matrimoni. Dopo il termine della relazione con Dougherty, e in seguito a un lungo periodo di corteggiamento, Marilyn sposò il celebre giocatore di baseball dei New York Yankees Joe DiMaggio con un matrimonio pubblico che, però, ebbe termine solamente nove mesi più tardi, con la donna che fu trovata fuori dalla sua casa a Beverly Hills in lacrime, dopo aver presentato istanza di divorzio.

Il matrimonio con DiMaggio fu costellato di polemiche. L'uomo è stato più volte descritto come iracondo, geloso e possessivo, nonché capace di atti violenti contro la moglie. Sembra, inoltre, che durante i due anni di corteggiamento del giocatore di baseball Marilyn abbia avuto almeno altri quattro uomini e che con uno di questi non sarebbe da escludere sia persino convolata a nozze in gran segreto. Il matrimonio con lo scrittore e drammaturgo Arthur Miller sembrò portare un'iniziale stabilità alla donna, che provò a ritirarsi a vita domestica in Connecticut per studiare pittura e a ricoprire il ruolo di musa ispiratrice per l'artista ma, segnala Lena Dunham in suo articolo su Vogue, il suo corpo, la sua bellezza e la sua femminilità saranno sempre al centro del discorso anche con il nuovo marito, come rivelerebbe una lettera inviata da lui stesso: "E mentre tu sarai lì a cucinare la colazione, io bacerò il tuo collo e la tua schiena e i dolci meloni del tuo sedere e il retro della tua ginocchia e ti farò voltare e bacerò i tuoi seni, e le uova bruceranno..."²²⁹.

²²⁹ L. Dunham, *What Marilyn Monroe Means to Me*, 2022, <https://www.vogue.com/article/what-marilyn-monroe-means-to-me-lena-dunham> (ultima consultazione il 14 febbraio 2023).

In seguito alla fine del matrimonio con Arthur Miller, Marilyn avrà numerosi ulteriori amanti e compagni, a partire da Frank Sinatra (già suo amante dopo la rottura con DiMaggio) e, in particolare, con il presidente degli Stati Uniti John Fitzgerald Kennedy e col fratello Robert. Del rapporto col primo si ricorda in particolare maniera la ripresa della famosa performance in onore del quarantacinquesimo compleanno del Presidente, nella quale la Monroe si esibisce nell'interpretazione di *Happy Birthday, Mr. President*, presentata sul palco come una musa da Peter Lawford che ne mette repentinamente in risalto la femminilità: "Signor presidente", lo sentiamo dire, "in occasione del suo compleanno, questa splendida signora non si limita a essere avvenente: è anche puntuale". Marilyn però non compare e Lawford riprende con ironia: "Signor Presidente, poiché nella storia del mondo dello spettacolo non c'è forse stata una donna che abbia significato tanto quanto lei, che abbia fatto di più...", suscitando l'ilarità del pubblico, per poi essere interrotto dall'arrivo dell'attrice. Una volta giunta sul palco, Marilyn si esibisce cantando "tanti auguri signor presidente", per poi continuare con dei versi inediti recitati sulla musica di *Thanks for the Memory*: "Grazie signor presidente per tutto quello che hai fatto, per le battaglie che hai vinto, per come affronti l'U.S. Steel, e i nostri problemi, a tonnellate, noi ti ringraziamo, tanto tanto"²³⁰.

Marilyn si mostra qui come oggetto fragile a disposizione dello sguardo dell'uomo più potente del mondo, in una performance diventata simbolo di seduzione, erotismo e malizia, tanto da aver scatenato nel corso del tempo una lunga serie di parodie, su tutti la sequenza di *Breaking Bad* in

²³⁰ A. Summers, *op.cit.*, pp. 296-297.

cui Skyler, la moglie del protagonista Walter White, si esibisce in un'imitazione di Marilyn per il compleanno del suo capo Ted Beneke, con il quale tradirà il marito da lì a poco. Il rapporto di sottomissione instauratosi tra J.F.K e Marilyn Monroe sarà poi ripreso nel suo *Blonde* da Andrew Dominik in una sequenza dall'alto valore allegorico, declinata anche qui in chiave pornografica, che ha destato scandalo, in cui si può vedere la Monroe, interpretata da Ana De Armas, praticare una fellatio a un Kennedy che la appella violentemente ("lurida porca", "brutta sporcacciona") mentre la ragazza tenta di trattenere un conato di vomito.

Pierpaolo Pasolini, toccato dalla precoce scomparsa dell'attrice e dalla sua vita segnata dall'abuso e dalle dipendenze, la descrisse nel suo film documentario *La rabbia* come il simbolo di una bellezza antica e pura crudelmente distrutta dalla cattiveria della società moderna²³¹, una società ossessionata con il suo aspetto fisico, con il suo corpo, divenuto oggetto del desiderio, delle brame di uno sguardo maschile che, anche seguendo la lettura data da Andrew Dominik nell'adattamento del romanzo di Joyce Carol Oates, l'avrebbe lentamente logorata.

Nel film vediamo un'immagine altamente esplicativa di questa dinamica: Marilyn ha da poco avuto un aborto spontaneo, e fatica a tornare alla sua vita da diva del cinema, ma è il giorno della prima di *A qualcuno piace caldo* e la donna si deve recare al cinema per la presentazione della pellicola. La donna si avvicina all'ingresso, trasportata da una limousine che viene accerchiata da una folla trepidante e inneggiante il suo nome. Uscita dalla macchina e investita dai flash dei fotografi la donna saluta il

²³¹ V. Codeluppi, *op.cit.*, p. 76.

pubblico con grazia, alzando gentilmente la mano destra, ripresa di schiena e di profilo in piano americano, dopodiché la steadicam sposta il suo fuoco sul pubblico con una carrellata orizzontale che sembra corrispondere allo sguardo di Norma Jane. I volti degli uomini che la circondano e si dimenano sono volti deformi, di esseri spaventosi e inquietanti, uomini che protendono le mani verso di lei come a volerla afferrare, dilaniare, possedere. Sono uomini-mostri, uomini-vampiro. I loro occhi sono enormi, a palla, privi di vita. Le loro mascelle sono spalancate in maniera innaturale, la lingua ipertrofica e piatta. I loro molari pronti a dilaniarne la carne. La camera torna poi con una panoramica orizzontale al volto di Norma Jane, ripresa col marito Arthur Miller, interpretato da Adrien Brody.

In conclusione, come per il film di Dominik, la rappresentazione mimetica di *Marilyn Monroe gokkun* sembra duplicare Marilyn Monroe per evidenziare le dinamiche dello sguardo a cui è stata soggetta lungo tutta la sua carriera, una componente fondamentale della sua esperienza di attrice e personaggio mediatico. L'opera attiva un piacere da ritrovarsi nel riconoscimento di parte della storia di Marilyn Monroe come oggetto mediale, chiarendo un aspetto storico della sua figura, informandoci sulla sua iper-sessualizzazione, sugli abusi ricevuti, sullo stretto rapporto che è involontariamente intercorso tra lei e l'ambito erotico/pornografico, sulla sua subordinazione allo sguardo maschile attraverso i media del Novecento.

Prende corpo in *Marilyn Monroe gokkun* una versione stilizzata dell'attrice, una realtà alternativa dove la donna non è diventata una star del cinema ma della pornografia per vendicarsi di un uomo; il piacere

mimetico non deriva tanto dal riconoscimento del volto della donna (sebbene la sovrapposizione sia ben realizzata per la maggior parte del video), o dall'esposizione della tecnica del *deepfake* (l'alternarsi del volto reale dell'attrice defunta e di quella pornografica), ma dalla comprensione del rapporto tra il suo corpo, la sua figura mediale, lo sguardo maschile e l'ambito pornografico. Il video dialoga a sua insaputa con la realtà sensibile, con la storia, reale e simbolica, di Marilyn Monroe, non rappresentando realisticamente e pedissequamente i dettagli di una realtà storica, ma cogliendone una verità profonda e allegorica, un aspetto recondito della sua essenza.

All'interno di questo mondo ipotetico si dà vita a una seconda natura per Marilyn Monroe, una natura che chiarisce alcuni aspetti del reale, informandoci sul soggetto della rappresentazione attraverso le "cose che potrebbero avvenire e cioè quelle possibili secondo la verisimiglianza o la necessità"²³², lontana da un rapporto strettamente referenziale con le vicissitudini della star o dalla precisa rappresentazione degli eventi storici che l'hanno vista coinvolta ma mettendo in luce un universale, la "marilynità" che ha affossato Norma Jane.

²³² Aristotele, *op.cit.*, p. 61.

The Andy Warhol Diaries: deepfake come “incremento d’essere”

Dei *deepfake audio* si è precedentemente discusso all’interno del capitolo dedicato alla tassonomia dei *deepfake*. Un *deepfake audio* è, sostanzialmente, un audio prodotto facendo ricorso all’intelligenza artificiale all’interno del quale viene riprodotta una specifica voce, con i suoi attributi timbrici, prosodici, il suo accento, la sua grana, per riprendere un termine caro a Roland Barthes²³³. In ambito musicale, artisti-ricercatori come Holly Herndon stanno sperimentando le potenzialità dell’intelligenza artificiale applicata alle melodie vocali. In particolare, l’autrice di *Proto*, da anni attiva nel campo dello studio di una possibile applicazione del *machine learning* in ambito musicale, ha sviluppato insieme all’azienda Never Before Heard Sounds, azienda che si occupa di produrre nuove tecnologie utili per i musicisti, il software Holly +.

All’utente del sito di Holly + è concessa la possibilità di fare l’upload su un sito dedicato di un file audio contenente una melodia registrata con la sua stessa voce o con qualsiasi voce questi voglia sostituire con l’unico limite di una durata massima di cinque minuti. L’intelligenza artificiale è in grado di riprodurre questa melodia sostituendo alla voce del cantante o della cantante la voce della stessa Holly Herndon utilizzando il suo modello vocale. Never Before Heard Sounds ha infatti allenato un modello su diverse ore di parlato di Holly Herndon stessa facendo in modo che intonazione e ritmo del file originale restino inalterati, ma che alla voce al

²³³ R. Barthes, *La grana della voce. Interviste 1962-1980*, Einaudi, Torino 1986.

suo interno venissero sostituite la tessitura e la timbrica di Holly Herndon. Presentato ufficialmente al Sonar di Barcellona, anche nella sua veste *live*, Holly+ permette così a chiunque di creare brani o qualsiasi tipo di opera con la sua voce²³⁴, invitandoci a riflettere sul futuro di una pratica che avrà sempre maggiore rilevanza all'interno della produzione musicale, ponendo una serie di quesiti legali, commerciali, etici, tecnici²³⁵...

Non solo in ambito musicale: i *deepfake audio* stanno lentamente venendo integrati anche all'interno di più o meno note produzioni audiovisive, installazioni videoartistiche od opere di varia natura. A titolo esemplificativo, nell'estate del 2021, il regista premio oscar Morgan Neville era finito al centro di un acceso dibattito in seguito alla scelta di fare ricorso all'intelligenza artificiale per ricostruire la voce di Anthony Bourdain per il suo *Roadrunner*. Qualche anno prima, nel 2018, aveva trovato gli onori delle cronache un progetto condotto dal The Times in collaborazione con le aziende Cereproc e Rothco e presentato al festival di Cannes dal nome *JFK Unsilenced*. Nell'opera la voce dell'ex presidente degli Stati Uniti d'America veniva ricostruita perché potesse pronunciare il discorso mai proferito il giorno della sua morte per mano di Lee Harvey Oswald il 22 novembre del 1963. Un pluripremiato video *deepfake* come *In Event of Moon Disaster*, prodotto dal MIT e diretto da Francesca Panetta e Halsey Burgund, utilizzava invece un *deepfake* vocale di Richard Nixon per accompagnarlo al *face-swap* dell'ex presidente degli Stati Uniti d'America, applicato sul corpo di un attore con la sua stessa corporatura²³⁶: riuscendo

²³⁴ <https://www.youtube.com/watch?v=Wk6T2WmhuJw&t=1760s>

²³⁵ Dal sito di Holly +, <https://holly.plus/>.

²³⁶ K. Hao, *op.cit.*

nel difficile intento di ricostruire in maniera certosa il volto e la parlata di Nixon – un’operazione per la quale, però, non deve assolutamente essere sottovalutata l’importanza della performance attoriale – si ricostruiva il reale discorso che Nixon avrebbe dovuto tenere se l’allunaggio avesse dato esito tragico, un discorso mai pronunciato.

Per questo tipologia di prodotti (nel suo articolo fa esplicito riferimento a Microsoft’s Rembrandt, un falso quadro di Rembrandt prodotto a partire dall’analisi di migliaia di dipinti del pittore olandese da parte di un software sviluppato da Microsoft), media manipolati, inautentici, che riproducono oggetti provenienti da fonti autentiche di cui viene imitato lo stile, Luciano Floridi ha utilizzato il termine *ectype*²³⁷.

Non abbiamo una parola per definire un artefatto come Microsoft’s Rembrandt. Quindi lasciate che io proponga *ectype*. La parola deriva dal greco e ha un significato sottile che è qui piuttosto utile: un *ectype* è una copia, ma non una copia qualsiasi, ma piuttosto una copia che ha un legame specifico con la sua fonte (l’origine della sua creazione), il suo archetipo. In particolare, un *ectype* è l’impressione lasciata da un sigillo. Non è la cosa reale, ma è chiaramente collegata alla cosa reale in maniera significativa, autentica²³⁸.

Un *ectype* sarebbe il prodotto, quindi, di documenti realmente esistenti attraverso cui generare *output* fittizi, pressoché indistinguibili da un originale, ma in realtà non autentici; le cose possono però essere ancora più complesse. Le tecnologie digitali sono in grado, secondo Floridi, di

²³⁷ Il concetto è traducibile in italiano come “ectipo” (impronta).

²³⁸ L. Floridi, *Artificial Intelligence, Deepfake and a Future of Ectypes*, in “Philosophy & Technology”, n. 31, 2018, pp. 319.

separare la fonte archetipica, ad esempio ciò che era nella mente dell'artista, dal processo (stile, metodo, procedura) che conduce dalla fonte all'artefatto. Così, si possono avere *ectype* che siano autentici nello stile e nel contenuto ma non originali nei termini della fonte archetipica (è il caso di Microsoft's Rembrandt). Ma possono esistere anche *ectype* che sono originali nei termini della fonte archetipica, quindi che provengono da dove sostengono di provenire ma non autentici nei termini della performance, della produzione o del metodo (non sono gli stessi utilizzati dal produttore per distribuire l'artefatto). Si possono in definitiva avere *ectype* che siano artefatti autentici ma non originali, o artefatti inautentici ma originali²³⁹.

Quest'ultimo è il caso della serie televisiva *The Andy Warhol Diaries*, prodotta da Ryan Murphy e diretta da Andrew Rossi, forse la prima grande produzione a utilizzare un *deepfake* vocale come perno centrale della narrazione. La serie tv Netflix distribuita sulla stessa piattaforma nel marzo del 2022 si distingue dai numerosi prodotti seriali dell'azienda fondata da Reed Hastings e Marc Randolph proprio per aver messo al centro della narrazione una tecnologia così recente e controversa, suscitando diverse riflessioni sulla maggior o minor liceità dell'operazione²⁴⁰. Nella serie si può ascoltare la voce dello stesso Warhol leggere i diari che l'artista aveva formato narrando storie e pensieri personali alla diarista Pat Hackett. Stralci di questi scritti sono stati letti da un attore, Bill Irwin, e dati in pasto all'algoritmo di Resemble AI, azienda

²³⁹ *Ibid.*

²⁴⁰ A. Watercutter, *Una nuova serie di Netflix ha ricreato la voce di Andy Warhol grazie all'intelligenza artificiale*, 2022, <https://www.wired.it/article/diari-andy-warhol-serie-netflix-intelligenza-artificiale/> (ultima consultazione il 14 febbraio 2023).

che lavora nel campo dell'intelligenza artificiale, in modo che si potesse ovviare alla scarsa quantità di dati audio nelle loro mani (3 minuti e 12 secondi) dei quali l'algoritmo aveva studiato ogni singolo fonema in modo da provare a predire i fonemi assenti. Il modello è stato poi caricato sulla piattaforma web dell'azienda dove Rossi ha potuto digitare le frasi di cui aveva bisogno e lavorare con intonazione ed emotività²⁴¹.

Nel secondo episodio di *The Andy Warhol Diaries* sentiamo la voce *deepfake* di Andy Warhol raccontare un episodio della sua vita: "Sono andato in ufficio perché stavano costruendo un robot di me stesso" e parallelamente sentiamo una voce narrante, probabilmente un estratto da un servizio informativo, affermare: "Una volta Andy Warhol ha detto "Voglio essere una macchina", e ora il suo desiderio sta per diventare realtà. Warhol sta per essere clonato meccanicamente". Sullo schermo si succedono una dopo l'altra inquadrature di Andy Warhol, con i suoi movimenti ripetitivi e legnosi, il particolare di un robot che serra una mano, un suo mezzobusto, un primo piano nel quale lo scheletro meccanico muove leggermente la bocca e il cranio e una foto di repertorio in cui questo, in piedi, stringe da dietro, come in un abbraccio, il collo appartenente al calco certosino del volto di Andy Warhol. Una delle mani del robot è ricoperta da un guanto che tenta di imitare la pelle umana. Successivamente si può assistere al titolo di un articolo di giornale: "Andy Warhol vive – in forma robotica". Dopodiché il *deepfake* vocale di Warhol continua: "Mi sono seduto sulla sedia portata dai tizi del robot. Siamo rimasti seduti per un'ora, così hanno studiato la mia faccia per vedere se

²⁴¹ *Ibid.*

diventerò un bel robot. Hanno fatto il retro della testa. Mi hanno messo una retina per parrucche. Hanno preso della poltiglia e mi hanno coperto occhi e orecchie. Mi hanno detto “dammi un pizzicotto, se vuoi toglierlo”. Mi dava la nausea. È stato terribile. Finalmente hanno tolto lo stampo, ma lo hanno fatto cadere, e hanno detto che forse dovevano farne un altro. Io ho risposto: “No, non ne farete un altro”. Pensavo: “Ma perché ho accettato di farlo?””. Mentre la voce sintetica di Warhol recita questi stralci dal diario, il *found footage* utilizzato all’interno del film ci mostra un Warhol impaurito, con gli occhi smarriti, mentre degli addetti misurano il suo cranio, alternando fotografie in bianco e nero dell’operazione così dolorosa per il corpo dell’artista, mostrato nell’atto di digrignare i denti o in pose che ne comunicano lo stato di palese sofferenza corporale (in una di queste viene mostrato il suo volto completamente ricoperto di una sostanza bianca e argillosa), oppure in video in cui Warhol parla al telefono, vestito con una camicia rosa, dando ancora sfoggio dei suoi movimenti robotici.

Nel 1982 Andy Warhol fu a tutti gli effetti trasformato in un robot per una produzione artistica che non ebbe mai luogo da un dipendente di Disney Imagineering. Questa versione robotica di Andy Warhol è rimasta per circa vent’anni all’AVG, un brand che disegna e fabbrica personaggi meccanici per il cinema, per poi essere venduto a un collezionista privato. La figura meccanica, l’automa, era stato creato perché potesse fare da protagonista all’interno di una produzione multimediale che avrebbe compreso anche uno show televisivo²⁴² coprodotta da Andy Warhol e dall’impresario di Broadway Lewis Allen.

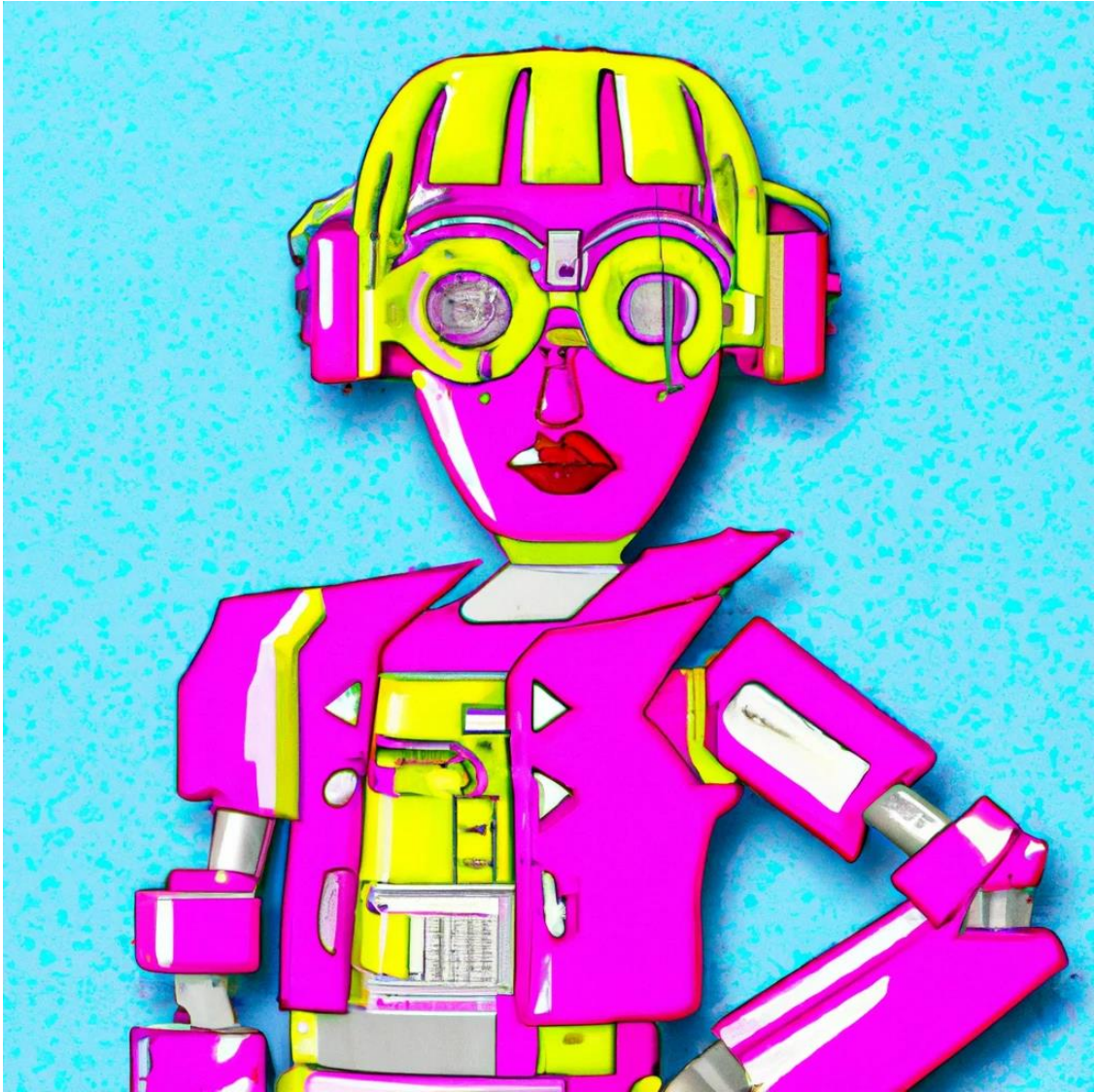
²⁴² L. Otty, P. Sellars, *The No Man Show: Technology and the Extension of Presence in the Work of Andy Warhol*, 2005.

L'opera fu provvisoriamente intitolata *Andy Warhol Overexposed: A no Man Show* e non ebbe mai vita per mancanza di fondi. Il costo per la produzione è stato di 400.000 dollari e ha previsto un lungo e faticoso lavoro sul corpo stesso di Andy Warhol: la sua voce monotona fu registrata, il suo stile di movimento ripreso su video, e il corpo fotografato da centinaia di angolazioni. Usando il riferimento fotografico John Davis, che all'epoca era a capo del reparto scultura, ha scolpito il cranio attorno a dei bulbi oculari personalizzati e una dentatura standard, creando stampe e pelli di gomma fusa oltre che un teschio di fibra che avrebbe sostenuto le parti più ossute del viso. Per il Warhol-robot fu creato un corpo di pannelli rimovibili in plastica, un guscio cavo che poi fu lasciato ai meccanici dell'AVG che avrebbero prodotto una muscolatura motorizzata. Una volta terminata la scultura completa, questa è stata aerografata, vestita e sormontata da una parrucca biondo platino in stile Warhol. Per l'opera non venne mai scritta nemmeno una sceneggiatura, per cui la prevista sincronizzazione del labiale non venne mai programmata, ma alcune funzioni erano già attive: tra queste, 54 movimenti del corpo, tra cui l'ondeggiamento del pomo d'Adamo e la possibilità per il robot di alzare le spalle. L'opera avrebbe dovuto rappresentare la quotidianità di Andy Warhol, mostrato a letto mentre dettava i propri diari al telefono e prevedeva anche la realizzazione di diversi bassotti robotici che rappresentassero i cani dell'artista²⁴³.

²⁴³ A. Ridenour, *The Automated Andy Warhol is reprogrammed*, Los Angeles Times, 2002, <https://www.latimes.com/archives/la-xpm-2002-may-16-wk-town16-story.html> (ultima consultazione il 14 febbraio 2023).

L'opera è nata da un'idea del regista Peter Sellars, che aveva l'intenzione di scrivere una sceneggiatura basata sui diari di Warhol e su ulteriori due suoi libri: *The Philosophy of Andy Warhol* e *Exposures*, ma Warhol non registrò mai un monologo e, in seguito alla sua morte, l'idea di far sì che un attore leggesse le battute o di provare a utilizzare le registrazioni esistenti dell'autore non ebbero seguito. L'idea di un Warhol robotico potrebbe però, in un certo senso, essere stata suggerita da Warhol stesso, nel suo esprimere il suo desiderio di diventare una macchina o un vero e proprio robot. A tal proposito, Thomas Sokolowski, direttore dell'Andy Warhol Museum di St. Peetersburg si è concentrato in un'intervista sulla figura già di per sé robotica di Warhol, come se fosse una parte della sua stessa essenza: "... guardate ad alcuni suoi frame mentre viene misurato per il progetto... i suoi movimenti sono molto legnosi, quasi robotici[...]. Pensava che finché qualcuno fosse sembrato Warhol e avesse suonato come Warhol, le persone non se ne sarebbero curate dal momento che stavano avendo la Andy Warhol Experience". Warhol stesso si diceva entusiasta dell'idea mentre il robot veniva costruito, sperando che questi potesse impersonarlo in una serie di discorsi pubblici, il che non era certo una novità per l'artista, il quale già nel 1967 aveva assoldato Alan Midgett per impersonarlo durante una lezione collegiale²⁴⁴.

²⁴⁴*Ibid.*



Andy Warhol robotico immaginato da DALL·E un algoritmo generativo in grado di produrre immagini a partire da un testo di input.

D'altronde l'artista ha da sempre mostrato un amore, se non una vera e propria ossessione per i processi meccanizzati o automatizzati, quasi come se i comportamenti ripetitivi fossero insiti nella sua stessa psicologia, nel suo modo di guardare al mondo. Il direttore dell'Andy Warhol Museum ha infatti affermato che, sebbene non ne abbia avuto certezza, crede che Warhol soffrisse della sindrome di Asperger, una forma di autismo molto lieve, per cui il fatto che Warhol continuasse a eseguire le stesse azioni in maniera continua e ripetitiva lungo tutta la settimana (come, ad esempio,

consumare un solo pasto al giorno o indossare lo stesso paio di mutande per più di trent'anni) assumerebbe un valore enorme nell'analisi del suo processo creativo²⁴⁵.

Più di ogni altro artista, Warhol ha infatti

coltivato la riconoscibilità delle immagini come interrogativo ontologico [...] rovesciando in ogni suo gesto le tradizionali nozioni di arte. Alla manualità ha opposto la ripetizione meccanica, sostituendo l'unicità dell'opera con la sua serializzazione. Ha selezionato immagini e temi della civiltà industrializzata rendendo inutile ogni separazione gerarchica della cultura e delle sue pratiche e in questo modo ha abdicato a ogni pretesa di immettere nel mondo nuove immagini riadattando quelle già esistenti²⁴⁶.

Il rapporto tra uomo e macchina è un elemento cruciale nella poetica di Warhol che, attraverso un abbondante ricorso alla tecnologia, problematizzò una lunga serie di questioni legate all'abilità artistica, da ritrovarsi nella dialettica tra presenza e assenza: un dipinto di Warhol non dipendeva direttamente dal suo tocco e avrebbe potuto essere prodotto anche in sua assenza, per esempio da un suo assistente. A titolo esemplificativo, *Marilyn Diptych* (1962) fu possibile grazie al ricorso a tecnologie come la serigrafia, in cui veniva utilizzata una rete per trasferire l'inchiostro o la tintura su uno specifico strato di una stampa, tranne nelle aree rese impermeabili all'inchiostro da uno stencil. A livello cinematografico, invece, in film come *Chelsea Girl* (1966), Warhol utilizzava lo split screen per far accedere lo spettatore a due spazi simultanei, e l'uso di una macchina da presa statica gli permetteva di non far sentire la sua

²⁴⁵ *Ibid.*

²⁴⁶ A. Mecacci, *Introduzione a Andy Warhol*, Laterza, Bari 2014 (edizione digitale).

mano, di auto-sopprimersi come regista. Inoltre, quando decise di dedicarsi alla letteratura, Warhol optò per una tecnologia come il registratore. *a: a novel* (1968) fu concepito proprio come la registrazione di una storia, un giorno nella vita di Andy Warhol ma senza l'immagine o la presenza di Andy Warhol²⁴⁷.

L'artista ha manifestato così per tutta la sua carriera la sua presenza attraverso il mezzo tecnico e la rimozione della sua figura che questo consentiva: Warhol si è sempre cancellato dal processo, ha utilizzato la tecnologia per raggiungere una telepresenza attraverso cui rimanere assente²⁴⁸ dando il là a una riflessione sul corpo che diventerà una prerogativa dell'intera pop-art da lì a venire, nella concezione di un corpo, di una figura umana che "non è più polo significativo attorno a cui si organizza l'ambiente"²⁴⁹, che diventa merce, un oggetto tra gli altri. Inoltre, nella ricerca pop, le immagini pubblicitarie, gli oggetti della scena urbana, le tecniche industriali, tutto viene recuperato nel moto di un'esperienza orientata dal visivo, dall'anonimo, dal quantitativo e dall'artificiale. L'esperienza dell'artista pop si dimostra quantitativa e impersonale, e l'attività del suo occhio si organizza non su schemi naturali, ma meccanici e automatici²⁵⁰. La pop art coglie e interpreta le trasformazioni apportate dai mass media e dai prodotti tecnologici all'esperienza quotidiana di tutti, cercando di penetrare nel comfort meccanizzato della modernità e trovando nella realtà e nella cultura industriale il luogo d'intervento dei

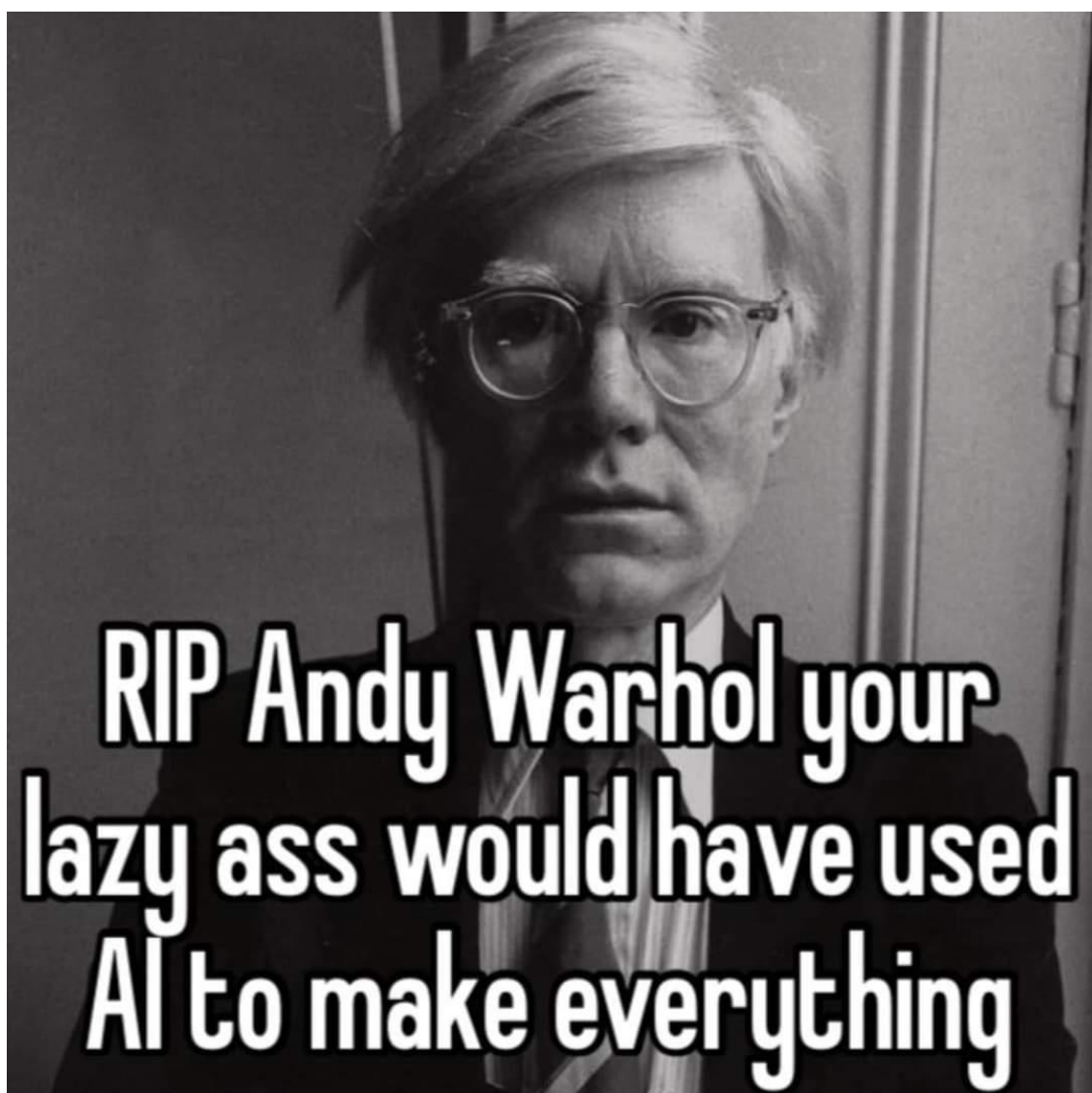
²⁴⁷ L. Otty, P. Sellars, *op.cit.*, p.11.

²⁴⁸ *Ivi.* p. 12.

²⁴⁹ C. Carriero, *Il consumo della Pop Art*, Jaca Book, Milano 2003, p. 89.

²⁵⁰ A. Boatto, *Pop Art*, Laterza, Bari 1983, pp. 68-69.

nuovi artisti²⁵¹. Non è un caso che sia la linea di vita americana il luogo privilegiato della ricerca della pop art, nata proprio nel paese più industrializzato del mondo dove ha messo in moto un processo di riappropriazione dell'unica esperienza possibile all'epoca, un'esperienza intersoggettiva e anonima all'interno di una dimensione di automatismi collettivi²⁵².



²⁵¹ *Ivi.*, p. 71.

²⁵² *Ivi.*, p. 73.

Un meme che recita “Riposa in pace Andy Warhol, il tuo culo pigro avrebbe usato l’intelligenza artificiale per produrre qualsiasi cosa”.

Procedendo lungo questa scia, al fine di inquadrare l’uso del *deepfake* nella serie *The Andy Warhol Diaries* come “incremento d’essere” di Andy Warhol, capace di informarci sulla sua essenza recondita e su parte della sua storia, ci è utile introdurre un altro concetto relativo alla *mimesis*, quello di “figura” proposto da Erich Auerbach in diversi suoi scritti, in particolare *Studi su Dante*, poi ripreso in *Mimesis. Il realismo nella letteratura occidentale*. Nel passaggio dedicato all’interno di *Studi su Dante* Auerbach segue l’evoluzione del termine “figura” nell’antichità pagana. A partire da Terenzio e dal suo significato originale di “formazione plastica”, il termine ha assunto diverse accezioni: in Cicerone era indice di diversi livelli stilistici; Ovidio lo attribuiva alla “copia”; Varrone lo utilizzava per fare riferimento all’“apparenza esteriore”, accanto all’idea che potessero esserci “figure” anche per il senso dell’udito²⁵³, idea che viene ripresa in Lucrezio, nelle accezioni più generiche del termine o come “punto geometrico” e, soprattutto, come “simulacro” e “ombra del morto”, idea a noi essenziale per l’analisi di *The Andy Warhol Diaries*.

Auerbach, infatti, terminato l’*excursus* iniziale, si concentra sul significato attribuito al termine dai Padri della Chiesa, in cui “figura” inizia ad acquisire il senso di “prefigurazione del futuro”. “Figura” è qui “qualche cosa di reale, di storico, che rappresenta e annuncia qualche altra cosa, anch’essa reale e storica”²⁵⁴. Diversa dall’allegoria, la “figura” è in

²⁵³ E. Auerbach, *Studi su Dante*, Feltrinelli, Milano 2017.

²⁵⁴ *Ibid.*

tutto e per tutto un elemento appartenente a una realtà storica, un dato concreto che prefigura un'ulteriore realtà storica, che la annuncia, la anticipa: a dialogare sono quindi il concetto di "figura" e il concetto di "adempimento", entrambi rimandanti a un futuro ancora in là da venire, un compimento definitivo. Questo tipo d'interpretazione ("interpretazione figurale") tendeva a guardare all'Antico Testamento come a un contenitore di figure o profezie reali della redenzione che sarebbe avvenuta nel Nuovo Testamento (così, per esempio, in Sant'Agostino).

L'interpretazione figurale stabilisce fra due fatti o persone un nesso in cui ognuno di essi non significa soltanto se stesso, ma significa anche l'altro, mentre l'altro comprende o adempie il primo. I due poli della figura sono separati nel tempo, ma si trovano entrambi nel tempo, come fatti o figure reali; essi sono contenuti entrambi, come si è già sottolineato più volte, nella corrente che è la vita storica...²⁵⁵.

L'interpretazione figurale ebbe larga diffusione anche nel Medioevo: questa, addirittura ne divenne la base generale della Storia, finendo con l'intervenire nell'interpretazione della stessa realtà quotidiana²⁵⁶. Auerbach si concentra infine sull'analisi delle figure ne *La divina Commedia*. Per Dante infatti, dice Auerbach, "il senso di ogni vita è interpretato" e, prendendo in considerazione Virgilio, afferma che nella *Commedia* "Virgilio è bensì il Virgilio storico, ma d'altra parte non lo è più, perché quello storico è soltanto "figura" della verità adempiuta che il poema rivela[...]" e questo perché "in Dante il personaggio è tanto più reale

²⁵⁵ *ibid.*

²⁵⁶ *ibid.*

quanto più integralmente interpretato”²⁵⁷. Sarebbe a dire che non esiste alcuna differenza, alcun aut-aut, tra il senso storico della persona e il suo senso recondito: per Dante, Virgilio era contemporaneamente poeta e guida, ruolo che ricopre nella Commedia perché fu storicamente guida di tutti gli altri poeti posteriori “infiammati dalla sua opera” e perché avendo descritto il regno dei morti, poteva guidare Dante nel suo percorso agli Inferi.

La serie lavora così sul desiderio di de-umanizzazione di Andy Warhol utilizzando il *deepfake* non come mero espediente tecnico e narrativo, ma come metafora. Il tentativo è stato infatti quello di ricercare parte dell’essenza di Warhol proprio nel suo rapporto con il macchinico, a partire dal suo stesso desiderio di liberarsi di quel corpo umano la cui bruttezza, viene ripetuto più volte durante la serie, era per lui fonte di profonda sofferenza. Tra il 1956 e il 1957 l’artista si sottopose addirittura a un intervento chirurgico estetico per la correzione del proprio naso, un intervento che diede pochi risultati e che “può essere interpretato come il primo tassello di quella costruzione artificiale della maschera dell’io artistico di Warhol”²⁵⁸ (nello stesso periodo l’artista inizierà a indossare delle parrucche bionde, successivamente argentate, e degli occhiali scuri che ne nascondessero gli occhi²⁵⁹).

Warhol diventa così un essere eternamente replicabile, serigrafato, e il *deepfake* vocale, le cui sfumature “robotiche” della voce sono evidenti sin dalle prime battute (in tal senso va segnalato come lo stesso Warhol si

²⁵⁷ *Ibid.*

²⁵⁸ A. Mecacci, *op.cit.*

²⁵⁹ *Ibid.*

sforzasse per donare un tono anodino alla sua voce durante le interviste) si fa mezzo per scavare tra le paure, i desideri e le speranze del padre della pop art. Il Warhol proposto in *The Andy Warhol Diaries* può quindi essere letto in senso gadameriano come un “accrescimento d’essere”. Il *deepfake* vocale dell’artista che legge i suoi stessi diari non è una semplice ripetizione della sua voce; è un evento che ha valore ontologico. La sua funzione non è quella di sopprimersi per mediare fedelmente un originale. Il *deepfake* vive in comunione con il Warhol reale, presentandolo sotto vesti inedite, capaci di cogliere una parte fondamentale della sua natura e invitarci a una rilettura della sua figura. In *The Andy Warhol Diaries* avviene una vera e propria trasformazione di Andy Warhol, il compimento grazie all’intelligenza artificiale, quindi attraverso possibilità fino a oggi non ancora sperimentate, di una traduzione robotica da lui ricercata per tutta la vita, se non della sua immagine, quantomeno della sua voce.

Warhol sembra così vivere ancora, e più pienamente, nella sua essenza storica, all’interno di una presenza disincarnata, robotica e priva di corpo, superficiale. È lui stesso a invitarci a questa lettura della sua figura in un’intervista del 1966:

Non leggo molto di quel che si scrive su di me, guardo solo le figure negli articoli, non importa che cosa dicono di me; leggo soltanto i segni delle parole. Guardo tutto in questo modo, la superficie delle cose, una specie di breille mentale, lascio scorrere solo le mie mani sulla superficie delle cose. [...] Se vuoi sapere tutto su Andy Warhol guarda soltanto la superficie: dei miei quadri, dei miei film e di me, io sono lì. Dietro non c’è niente²⁶⁰.

²⁶⁰ *Ibid.*

Riferimenti bibliografici

Agarwal S., El-Gaaly T., Farid H., Lim S., *Detectin Deep-Fake Videos From Phoneme-Viseme Mismatches*, in “*IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*”, pp. 660-661 2020.

Ajaka N., Samuels E., Kessler G., *Seeing isn't believing: The Fact Checker's guide to manipulated video*, <https://www.washingtonpost.com/news/fact-checker/wp/2016/11/22/the-fact-checkers-guide-for-detecting-fake-news/> (ultima consultazione il 14 febbraio 2023).

Albahar M., Almalki J., *Deepfakes: Threats and Countermeasures Systematic Review*, in “*Journal of Theroretical and Applied Information Technology*”, n.97, vol. 22, pp. 3242-3250 2020.

Arcagni S., *L'occhio della macchina*, Einaudi, Torino 2018.

Aristotele, *Poetica*, Giulio Einaudi Editore, Torino 2008.

Auerbach E., *Mimesis. Il realismo nella letteratura occidentale*, Einaudi, Torino 2000.

Auerbach E., *Studi su Dante*, Feltrinelli, Milano 2017.

Barale A., *Arte e intelligenza artificiale: alcune domande*, in A. Barale (a cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020.

Barthes R., *La grana della voce. Interviste 1962-1980*, Einaudi, Torino 1986.

Barcellona G., *Impressions, l'app per creare deep fake credibili. Un pericoloso salto di qualità per i falsi*, 2020, https://www.repubblica.it/tecnologia/social-network/2020/04/26/news/impressions_app_deep_fake-254939338/?refresh_ce (ultima consultazione il 14 febbraio 2023).

Benjamin W., *Sulla facoltà mimetica*, in *Angelus Novus*, Einaudi, Torino 1962.

Benjamin W., *L'opera d'arte nell'epoca della sua riproducibilità tecnica*, Giulio Einaudi Editore, Torino 2011.

Bakker P., Taalas S., "Il buisness del'adult content: una rilettura del concetto di innovazione", in E. Biasin, G. Maina, F. Zecca (a cura di), *Il porno espanso: dal cinema ai nuovi media*, Milano-Udine, Mimesis Edizioni.

Bertetto P., *Lo specchio e il simulacro. Il cinema nel mondo diventato favola*, RCS Libri, Milano 2007.

Bettetini M., *Contro le immagini. Le radici dell'iconoclastia*, Edizioni Laterza, Roma-Bari 2006.

Boatto A., *Pop Art*, Laterza, Bari 1983

Bode L., *Grave Robbing or Career Comeback? On the digital resurrection of dead stars*, in Kallioniemi K., Kärki K., Mäkelä J., Salmi H. (a cura di), *History of Stardom Reconsidered*, 2007.

Bode L., Lees D., Golding D., *Editorial the digital face and deepfakes on screen*, in "Convergence: The International Journal of Research into New Media Technologies", n. 27 (4).

Boehm G., *Il ritorno delle immagini*, in: A. Pinotti, A. Somaini, (a cura di), *Teorie dell'immagine. Il dibattito contemporaneo*, Raffaello Cortina Editore, 2008.

Burkell J., Gosse C., *Nothing New Here: Emphasizing the social and cultural context of deepfakes*, in "First Monday", n. 24, vol. 12, 2019.

Bridle J., *Nuova era oscura*, Nero, Roma 2019.

Brown W., D.H. Fleming, *Celebrity Headjobs: or ooziingsquid sex with a framed-up leaky {Scar-Jø}*, in *Porn Studies*, Vol. 7, n. 4, 357-366.

F. Casetti, *La galassia Lumière: sette parole chiave per il cinema che viene*, Bompiani, Milano 2015.

Carriero C., *Il consumo della Pop Art*, Jaca Book, Milano 2003

Cavarero A., *A più voci: per una filosofia dell'espressione vocale*, Feltrinelli, Milano 2003.

Codeluppi V., *Il divismo. Cinema, televisione, web*, Edimill, Bologna 2016.

Cole S., *For the Love of God, Not Everything Is a Deepfake*, 2018, <https://www.vice.com/en/article/7kzgg9/joe-biden-tongue-gif-twitter-deepfake> (ultima consultazione il 14 febbraio 2023).

Cole S., *This Horrifying App Undresses a Photo of Any Woman With a Single Click*, 2019, <https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman> (ultima consultazione il 14 febbraio 2023).

D'Elia D., *Tutto su ChatGPT: che cos'è, come si usa e cosa permette di fare*, 2023, <https://www.wired.it/article/chatgpt-guida-utilizzo/> (ultima consultazione il 14 febbraio 2023).

Dale R., *GPT-3: What's it good for?* in "Natural Language Engineering", n. 27, vol. 1, pp.113-118 2021.

Dasilva J.P., K.M. Ayerdi, T. Mendigueren Galdospin, *Deepfakes on Twitter: Which Actor Control Their Spread*, Disinformation and Democracy: Media Strategies and Audience Attitudes, Vol. 9, n.1.

Diresta R., *AI-Generated Text Is The Scariest Deepfake of All*, 2020, <https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/> (ultima consultazione il 14 febbraio 2023).

De Ruyter A., *The Distinct Wrong of Deepfakes*, in "Philosophy and Technology", 2021, n. 34, 1311-1332.

Dell'Oso L., Dalle Luche R., *L'altra Marilyn. Psichiatria e psicoanalisi di un cold case*, Le Lettere, Firenze 2016.

Dickson B., *The Guardian's GPT-3-written article misleads readers about AI. Here's why*, 2020, <https://bdtechtalks.com/2020/09/14/guardian-gpt-3-article-ai-fake-news/> (ultima consultazione il 14 febbraio 2023).

Dyer G.W., *Le GAN e la mimesi*, in A. Barale (a cura di), *Arte e intelligenza artificiale: be my GAN*, Editoriale Jaca Book, Milano 2020.

Dunham L., *What Marilyn Monroe Means to Me*, 2022, <https://www.vogue.com/article/what-marilyn-monroe-means-to-me-lena-dunham> (ultima consultazione il 14 febbraio 2023).

Eugeni R., *Capitale algoritmico: cinque dispositivi postmediali (più uno)*, Scholè, Brescia 2021.

Fagni T., Falchi F., Gambini M., Martella A., M. Tesconi, *TweepFake: About detecting deepfake tweets*, PLoS ONE, n. 16, vol. 5 2021.

Fletcher J., *Deepfakes, Artificial Intelligence, and some Kind of Dystopia: The New Faces of Online Post-Fact Performance*, in "Theater Journal", vol. 70, n.4 2018.

Floridi L., *Artificial Intelligence, Deepfake and a Future of Ectypes*, in "Philosophy & Technology", n. 31, 2018.

Floridi L., Chiriatti M., *GPT-3: It's Nature, Scope, Limits, and Consequences*. In "Minds and Machines", n. 30, pp. 681-694 2021.

Gadamer H., *Verità e metodo*, Rizzoli/Bompiani, Milano 2000.

Gadamer H., *Studi Platonici*, Marietti Editore, Torino 1998.

Goodfellow I., Pouget-Abadie J., et al., *Generative Adversarial Nets*, arXiv, 2014, p.1.

Frum D., *The Very Real Threat of Trump's Deepfake*, 2020, <https://www.theatlantic.com/ideas/archive/2020/04/trumps-first-deepfake/610750/> (ultima consultazione il 14 febbraio 2023).

Gadamer H., *Studi Platonici*, Marietti Editore, Torino 1998.

Giloldi R., *La mimesis e le sue peripezie*, in *La Poetica e le sue interpretazioni. Aristotele tra filosofia, letteratura e arti*, Pendragon, Bologna 2020.

Grothaus M., *Trust No One: Inside the World of Deepfakes*, Hodder & Stoughton, Londra 2021.

GPT-3, *A Robot wrote this entire article. Are you scared yet, human?*, 2020, <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> (ultima consultazione il 14 febbraio 2023).

S. Halliwell, *L'estetica della mimesis. Testi antichi e problemi moderni*, Aesthetica Edizioni, Palermo 2009.

Hao K., *A deepfake bot is being used to "undress" underage girls*, 2020, <https://www.technologyreview.com/2020/10/20/1010789/ai-deepfake-bot-undresses-women-and-underage-girls/> (ultima consultazione il 14 febbraio 2023).

Harris T.B., *The Building of Popular Images: Grace Kelly and Marilyn Monroe*, in "Studies in Public Communication", n. 1. 1957, pp. 45-58.

Hern A., *'Deepfake' face-swap porn videos banned by Pornhub and Twitter*, The Guardian, 2018, <https://www.theguardian.com/technology/2018/feb/07/twitter-pornhub-ban-deepfake-ai-face-swap-porn-videos-celebrities-gfycat-reddit> (ultima consultazione il 14 febbraio 2023).

Hern A., *New AI fake text generator may be too dangerous to release, say creators*, 2019, <https://www.theguardian.com/technology/2019/feb/14/elon-musk->

backed-ai-writes-convincing-news-fiction (ultima consultazione il 14 febbraio 2023).

Kätsyri J. et al., *A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness*, in "Frontiers in psychology 6.1", 2015.

Karandikar A., Deshpande V., *Deepfake Video Detection Using Convolutional Neural Network*, in "International Journal of Advanced Trends in Computer Science and Engineering", Vol. 9, n.2, 2020.

Khanjani Z., Watson G., Janeja V.P., *How Deep Are the Fakes? Focusing on Audio Deepfake: A Survey*, arXiv, 2021.

Kelleher J.D., *Deep Learning*. The MIT Press, Cambridge, Londra 2019.

Klossowski P., *Simulacra. Il processo imitativo nell'arte*, Mimesis, Milano 2002.

Korngiebel D.M., Mooney S.D., *Considering the possibilities and Pitfalls of Generative Pretrained Transformer 3 (GPT-3) in healthcare delivery*, in "npj. Digital Medicine" 4 (93) 2021.

Korshunov P., Marcel S., *DeepFakes: a New Threat to Face Recognition? Assessment and Detection*, arXiv:1812.08685v1 2018.

Kushner D., "A Brief History of Porn on the Internet", 2020, <https://www.wired.com/story/brief-history-porn-internet/> (ultima consultazione il 14 febbraio 2023).

Levi Strauss, *Perché crediamo alle immagini fotografiche?*, Johan & Levi Editore, Milano 2021.

Lucy L., D. Bamman. *Gender and Representation Bias in GPT-3 Generated Stories*, in "Proceedings of the 3rd Workshop on Narrative Understanding, pp. 48-55 2021.

Manovich L., *Software Takes Command: extending the language of new media*, Bloomsbury USA Academic, New York 2013.

Moruzzi C., *Alla ricerca della creatività: le GAN come paradigma dell'autonomia nel software per la composizione musicale*, in A. Barale (A cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020, p.156.

Öhman C, *Introducing the pervert's dilemma: a contribution to the critique of Deepfake pornography*, in "Ethics and Information Technology", n. 22 2020.

Lawson E., "Pixels, porn and private selves: intimacy and authenticity in the celebrity nude photo hack", in *Celebrity Studies*, Vol. 6, n. 4, pp. 607-609.

Lees D., Keppel-Palmer M., Bashford-Rogers T., *The digital resurrection of Margaret Thatcher: Creative, technological and legal dilemmas in the use of*

deepfakes in screen drama, in "Convergence: The International Journal of research into New media Technologies", n.0(0), 2021, pp. 954- 973.

Maddison S., "Le mitologie pornografiche e i limiti del piacere. Max Hardcore e il porno estremo", in E. Biasin, G. Maina, F. Zecca, (a cura di), *Il porno espanso: dal cinema ai nuovi media*, Mimesis, Milano-Udine.

Maddocks S., "A Deepfake Porn Plot Intended to Silence Me": Exploring continuities between pornographic and political deep fakes", in "Porn Studies", Vol. 7, pp. 415-423.

Maras M., Alexandrou A., *Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos*, in "The International Journal of Evidence & Proof", Vol. 23 (3), p. 255 2018.

Martinez V.C., Castillo G.P., *Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso*, in "Historia y comunicación social", n. 24(2).

Matern F., Riess C., Stamminger M., *Exploiting Visual Artificats to Expose Deepfakes and Face Manipulations*, in "IEEE Winter Applications of Computer Vision Workshops", pp. 83-92 2019.

McCarthy J., Minsky M.L., Rochester N., Shannon C.E., *A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955*", in "AI magazine", n. 27 (4), p.12 2006.

McLhuan M., *Gli strumenti del comunicare*, il Saggiatore, Milano 2008.

Mecacci A., *Introduzione a Andy Warhol*, Laterza, Bari 2014 (edizione digitale).

Metz R., *These people do not exist. Why websites are churning out fake images of people (and cats)*, 2019, <https://edition.cnn.com/2019/02/28/tech/ai-fake-faces/index.html> (ultima consultazione il 14 febbraio 2023).

Mihailova M., "To Dally with Dalì: Deepfake (Inter)faces in the Art Museum", in *Convergence: The International Journal of Research into New Media Technologies*, Vol. 0(0), 2021, pp. 1-17.

Mitchell M., *Artificial Intelligence: A Guide for Thinking Humans*, Penguin Random House, New York 2019.

Mitchell W.J.T., , *I quattro concetti fondamentali della scienza dell'immagine*, in: W.J.T. Mitchell; M. Cometa, V Cammarata (a cura di), *Pictorial Turn. Saggi di cultura visuale*, Raffaello Cortina Editore Milano, 2017.

Mori M., MacDorman K.F., Kageki N., *The Uncanny Valley* [from the field], in "IEEE Robotics & Automation Magazine", 19.2, 2012, pp. 98-100.

Mitchell W.J.T.,, *Cloning Terror: The War of Images, 9/11 to the Present*, University of Chicago Press, 2011.

Naruniec J., Helminger L., Schroers C., Weber R.M., *High-Resolution Neural Face Swapping for Visual Effects*, in “Eurographic Symposium on Rendering 2020”, n. 34, vol. 39 2020.

Newton O., Stanfill M., “My NSFW video has partial occlusion: deepfakes and the technological production of non-consensual pornography”, in *Porn Studies*, Vol. 7 (4).

Otty L., Sellars P., *The No Man Show: Technology and the Extension of Presence in the Work of Andy Warhol*, 2005.

Poma L., *Il sex appeal dei corpi digitali. Seduzione, amori, tradimenti, malattie e immortalità dei nostri digital body*, Milano, Franco Angeli 2016.

Paris B., Donovan J., *Deepfakes and cheap fakes: the manipulation of audio and visual evidence*, Data & Society 2019.

Perlow J., ‘*The Beatles: Get Back shows that deepfake tech isn’t always evil*,’ 2021, <https://www.rollingstone.it/musica/peter-jackson-parla-di-get-back-non-sono-i-beatles-che-si-stanno-sciogliendo/597260/#Part1> (ultima consultazione il 14 febbraio 2023).

Pinotti A., *Alla soglia dell’immagine. Da Narciso alla realtà virtuale*, Einaudi, Torino 2021.

Pinotti A., Somaini A., *Cultura visuale*, Einaudi, Torino 2016.

Platone, *La Repubblica*, Mondadori, Milano 2016.

Preciado B., *Pornotopia. Playboy: architettura e sessualità*, Fandango, Roma 2011.

Ridler A., *Set di dati e decadenza: Fall of the House of Usher*, in A. Barale (a cura di), *Arte e intelligenza artificiale: Be my GAN*, Editoriale Jaca Book, Milano 2020.

Ridenour A., *The Automated Andy Warhol is reprogrammed*, Los Angeles Times, 2002, <https://www.latimes.com/archives/la-xpm-2002-may-16-wk-town16-story.html> (ultima consultazione il 14 febbraio 2023).

Sallis J., *On the Manifold Senses of mimesis. From Plato to Gadamer and Beyond*, in "The Blackwell Companion to Hermeneutics", 2016.

Schick N., *Deep Fakes and the Infocalypse: What You Urgently Need to Know*, Octopus Publishing Group, Londra 2020.

Scott K., *Microsoft Teams up with Open AI to exclusively license GPT-3 language model*, in "Official Microsoft Blog" 2020, <https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/> (ultima consultazione il 14 febbraio 2023).

Simonite T., *Deepfakes are becoming the Hot new Corporate Training Tool*, <https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/> (ultima consultazione il 14 febbraio 2023).

Smith A., Mansted K., *Weaponised deep fakes*, 2020, <https://www.aspi.org.au/report/weaponised-deep-fakes> (ultima consultazione il 14 febbraio 2023).

Smith H., Mansted K., *Weaponised Deep Fakes: National Security and Democracy*, 2020, <https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-04/Weaponised%20deep%20fakes.pdf?lgwT9eN66cRbWTovhN74WI2z4zO4zJ5H> (ultima consultazione il 14 febbraio 2023).

Summers A., *Marilyn. Tutti i segreti di una vita*, Bompiani, Milano 1985.

Uva C., *Ultracorpi. L'attore cinematografico nell'epoca della digital performance*, Bulzoni Editore, Roma 2011.

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

Vegetti M., *Introduzione*, in Platone, *La Repubblica*, Mondadori, Milano 2016.

Vincent J., *This app uses neural networks to put a smile on anybody's face*, 2017, <https://www.theverge.com/tldr/2017/1/27/14412814/faceapp-neural-networks-ai-smile-image-manipulation> (ultima consultazione il 14 febbraio 2023).

Vincent J., *OpenAI has published the text-generating AI it said it was too dangerous to share*, 2019, <https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters> (ultima consultazione il 14 febbraio 2023).

Virilio P., *La bomba informatica*, Raffaello Cortini Editore, Milano 2000.

Wagner T., A. Blewer, *"The word real is no longer real": Deepfakes, Gender, and the Challenges of AI-Altered Video*, in *"Open Information Science"*, n. 3, 2019, pp. 32-39.

Watercutter A., *Una nuova serie di Netflix ha ricreato la voce di Andy Warhol grazie all'intelligenza artificiale*, 2022, <https://www.wired.it/article/diari-andy-warhol-serie-netflix-intelligenza-artificiale/> (ultima consultazione il 14 febbraio 2023).

Westerlund M., *The Emergence of Deepfake Technology: A Review*, in *"Technology Innovation Management Review"*, n.9, vol. 11, 2019, pp. 39-52.

Wiggers K., *Open AI's Massive GPT-3 is impressive, but size isn't everything*, VentureBeat, 2020, <https://venturebeat.com/2020/06/01/ai-machine-learning-openai-gpt-3-size-isnt-everything/> (ultima consultazione il 14 febbraio 2023).

Woolley S., *The Reality Game: How the next wave of technology will break the truth and what we can do about it*, Octopus Publishing Group, Londra 2020.

Yang C., Ding L., *Defending against GAN-based Deepfake Attacks via Transformation-aware Adversarial Faces*, arXiv, 2020.

Zhang M., Li J., *A Commentary of GPT – 3 in MIT Technology Review 2021*, in "Fundamental Research", n. 1, 2021, pp. 831 -833.

Zhok A., *Rappresentazione e realtà. Psicologia fenomenologica dell'immaginario e degli atti rappresentativi*, Mimesis, Milano 2014.

Ringraziamenti

Questa tesi non sarebbe stata possibile senza la cura del professor Canova, la pazienza del professor Montani, e il conforto di Giulio Sangiorgio.