



Università
Ca' Foscari
Venezia

Corso di Laurea
Magistrale
in Scienze
Filosofiche

Tesi di Laurea

Arte e intelligenza artificiale nell'era dei deepfake

Relatore

Ch. Prof. Pietro Conte

Laureanda

Claudia Spinato

Matricola 851923

Anno Accademico

2020 / 2021

INDICE

<i>Introduzione</i>	<i>3</i>
<i>Capitolo 1. Iperrealismo e fenomenologia</i>	<i>6</i>
1.1. Il problema dell'iperrealismo.....	6
1.1.1. Approccio fenomenologico alle teorie dell'immagine	9
1.1.2. Il perturbante	14
1.2. I new media	21
1.2.1 Fenomenologia dei new media	21
1.2.2. L'epoca della simulazione.....	25
1.2.3. New media e fake news	32
<i>Capitolo 2. Deepfake: risvolti negativi nell'uso dell'intelligenza artificiale.</i>	<i>38</i>
2.1. Introduzione all'argomento: storia del deepfake	38
2.2. Gans (Generative Adversarial Networks).....	44
2.2.1. Un esempio concreto dell'uso delle GAN: thispersondoesnotexist	49
2.3. Problemi e criticità	54
2.3.1. Applicazioni a livello politico e fake news	56
2.3.2. Fake porn	59
2.4. Possibili soluzioni	61
<i>Capitolo 3. Usi alternativi (e positivi) dell'intelligenza artificiale: la via dell'arte.....</i>	<i>68</i>
3.1. La Gan Art e le Creative Adversarial Networks.....	68
3.1.1. Introduzione ad un diverso uso delle GANs	68
3.1.2. Le Creative Adversarial Networks: dall'emulazione alla creatività	71
3.1.3. I diversi campi artistici in cui sono state impiegate le Creative Adversarial Networks	77

3.2. <i>Gan Artists</i> : gli artisti e l'uso delle Generative Adversarial Networks.....	85
3.2.1. <i>Obvious</i>	85
3.2.2. Mario Klingemann	90
3.2.3. Anna Ridler	96
3.2.4. Robbie Barrat	102
3.2.5. Mauro Martino	107
3.2.6. Chris Peters.....	111
3.2.7. Tamiko Thiel.....	113
3.2.8. Memo Akten	116
3.3. Creatività e <i>machine learning</i> : un dibattito ancora aperto.....	120
<i>Conclusioni</i>	126
<i>Appendici: Intervista a Mauro Martino</i>	127
<i>Bibliografia</i>	139
<i>Elenco delle illustrazioni</i>	148
<i>Ringraziamenti</i>	150

Introduzione

Con lo sviluppo dell'intelligenza artificiale e dei nuovi media diventa sempre più difficile al giorno d'oggi orientarsi nel mondo digitale e della disinformazione. Il confine tra realtà e illusione è labile: i deepfake nell'ultimo decennio sono un pericolo sempre maggiore e spesso difficile da riconoscere. L'elaborato, basandosi sul concetto di iperrealismo secondo alcuni principali filosofi e affrontando il problema filosofico della distinzione tra immagine e realtà, ne segue lo sviluppo attraverso la nascita dei deepfake e dell'intelligenza artificiale (nello specifico delle GANs), individuando, tuttavia, un campo in cui questo sviluppo può essere innovativo e avere un risvolto positivo: quello artistico. Se l'uso iperrealistico dell'intelligenza artificiale può avere un risvolto negativo, con lo scopo di ingannare lo spettatore, l'uso della stessa tecnologia in maniera creativa potrebbe invece aprire una nuova strada?

Nel primo capitolo verranno dati i fondamenti del discorso da un punto di vista prettamente filosofico: si tratterà, infatti, la questione dell'iperrealismo, dell'eccesso di somiglianza tra immagine e modello e dunque della difficoltà a distinguere tra l'uno e l'altro. Seguendo questa linea, verrà prima esposta la concezione fenomenologica: l'approccio percettivo all'immagine secondo il filosofo Husserl (in questo modo verrà chiarita la difficoltà della distinzione tra immagine e modello a cui si riferisce, citando l'esemplare scena della dama husserliana). Si passerà, dunque, alla definizione di perturbante, per esporre la sensazione che si genera quando l'illusione cede il posto alla capacità di percepire adeguatamente ciò che ci si presenta di fronte, quando l'imitazione diventa così verosimile, così iperrealistica, da non lasciare spazio ad una adeguata comprensione del fenomeno. Il perturbante è proprio quella sensazione che si genera con l'eccessiva familiarità del modello, la coscienza di trovarsi davanti a qualcosa di reale, seguita da una comprensione di essere davanti ad un artefatto. Il discorso dell'eccessiva somiglianza e dell'iperrealismo verrà dunque trasposto ai nostri giorni, all'epoca attuale: seguendo il discorso del filosofo Jean Baudrillard si vedrà fino a che punto l'iperrealismo abbia raggiunto il suo massimo sviluppo, con l'avvento dei nuovi media immersivi. Ci troviamo, appunto, nell'epoca della simulazione: epoca in cui diventa estremamente difficile, spesso, distinguere ciò che è reale

da ciò che è virtuale. L'esempio concreto di questo fenomeno sono le fake news, fenomeno in cui lo spettatore non è in grado di orientarsi adeguatamente e di distinguere fino in fondo la verità dall'illusione.

Nel secondo capitolo si passerà ad un nuovo sviluppo dell'iperrealismo: non più quello basato sulla difficoltà a distinguere un'opera dalla realtà (una donna reale da una dama di cera, ad esempio, come nel caso di Husserl citato nel primo capitolo), bensì sulla distinzione tra immagini vere e immagini false. Con lo sviluppo dell'intelligenza artificiale, dei nuovi media e delle tecnologie a nostra disposizione diventa sempre più facile ingannare lo spettatore e falsificare la realtà. L'esempio concreto sono i deepfake: video creati usando dei sofisticati algoritmi chiamati GAN, tramite cui è possibile far circolare ancora più facilmente fake news e fake data. In questo capitolo verrà quindi trattato uno sviluppo dell'iperrealismo adattato ai nostri giorni e una sorta di "risvolto negativo" dell'uso dell'I.A. e delle nuove tecnologie: quello volto ad ingannare lo spettatore. Attraverso l'uso di articoli scientifici e di attualità verranno qui definiti: il significato di deepfake, i principali esempi in campo politico e di attualità, il significato di GAN e alcune possibili soluzioni per "difendersi" dalla circolazione e diffusione di video falsi.

Nel terzo capitolo, riprendendo il discorso dell'intelligenza artificiale, dopo averne esposto l'uso negativo, verrà invece mostrata una possibile "via positiva": un impiego non iperrealistico e non volto ad ingannare, bensì ad esprimersi in maniera innovativa. Questa possibile via è rappresentata dall'arte. Le GAN verranno intese sotto il concetto di CAN: gli stessi algoritmi esposti nel secondo capitolo, impiegati però in maniera creativa. Dall'uso iperrealistico, emulativo ed imitativo si passerà, infatti, all'uso creativo. Ci si chiederà in seguito se le macchine possano davvero considerarsi dotate di creatività, affrontando il dibattito contemporaneo, e verranno elencati alcuni artisti attuali che fanno uso di queste tecnologie in maniera espressiva, artistica, mostrando quindi le loro opere più esemplificative.

La metodologia seguita nell'elaborazione di questo lavoro è stata l'analisi di fonti scritte (libri e monografie, articoli filosofici, scientifici e di attualità presenti online, sitografia) e visuali (immagini e video online). Sono stati inoltre consultati alcuni esperti in

materia di informatica e arte (segue, a fine tesi, un'intervista ad un esperto in materia di I.A., algoritmi e applicazione di essi in campo artistico e scientifico).

La scelta dell'argomento è stata orientata principalmente dagli interessi personali e dall'esperienza pratica: ho cercato di unire quelli che sono stati fin dal principio del percorso universitario i miei interessi (la filosofia, l'intelligenza artificiale e le applicazioni di quest'ultima nell'epoca attuale, in ambiti concreti) e l'esperienza pratica del tirocinio effettuato in una galleria d'arte (seguendo, anche in questo caso, un interesse personale).

L'elaborato aspira inoltre ad unire l'analisi filosofica e il pensiero critico ad uno sguardo attuale su temi contemporanei, cercando dunque di non mantenere il pensiero filosofico su un piano "astratto", impersonale, inattuale, non applicabile dunque alla realtà concreta, bensì a farlo diventare parte integrante e in qualche modo utile per orientarci nell'epoca in cui viviamo.

Capitolo 1. Iperrealismo e fenomenologia

1.1. Il problema dell'iperrealismo

Vagando in un museo delle cere ci imbattiamo lungo una scalinata in una bella sconosciuta che ammicca verso di noi – uno scherzo ben noto. Si tratta di un manichino che per un istante ci ha tratti in inganno. Per tutto il tempo in cui viviamo in questa illusione abbiamo una percezione non diversa da qualsiasi altra. Vediamo una donna, non un manichino. Non appena ci accorgiamo dell'inganno, accade l'inverso: ora vediamo un manichino che *rappresenta* una donna.¹

Percezione: sono le percezioni a costruire il mondo in cui viviamo. Viviamo in un mondo oggettivo, viviamo tutti nella stessa realtà. Siamo soggetti differenti, ma il mondo in cui viviamo è lo stesso. Eppure, le percezioni che abbiamo di questo mondo sono spesso differenti per ognuno di noi. Cosa cambia da un soggetto all'altro? Perché lo stesso identico stimolo può essere recepito in maniera così differente da una persona all'altra? Perché qualcuno riesce a cogliere stimoli diversi, ad interpretare un'identica situazione in maniera completamente differente? Grazie alle percezioni. Noi vediamo lo stesso oggetto, ma ciò grazie a cui riusciamo ad interpretare quell'oggetto è la nostra percezione di esso. Oggetto, stimolo, percezione, interpretazione, risposta. Grazie a tutto questo, noi abbiamo cognizione della realtà. Grazie a tutto questo, ogni soggetto può costruire la propria realtà, diversa da un altro. Ma cosa può accadere quando le nostre percezioni non funzionano? Cosa può accadere quando la nostra percezione dell'oggetto ci porta su un terreno completamente sbagliato, conducendoci fuori strada? In questo caso, l'uomo è vittima di un *inganno*. È vero, d'altra parte, che non siamo esseri infallibili: siamo soggetti all'*illusione*. Ed è proprio questo che si verifica quando la realtà che ci si presenta davanti agli occhi, in apparenza così indubitabile, in superficie così chiara e distinta, non è ciò che sembra.

Non è ciò che sembra: è ciò che accade ad Husserl, nel frammento sopra citato, davanti alla dama di cera. Egli crede che sia reale, crede di trovarsi davanti ad una persona in carne ed ossa, eppure non si tratta che di un manichino. È convinto che ciò che vede sia reale,

¹ E. Husserl, *Ricerche logiche* (1900-1901), tr. it. e cura di G. Piana, 2 voll., il Saggiatore, Milano 2001, vol. 2, pp. 229-230.

eppure è semplicemente vittima di un'illusione. L'autore, nell'esempio sopra citato, realizza subito quello che sta accadendo: bastano pochi minuti, infatti, per comprendere che ciò che vede non è che un manichino in cera, che non si tratta di una dama vera e propria. Questo lo porterà ad una lunga riflessione, di cui parleremo in seguito. Ma è importante soffermarsi poco prima di questa riflessione, al momento vero e proprio dell'illusione in atto. Cosa accadrebbe se non fosse possibile distinguere tra ciò che è vero e ciò che non lo è? Se ci trovassimo davanti a qualcosa di così *iperreale* da non riuscire ad averne una corretta percezione? Non saremmo in grado di dire “quella dama che credevo reale è invece un manichino”, poiché resteremmo intrappolati nell'illusione.

Le nostre percezioni sono involontarie, ogni giorno in base a differenti stimoli che riceviamo abbiamo differenti percezioni. Siamo necessitati ad interpretare la realtà, a darne una nostra definizione, a darvi un senso in base a ciò che siamo. Se non lo facessimo, non potremmo nemmeno compiere le scelte di tutti i giorni. Se non potessimo spiegare e dare un senso a ciò che ci accade e che ci circonda (più o meno razionalmente, in modo più o meno complesso a seconda di ognuno) non potremmo nemmeno scegliere ciò che è bene e ciò che è male per noi, non potremmo più agire in alcun modo sulla realtà. Cadremmo in uno stato di aporia totalizzante. Per questo è importante interpretare, per questo è importante dare continuamente la nostra visione della realtà. E proprio per questo diventa paralizzante l'idea di essere ingannati, l'idea di essere vittime di una qualche inconsapevole illusione.

Nello specifico di questo elaborato, ci troveremo ad affrontare il problema dell'iperrealismo: in particolare ci chiederemo cosa accada quando un'immagine non assomigli più al suo modello, ma a livello percettivo si identifichi con esso. Cosa accada, quindi, quando la somiglianza lasci spazio all'identità.

Affrontando questo argomento ci troviamo “sullo scivoloso terreno della somiglianza, o più precisamente (e più pericolosamente) dell'eccesso di somiglianza”². Il tema è, infatti, “l'iperrealismo [...] nella cornice teorica e metodologica dell'estetica ma chiamandola a dialogare – ferme restando le inaggirabili identità disciplinari – con la storia dell'arte, l'antropologia delle immagini e quella variegata costellazione di studi che va sotto

² P. Conte, *In carne e cera. Estetica e fenomenologia dell'iperrealismo*, Quodlibet, Macerata, 2015, p.9.

il nome di “cultura visuale”³. La questione può essere analizzata in varie forme: nel caso della ceroplastica possiamo parlare di raffigurazioni indistinguibili dai soggetti raffigurati. Secondo Pietro Conte:

Annullando quella distanza rispetto al fruitore che è insita nel concetto stesso di <<raffigurazione>> e che impedisce che si generi confusione tra sfera iconica e mondo della vita, l'immagine si nega in quanto tale e cede il posto al doppio e all'illusione: quando il realismo diventa iperrealismo, la somiglianza si fa identità⁴.

Ci si chiede, tuttavia, quali siano questi processi che portano a percepire un manichino come una persona in carne ed ossa. Questione fondamentale non solo dal punto di vista della nostra percezione (e conseguente illusione), ma anche dal punto di vista artistico: se parlare di arte è parlare per e con immagini, se le immagini sono strutture che rinviano chiaramente a qualcos'altro che non è immagine⁵, cosa accade quando questo rinvio non c'è (in quanto manca la distanza)? Se le figure iperrealistiche non sono immagine, si può ancora parlare di arte? La questione dell'iperrealismo diventa fondamentale per l'arte e per l'estetica.

Bisogna inoltre precisare che con l'avvento della cera abbiamo potuto notare una sorta di rivoluzione nella storia dell'arte (parlando appunto di verosomiglianza e iperrealismo), ma questo accadeva all'incirca un secolo fa. Ci troviamo però, ad oggi, in un'epoca in cui l'iperrealismo ha raggiunto una dimensione ancora più ampia, con l'avvento dei nuovi media digitali e del virtuale. Verrà approfondito, dunque, l'iperrealismo sotto questa luce, che possiamo definire una sorta di “*iperrealismo 2.0*”: quello sviluppatosi alla luce e attraverso l'uso dei new media e delle nuove tecnologie immersive, portandoci in qualche modo “un passo oltre la cera”.

Prima di analizzare più approfonditamente la questione, è però necessario capire innanzitutto quali siano questi processi che ci portano ad essere vittime di illusione e di inganno, a percepire ad esempio (come nella scena sopracitata) un manichino come una persona in carne ed ossa. Quali siano, dunque, i modi in cui percepiamo la realtà. Ci troviamo, ora, nel terreno della fenomenologia.

³ *Ibidem.*

⁴ *Ivi*, pp. 10-11.

⁵ *Ivi*, p. 11.

1.1.1. Approccio fenomenologico alle teorie dell'immagine

“L'immagine non è un'illusione”⁶, sentenza Husserl senza mezzi termini. E l'affermazione può essere capovolta: l'illusione non è un'immagine”⁷.

Precedentemente abbiamo citato l'esempio di Husserl e la dama di cera. Torniamo al momento esatto in cui l'autore non ha ancora realizzato il fatto di essersi imbattuto in un manichino, al momento in cui egli è convinto di trovarsi di fronte ad una donna in carne ed ossa. Dunque, all'istante poco prima della realizzazione. Siamo ancora immersi nell'illusione e finché essa persiste, quella che abbiamo è una comune percezione. Se l'illusione continua, anche la percezione continuerà. Se l'illusione cessa improvvisamente, nell'istante preciso in cui questa smette di esistere e l'autore realizza di trovarsi davanti ad un fantoccio, la percezione viene percepita come errata: egli credeva di trovarsi davanti ad una donna, che in realtà era un manichino.

“*L'immagine non è un'illusione*” afferma correttamente Husserl: quando ci troviamo davanti ad un'immagine abbiamo ben chiaro ciò che vediamo. Una qualsiasi immagine è, infatti, una struttura di rinvio, qualcosa che rimanda ad altro. *Rappresenta* qualcos'altro, ma abbiamo ben chiaro questa distinzione nel momento in cui la vediamo. C'è una distanza tra lo spettatore e l'immagine (solitamente anche una cornice) ed è evidente la struttura di rinvio. Per questo non può esserci né illusione né inganno: sappiamo ciò che vediamo. In questo caso, invece, non è evidente il rinvio ad altro, perché il rappresentante e il rappresentato si confondono, si perdono l'uno nell'altro, in un'indistinzione tale che ci porta ad essere ingannati. Cosa accade nel passaggio dalla percezione della donna alla consapevolezza di trovarsi in presenza di un manichino? Quale differenza si instaura tra il primo atto intuitivo e il secondo? Affrontando questa analisi ci troviamo sul terreno della fenomenologia, ovvero dell'indagine sulle modalità tramite cui la coscienza intenzionale apprende gli oggetti.

⁶ E. Husserl, *Zur Lehre vom Bildbewusstsein und Fiktumbewusstsein* (1912), in *Husserliana*, vol. 23, cit., pp. 486-494, qui p. 486.

⁷ P. Conte, *In carne e cera. Estetica e fenomenologia dell'iperrealismo*, Quodlibet, Macerata, 2015, pp. 15-16.

“La sfera fenomenologica è la sfera di ciò che è dato veramente, di ciò che è rinvenibile in modo adeguato e delle sue parti effettive” afferma Husserl, aggiungendo che “per contro, il vissuto di fantasia, la cosiddetta rappresentazione di fantasia, è un dato fenomenologico”⁸.

Parlando dell’immagine fisica, Husserl distingue tra “cosa” e “immagine”. Afferma, infatti, che vi sia una duplicità nel concetto di immagine: abbiamo l’immagine in quanto “cosa fisica” (*Bildding*), trattasi ad esempio della tela di un dipinto, l’immagine in quanto “oggetto-immagine” (*Bildobjekt*), che si può manifestare in diversi modi (diverse conformazioni, diverse colorazioni). Vi è infine il “soggetto-immagine” (*Bildsujet*), che sarebbe “l’esatto analogo dell’immagine di fantasia, vale a dire l’oggetto che si manifesta e che è il rappresentante del *sujet-immagine*”⁹.

Husserl ci pone un esempio concreto per chiarire l’idea alla base di questo concetto: se prendessimo la fotografia di un bambino, sapremmo che la fotografia non è il bambino, bensì che *rappresenta* un bambino; sapremmo, dunque, che si tratta di un’immagine fotografica di un soggetto: “la fotografia in quanto cosa è un oggetto reale e come tale viene assunta nella percezione” (...) “dall’immagine fisica distinguiamo pertanto l’immagine che funge da rappresentante, l’oggetto che si manifesta, che ha funzione raffigurativa: attraverso di essa viene raffigurato il *sujet-immagine*”¹⁰.

D’altra parte, questi oggetti-immagine vanno a loro volta differenziati rispetto agli oggetti raffigurati: infatti il bambino della fotografia non mostra gli stessi colori del bambino reale (quello a cui l’oggetto fisico rimanda), bensì mostra colori fotografici. Come affermerà in seguito l’autore, tutto ciò non viene assunto tramite una riflessione a posteriori: è qualcosa che si sa fin dal principio.

Qui arriviamo al punto fondamentale della riflessione husserliana (che sarà fondamentale per la comprensione del discorso sull’iperrealismo): sono le differenze tra l’immagine fisica e la cosa che essa rappresenta a darci la *coscienza d’immagine*. Ma se queste differenze venissero meno, non potremmo più avere questa consapevolezza: “se l’immagine che si manifesta fosse, dal punto di vista fenomenico, assolutamente identica

⁸ E, Husserl, *Fantasia e coscienza d’immagine*, a cura di C. Rozzoni, Rubbettino Editore, Soveria Mannelli, 2017, pp. 4-5.

⁹ Ivi, p.23.

¹⁰ *Ibidem*.

all'oggetto inteso o, meglio, se la manifestazione d'immagine non si distinguesse per nulla dalla manifestazione percettiva dell'oggetto stesso, allora difficilmente si giungerebbe ad una coscienza del carattere d'immagine"¹¹. Per Husserl è necessario che ci sia una *coscienza della differenza*: l'oggetto vale solo in quanto rappresentante di un altro oggetto.

Torniamo ora, dopo aver esaminato nello specifico l'analisi fenomenologica e delle modalità apprensionali della coscienza, all'esempio iniziale della dama di cera: come possiamo farci ingannare in questo modo? Cos'è che permette all'oggetto di falsare a tal punto le nostre percezioni? Proprio questo: l'assenza della *coscienza della differenza*. L'oggetto fisico rappresentante si confonde con il soggetto rappresentato, diventa un tutt'uno. Non è più chiaro il confine tra le due dimensioni, non vi è più una chiara differenza. L'oggetto e il *sujet*-immagine si confondono inevitabilmente: siamo caduti nell'illusione, vittime d'inganno.

La *presentazione* è il carattere fondamentale della percezione: si tratta dell'"ostensione delle proprietà di un oggetto o di una molteplicità di oggetti che occupano lo stesso ambiente in cui si colloca l'osservatore e che si estendono nella stessa durata attuale in cui si susseguono le operazioni percettive"¹². Il potere dell'immagine sta proprio in questo: nel rendere presente un assente. L'immagine ha il potere di presentificare: non creando un'illusione, bensì mantenendo ben distinte la percezione attuale e la coscienza figurale inattuale. L'immagine è in questo senso un medium attraverso cui l'oggetto assente viene presentificato.

"Che si sia vittime di un'illusione non lo si capisce finché si è irretiti nell'illusione stessa"¹³: siamo nell'illusione della dama di cera. Non vi è cornice, non vi è distanza, non vi è nemmeno contrasto. Non possiamo avere una corretta coscienza d'immagine, non riusciamo ad averne una fruizione estetica: siamo irretiti nella dimensione dello scambio tra immagine e realtà¹⁴. Non ci sono differenze fenomeniche tra l'immagine e il suo referente,

¹¹ Ivi, p. 25.

¹² C. Calì, *Husserl e l'immagine*, in *Aesthetica Preprint Supplementa*, Centro Internazionale Studi di Estetica, 10 aprile 2002, p.35.

¹³ P. Conte, *In carne e cera. Estetica e fenomenologia dell'iperrealismo*, cit. p.30.

¹⁴ Husserl, a proposito dell'illusione, affermerà che: "Quando si cade vittima di un'illusione (intesa come illusione percettiva), questa si inquadra in un contesto percettivo posizionale: vedo la manifestazione in quanto realtà. Posso poi accorgermi dell'inganno e allora solo a quel punto sorge la coscienza dell'illusione: ciò che

ma quando mancano le differenze non vi è più alcuna immagine: abbiamo solo un'apparizione identica alla percezione dell'oggetto. L'immagine deve essere struttura di rinvio: bisogna essere in grado di cogliere il medesimo nella differenza. Quando, però, manca il contrasto tra oggetto iconico e soggetto iconico l'immagine cede il posto all'illusione: nel caso della dama di cera, non sappiamo dove finisce la cera e inizia ciò che intende riprodurre: lo spettatore vede nel manichino qualcosa di reale, la dama è una persona in carne ed ossa, qualcosa di reale. Husserl utilizza a questo proposito il termine "*Fiktum*"¹⁵ per indicare "ciò che è stato plasmato, modellato" e anche "invenzione-finzione-simulazione": l'immagine è tale solo se il Fiktum è esplicitamente un Fiktum.

Abbiamo compreso a questo punto in maniera più approfondita come si crea l'illusione, facciamo un passo oltre.

Ad un certo punto si insinua nel soggetto il dubbio: si svela l'illusione, qualcosa non torna: è vittima di un inganno. Cosa succede nell'esatto istante in cui se ne rende conto? Cos'è che fa svelare l'inganno, svegliando improvvisamente lo spettatore dal limbo dell'illusione, come fosse bruscamente risvegliato da una dimensione placida e sognante? Nonostante l'abilità nell'illudere c'è qualcosa che manca: la capacità di movimento. Qui inizia a vacillare la familiarità con l'oggetto, con una normale percezione, qui l'illusione percettiva deve vacillare. La statua di cera è figura e non persona, l'errore viene svelato. Si instaura un elemento di sorpresa e spiazzamento: l'esperienza passata e presente non coincidono, le attese future vengono disattese. Non vi è concordanza nel susseguirsi di queste esperienze. Il contrasto, tuttavia, può continuare: la coscienza d'immagine dopo aver scacciato quella percettiva ne subisce continuamente il ritorno e deve continuare la sua lotta per imporsi, per ricordarci di essere vittime di un'illusione. Ecco il ritorno, rieccoci intrappolati nell'inganno: "Non possiamo farci nulla: "sappiamo" che si tratta di parvenza, e tuttavia vediamo un essere umano"¹⁶.

prima consideravo realtà «ha» ora «il carattere di realtà soppressa» (Husserl, *Fantasia e immagine*, a cura di C. Rozzoni, Rubbettino Editore, Soveria Mannelli, 2017, p.49).

¹⁵ E. Husserl, *Fantasia e coscienza d'immagine*, a cura di C. Rozzoni, Rubbettino Editore, Soveria Mannelli, 2017, p. 54.

¹⁶ Ivi, p. 40.

La coscienza percettiva non può imporsi totalmente a causa del dubbio che lentamente si insinua, turbando il normale decorso della percezione; viceversa, però, la coscienza d'immagine non riesce a celebrare il suo trionfo a causa dell'incredibile verosomiglianza della cera, della sua capacità di somigliare a tal punto alla figura in carne ed ossa, che fa ricadere in continuazione il soggetto nell'errore e riporta l'apprensione figurale a quella percettiva.

Ecco il paradosso della scena originaria, di un'immagine che non è un'immagine: la coscienza percettiva non riesce ad imporsi stabilmente (come abbiamo visto, infatti, a causa dell'assenza di movimento della figura in cera il dubbio riesce ad insinuarsi e a farci sospettare dell'inganno delle nostre percezioni); d'altra parte nemmeno la coscienza d'immagine riesce ad avere la meglio.

Nonostante il dubbio, la somiglianza così iperrealistica della cera fa sì che di nuovo la coscienza ricada nell'errore delle percezioni: sappiamo razionalmente che non è così, ma le nostre percezioni dicono altro. È un continuo oscillare tra percezione, dubbio, consapevolezza e, tuttavia, di nuovo contrasto con la percezione. Come uscire da questa situazione di stallo? Come superare l'aporia con cui il nostro cervello e le nostre percezioni si devono rapportare? È possibile farlo, o ci troviamo davanti ad una situazione da cui facciamo fatica a svincolarci, ricadendo in una valle del perturbante?¹⁷

Per rispondere a queste domande, sarà necessario fermarsi ancora un attimo in questa fase del contrasto tra le nostre percezioni, nello stallo tra coscienza percettiva e coscienza d'immagine, cercando di definire la sensazione che l'identità tra il modello e l'oggetto a cui si riferisce e la mancanza di contrasto, portano nel soggetto che si scopre vittima di illusione.

Fino a questo punto, infatti, abbiamo trattato uno specifico caso: quello della cera e del suo scambio per una realtà in carne ed ossa. Manteniamo il punto fermo che abbiamo esposto inizialmente: il discorso sull'arte ed iperrealismo, che se un tempo poteva essere applicato alla dimensione delle statue in cera e dei manichini iperreali, nell'epoca attuale ha assunto nuove forme e nuove applicazioni: siamo arrivati ad una sorta di "*iperrealismo 2.0*" con l'evoluzione dei nuovi media digitali: ora lo scambio non è più tra l'immagine e l'oggetto

¹⁷ Si vedrà a riguardo, in maniera più dettagliata, la definizione di *perturbante* che dà Freud nell'omonimo saggio del 1919.

reale (immagine fisica e oggetto a cui essa si riferisce nella realtà), ma tra immagini (foto e video) che vengono scambiate per altre immagini reali. Non si scambia la realtà in carne ed ossa: l'inganno si insinua tra le immagini stesse. Questo inganno dà vita ad uno specifico effetto nello spettatore: il *perturbante*.

1.1.2. Il perturbante

“Che l'opera d'arte, e in particolare il ritratto, sia dotata di vita, è una delle convinzioni più primitive con cui gli spettatori ingenui cercano di raccapezzarsi di fronte alla creazione artistica”¹⁸.

Così afferma Schlosser in “*Storia del ritratto in cera*”. L'autore, uno dei rappresentanti della Scuola di Vienna, fonda il testo su un assunto teorico fondamentale: l'incontrollabilità delle reazioni umane di fronte alle opere artistiche, alle immagini e in particolare alle immagini iperrealistiche. Nello stesso periodo in cui Schlosser pubblica questo testo, Freud elabora un saggio basato sul concetto di “*perturbante*”¹⁹: questo effetto viene suscitato quando qualcosa con cui si ha familiarità, qualcosa che dovrebbe rassicurare e mettere a proprio agio, subisce un cambiamento e si fa improvvisamente estraneo e, di conseguenza, inquietante. L'effetto di straniamento si stabilisce tra la dimensione della familiarità (percepisco qualcosa che sento come vicino, conosciuto, che dovrebbe mettere a proprio agio, dunque rassicurante) e il contrasto con l'estraneità dello stesso. Come esempio concreto Freud porta proprio quello delle statue in cera:

Spesso e volentieri ci troviamo esposti a un effetto perturbante quando il confine tra fantasia e realtà si fa labile, quando appare realmente ai nostri occhi qualcosa che fino a quel momento avevamo considerato fantastico, quando un simbolo assume pienamente la funzione e il significato di ciò che è simboleggiato, e via di questo passo²⁰.

¹⁸ J. Von Schlosser, *Storia del ritratto in cera* (1911), a cura di P. Conte, intr. di G. Didi-Huberman, postfaz. di P. Conte, Quodlibet, Macerata 2011, p.41.

¹⁹ S. Freud, *Il perturbante* (1919), tr. It. di S. Daniele, in Id., *Opere*, a cura di C.L. Musatti, 12 voll., Bollati Boringhieri, Torino 1977.

²⁰ S. Freud, *Il perturbante*, cit., p.105.

Come abbiamo già visto precedentemente, l'esempio della statua in cera richiama la teoria della percezione: il contrasto tra le percezioni, il fatto che le nostre aspettative future vengano disattese (discordanza tra passato, presente e futuro) porta ad un effetto di disorientamento, il dubbio paralizza lo spettatore: egli non comprende più il confine tra immagine e realtà, o addirittura la possibilità di un'animazione dell'immagine stessa.

Andiamo avanti di qualche decennio ed applichiamo il concetto freudiano di perturbante dall'esempio delle statue di cera a qualcosa di ancora più attuale: nel 1970 Masahiro Mori, studioso giapponese di robotica, pubblica un saggio nella rivista *Energy*, esponendo un concetto chiamato “valle del perturbante” o “zona perturbante” (*Uncanny valley*)²¹. In breve, egli analizza come la sensazione di familiarità provata da un campione di persone a contatto con degli androidi in tutto e per tutto simili a persone in carne ed ossa possa aumentare man mano che queste somiglino all'uomo. D'altra parte, al culmine di questo realismo corrisponde un calo delle emozioni positive, sostituito da sensazioni spiacevoli (inquietudine e perturbamento).

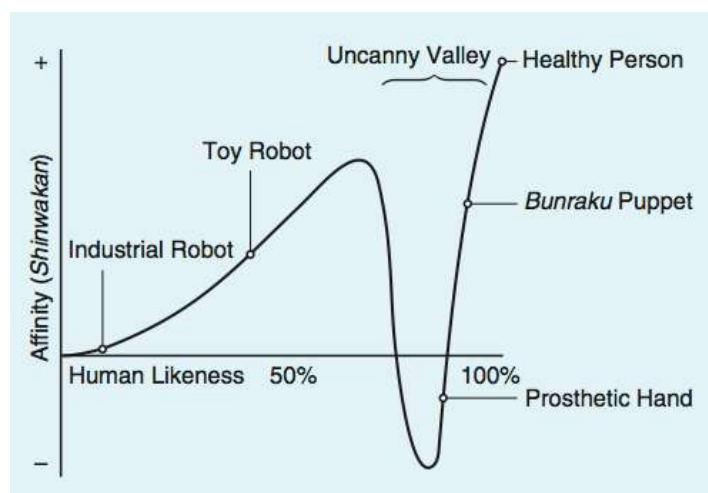


Fig.1: M. Mori, Uncanny valley.

²¹ Si veda a riguardo: M. Mori, *The Uncanny valley* (1970), tr. ingl. di Karl F. MacDorman and Norri Kageki, IEEE Spectrum, 2017.

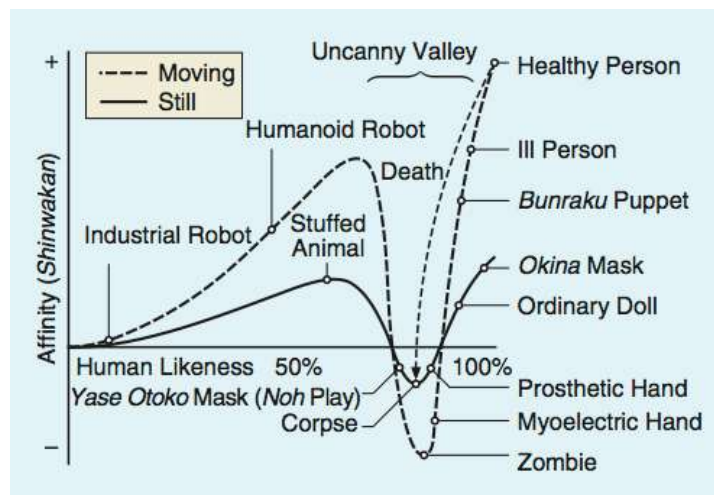


Fig. 2: M. Mori, Uncanny valley, schema grafico del movimento.

Nello specifico, l'ingegnere giapponese a contatto con le macchine comprende in che misura più la macchina assomigli all'uomo, più quest'ultimo riesca ad avere confidenza con la essa. L'andamento è proporzionale: maggiore è la somiglianza, maggiore sarà a sua volta anche la familiarità. Le persone prese in campione si rapportano agli androidi non come se essi fossero plastica e acciaio, ma come fossero essi stessi persone in carne ed ossa. Per Mori, tuttavia, la correlazione tra "Human likeness" e "familiarity" è solo in apparenza biunivoca: arrivati ad un certo punto, al culmine della somiglianza (come possiamo vedere dal grafico, fig. 1), non corrisponde più una crescita del livello di familiarità, bensì un crollo improvviso. L'empatia positiva diventa empatia negativa, quando la somiglianza raggiunge livelli troppo elevati si verificherà l'effetto opposto: la caduta dei livelli di familiarità produrrà proprio questa "Uncanny valley". Parlando del movimento, possiamo notare come esso estremizzi i picchi in entrambe le direzioni (fig. 2): un androide dotato di movimento sarà ancora più simile all'essere umano, ma proprio per questo ancora più perturbante.

Torniamo alla scena originaria: abbiamo le prove, sappiamo di essere vittime di inganno, di illusione, ma ricadiamo ogni volta nell'errore: la statua continua a turbare le nostre percezioni. Sta proprio qui il significato del perturbante: quando la somiglianza è così netta, quando l'iperrealismo raggiunge livelli così alti, diventa difficile realizzare completamente l'inganno, prendere le distanze da ciò che vediamo. La verosomiglianza aumenta, proporzionalmente aumenta la nostra familiarità, allo stesso modo aumentano le

sensazioni perturbanti in noi nel momento in cui realizziamo di trovarci davanti a qualcosa di fittizio (che si tratti della statua di cera husserliana, oppure che si tratti degli androidi di Mori). È come se freudianamente ci trovassimo in un sogno, in una dimensione illusoria da cui non riusciamo completamente a svegliarci e distaccarci. Gli indizi ci sono, il dubbio si insinua, ma come possiamo distinguere le nostre percezioni? Quali sono le percezioni veritiere e quali quelle false? Cosa ci può dire il nostro raziocinio davanti a ciò, come può discernere tra le percezioni ed elaborarle in maniera coerente, interpretandole correttamente? È una situazione paralizzante.

Per quanto riguarda le statue di cera possiamo dire che:

L'effetto perturbante che ineluttabilmente riescono a generare è dovuto proprio ad un inaggirabile scarto: da un lato sono troppo simili all'uomo per non essere prese per vere, dall'altro però alcune loro caratteristiche – tra cui in primo luogo l'assenza, o quanto meno la meccanicità dei movimenti – fanno sì che vengano intese come statue, artefatti, e non persone vere e proprie.²²

Un continuo contrasto tra iperrealismo e, d'altra parte, presenza di differenze pronte a smascherarlo. Questo fenomeno è stato spiegato da alcuni ricercatori che hanno sostenuto e rielaborato la teoria di Mori²³: il nostro cervello è costituito in maniera tale da formulare ipotesi sulla base di indizi e dati apprensionali simili. Una volta vissuta una determinata esperienza, il nostro cervello sarà in grado di formulare ipotesi “predittive”. Non solo registrerà e memorizzerà stimoli, ma li rielaborerà in modo tale da analizzare l'ambiente circostante formulando previsioni coerenti. Nel momento in cui le nostre previsioni non siano più coerenti con gli stimoli o, addirittura, vi siano delle contraddizioni tra gli stimoli dello stesso ambiente, si verificherà un fenomeno chiamato “dissonanza cognitiva”²⁴. Questo lascerà il soggetto in una condizione di disorientamento nell'ambiente circostante. Prendendo l'esempio dell'androide: ci aspettiamo che si comporti in un certo modo, che dia atto a una serie di azioni e invece si verifica qualcosa di inatteso, che contraddice le nostre aspettative e le nostre previsioni.

²² P. Conte, *Unheimlich*. Dalle figure di cera alla Uncanny Valley, PSICOART n. 2, 2011-12, pp.13-14.

²³ Tra cui Karl F. Macdorman, uno dei primi a fornire una traduzione in inglese della teoria di Mori (2005) dando vita ad un intenso dibattito a riguardo.

²⁴ Cfr. P. Conte, *In carne e cera. Estetica e fenomenologia dell'iperrealismo*, p.47.

Facciamo ancora un passo avanti, che ci porterà ad approfondire in seguito il nostro discorso: al giorno d'oggi lo sviluppo delle immagini digitali ha raggiunto dei progressi notevoli. I nuovi media immersivi e i mezzi per elaborare le animazioni digitali hanno raggiunto livelli molto alti di realismo e verosomiglianza, grazie allo sviluppo della tecnologia. Lo vediamo in diversi campi, nel cinema come nei video su piattaforme digitali. Anche in questo caso possiamo notare come il lavoro di elaborazione digitale e di animazione renda più facile lo sviluppo di personaggi sempre più realistici, simili agli esseri umani ma diversi per quanto riguarda ad esempio l'espressione dell'emotività: ancora, ricadiamo nel perturbante e nella *uncanny valley*. La sensibilità per le differenze diventa ancora maggiore una volta che aumentano i livelli di verosomiglianza, allo stesso modo maggiore diventa l'effetto di spaesamento, disorientamento, perturbamento.

Mori, infatti, nel suo saggio per la rivista *Energy* afferma: “I have noticed that, in climbing toward the goal of making robots appear human, our affinity for them increases until we come to a valley, which I call the *uncanny valley*”²⁵.

Altri autori (come, ad esempio, K. F. Macdorman²⁶) affermano che esistono dei modi per evitare questo effetto perturbante: “The effect can be reduced by not creating robots or computer-animated characters that combine features on different sides of a boundary – for example, human and nonhuman, living and nonliving, or real and artificial”²⁷.

Evitare, dunque, alcune caratteristiche che rendono ancora più verosimile l'androide (superando i confini tra reale e artificiale) come, appunto, il movimento e l'emotività.

Traendo le conclusioni di questo discorso, possiamo affermare che il concetto di somiglianza eccessiva mette in seria discussione il concetto di rappresentazione: seguendo il discorso di Husserl abbiamo, infatti, compreso come per avere una rappresentazione

²⁵ M. Mori, tr.: “Ho notato che, andando verso l'obiettivo di far apparire umani i robot, la nostra affinità verso di essi aumenta man mano finché non arriviamo ad una valle, che chiamo la valle del perturbante”.

²⁶ Professore associato di robotica e “interazione uomo-computer” presso l'Università dell'Indiana, che ha a lungo studiato la teoria di Mori.

²⁷ “L'effetto può essere ridotto evitando di creare robot o personaggi animati al computer che combinano caratteristiche di entrambe le parti, superandone i confini - come ad esempio umano/non umano, vivente/non vivente o reale/artificiale. Si veda a riguardo Rina Diane Caballar, *What is the uncanny valley?* Su IEEE Spectrum, 06 Novembre 2019 (<https://spectrum.ieee.org/what-is-the-uncanny-valley>) URL visitato in data 14 agosto 2021.

adeguata sia necessaria una distanza tra il rappresentante e ciò che viene rappresentato, così come tra l'opera e il fruitore dell'opera²⁸. Nel caso delle statue di cera, degli androidi e delle immagini virtuali iperrealistiche, questo confine viene varcato, la distanza non è più presente, l'immagine e il modello si confondono a tal punto da diventare un tutt'uno. Non si possono prendere le misure, non ci si può orientare. Ogni confine tra reale e virtuale sembra varcato.

Cosa ne sarà dell'arte a questo punto? E cosa accadrà, appurate queste condizioni, con l'uso dei nuovi media digitali? Con lo sviluppo dei mezzi e della tecnologia gli interrogativi aumentano, l'indagine continua.



Fig. 3: Sophia, robot umanoide creato da David Hanson di Hanson Robotics. È dotata di una gamma di espressioni emotive e ha la capacità di processare ed elaborare naturali proprietà di linguaggio.

²⁸ Importante la questione della distanza e della cornice. Secondo Simmel, la cornice genera isolamento e concentrazione: "Quel che la cornice procura all'opera d'arte è il fatto che essa simboleggi e rafforzi questa doppia funzione del suo confine. Essa esclude l'ambiente circostante, e dunque anche l'osservatore, dall'opera d'arte e contribuisce a porla a quella distanza in cui soltanto essa diventa esteticamente fruibile" (G. Simmel, *La cornice del quadro. Un saggio estetico*, in *I percorsi delle forme, i testi e le teorie*, a cura di Maddalena Mazzocut-Mis, Bruno Mondadori, Milano, 1997, p. 210).



Fig. 4: Geminoid Hi, creazione di Hiroshi Ishiguro.



Fig. 5: Telenoid, dispositivo di comunicazione robotica progettato da Hiroshi Ishiguro. Le sue sembianze lo fanno sembrare più alieno che umano. È, infatti, tra i primi posti in classifica tra i robot più “perturbanti” della Robots Guide di IEEE Spectrum.

1.2. I new media

1.2.1 Fenomenologia dei new media

Dalla scena originaria husserliana siamo arrivati ad un nuovo sviluppo, un nuovo stadio: l'iperrealismo dei new media. Il cammino di questi studi fenomenologici, le analisi sopra esposte a livello teorico (ovvero come le nostre percezioni si applichino al mondo dell'arte e, nello specifico, all'uso iperrealistico della cera) verranno direzionati nei confronti degli ambienti virtuali generati con l'ausilio delle più recenti tecnologie dell'immagine. I punti di forza e i punti critici di questi approcci teoretici, dunque, verranno ora messi in luce nell'approccio alla rapida evoluzione dei nuovi media immersivi.

Lambert Wiesing negli anni 2000 scrive un testo che è una raccolta di studi sulla filosofia dell'immagine. *“Artificial presence”*: come le immagini rendano possibile la presenza artificiale delle cose. Questi studi di fenomenologia sulle immagini vengono però applicati ai tipi estremi di immagini virtuali per il lavoro filosofico sul concetto di immagine. L'oggetto, ovvero la presenza artificiale dell'immagine, non è soggetto alle leggi della fisica: Wiesing prende le distanze dall'approccio semiotico dell'immagine da un punto di vista fenomenologico. Il paradigma semiotico concepisce un soggetto-immagine ed un oggetto-immagine. Vi è, inoltre, la possibilità di utilizzare un'immagine come segno. Il paradigma semiotico da Wiesing viene inteso come plausibile, ma allo stesso tempo non essenziale e soprattutto non tipico dell'immagine, in quanto applicabile agli oggetti in generale. Anche il paradigma mimetico viene considerato applicabile agli oggetti reali che si assomigliano. L'alternativa semiotica viene scartata poiché descrive la relazione tra “oggetto-immagine” e “soggetto-immagine”, ma non viene revocata poiché descrive ugualmente la relazione tra immagine fisica e oggetto immagine. Anche il paradigma mimetico viene considerato uno dei modi per descrivere il segno e, tuttavia, l'approccio fenomenologico risulta il più adeguato a comprendere la relazione tra immagine fisica e oggetto immagine. In che modo?

Abbiamo accennato come il paradigma fenomenologico sottolinei la “presenza artificiale” dell'oggetto-immagine. Wiesing descrive la “presenza reale” in questo modo:

“Un oggetto presente nel mondo reale è necessariamente un oggetto che agisce secondo le leggi della fisica, il che significa che può agire fisicamente sul corpo umano”²⁹.

Per sottolineare come questo concetto di *presenza artificiale* sia inteso come controparte di una comprensione mimetica dell’immagine, Wiesing aggiunge: “L’immagine non viene mostrata per somiglianza a immagini, ma mostrando cose che non sono sottostanti alle leggi della fisica”³⁰.

Wiesing cerca anche di rispondere alla domanda di come la *presenza artificiale* venga raggiunta: si riferisce alla distinzione husserliana tra “*Geltung*”³¹ e “*Genesis*” e descrive come questo fenomeno sia tipico dei media: un segno che ha un significato, oppure un’immagine fisica che rivela un’immagine-oggetto.

Inoltre, l’autore afferma che mentre l’oggetto-immagine rappresentato dall’immagine è sottoposto all’invecchiare del tempo, l’immagine che lo rappresenta non lo è. Aggiunge, inoltre, che un’immagine bidimensionale non cambia vista da differenti angoli. Questo dimostra sempre la differenza tra immagine fisica e oggetto immagine: l’oggetto immagine non cambia quando la nostra posizione rispetto all’immagine fisica varia. Wiesing stesso scrive che, al contrario dell’immagine su uno schermo cinematografico che è uguale per tutti, la prospettiva su un teatro varia a seconda dello spettatore³².

Cerchiamo di dare un quadro complessivo del nostro discorso: Wiesing rifiuta l’approccio semiotico all’immagine, poiché essa può essere utilizzata come segno, ma è principalmente un fenomeno a sé stante. Egli rifiuta anche l’approccio mimetico all’immagine, non riuscendo a capire come una mera apparenza visiva sia sufficiente a suggerire la piena presenza di un oggetto immagine. Wiesing sostiene, inoltre, che l’oggetto immagine non sia soggetto alle leggi della fisica e che sia “pura visibilità”, trascurando il

²⁹L. Wiesing: *Artifizielle Präsenz Studien zur Philosophie des Bildes*, Suhrkamp, 2005. tr.: “Ein Gegenstand in der Welt mit realer Präsenz ist Notwendigerweise ein Gegenstand, welcher sich nach den Gesetzen der Physik verhält, was wiederum bedeutet, dass er physisch auf den menschlichen Körper einwirken kann”, cit. p.31.

³⁰Ivi, tr.: “Bilder werden nicht durch Ähnlichkeit zu Bildern, sondern durch das Zeigen von Dingen die nicht den Gesetzen der Physik unterlegen”, cit. p. 59.

³¹ A tale proposito occorre ricordare come per Husserl anche gli oggetti di fantasia (“*ficta* della fantasia”) abbiano una validità (*Geltung*) a livello intersoggettivo: anche mondi fantastici come quelli della letteratura, delle fiabe e del cinema non sono un mero prodotto della mente di un singolo individuo, non sono “un nulla”, bensì esibiscono una validità.

³² Ivi, p. 161.

fatto che l'oggetto immagine abbia un doppio "status ontologico": esso è in parte reale e in parte artificiale. Solo questo doppio status ontologico può spiegare la "comparsa" dell'oggetto immagine come "presenza artificiale". Sia la fenomenologia sia la semiotica negano che la somiglianza sia la nozione chiave per comprendere l'immagine: esse confondono la somiglianza tra soggetto dell'immagine e oggetto-immagine come relazione mimetica (modello). Tuttavia, anche se alcuni oggetti-immagine assomigliano ad oggetti del mondo reale, non si tratta della somiglianza che costituisce un'immagine (anche perché non tutte le immagini hanno un oggetto come modello nel mondo reale). Costitutiva dell'immagine è la somiglianza in parte tra il mezzo mimetico e l'oggetto immagine.

Abbiamo visto l'impianto teorico, facciamo ora un passo in avanti. Wiesing, parlando di "virtual reality" e "new media" espone la necessità di distinguere tra due tipi di realtà virtuale: quella immersiva (come la conosciamo dal "cyberspazio") e quella non immersiva (come la conosciamo dai videogames)³³.

Per l'autore la realtà virtuale immersiva emerge dall'assimilazione della percezione dell'immagine-oggetto (Bildobjekt) con la percezione di una cosa reale, d'altra parte invece la realtà virtuale non immersiva emerge dall'assimilazione dell'oggetto immagine con l'immaginazione. Questa assimilazione degli oggetti nell'immaginazione è la novità delle immagini prodotte dai new media.

Wiesing distingue le immagini facendo due tipi di distinzioni³⁴: immagini interne ed esterne. Il confronto esterno andrà fatto con fenomeni che non sono immagini, il mezzo dell'immagine andrà confrontato ad esempio con il linguaggio concettuale, le percezioni e le capacità cognitive; il confronto interno invece andrà fatto con i diversi tipi di immagini (fotografia, dipinti, animazioni, film ecc.) Bisogna però avere un sicuro metodo per distinguere tra interno ed esterno. L'autore ci porta un esempio di come differenti tipi di immagine possano essere collegati ad un fenomeno esterno.

Wiesing porta l'esempio di una rappresentazione di un cavallo: abbiamo tre cavalli, ovvero quello reale, quello fantasioso e quello dipinto in immagine. In tutti e tre i casi la coscienza è intenzionalmente orientata ad un cavallo. Anche se si tratta di tre atti diversi dal

³³ Ivi, p. 89.

³⁴ Ivi, cfr. p. 90.

punto di vista della coscienza, si tratta pur sempre di un atto di coscienza. Quello che manca per una descrizione precisa di ogni atto intenzionale è la descrizione della differenza specifica con cui ogni atto intenzionale della coscienza si distingue dall'altro. Quando vediamo un cavallo, o lo immaginiamo, o lo vediamo in un'immagine, siamo, in diversi modi, consapevoli di un cavallo. Dunque, il compito è determinare il modo in cui si dà ciascuna di queste modalità della coscienza. Questo diventa chiaro per la "coscienza della percezione". La questione più interessante della teoria dei media è la posizione intermedia dell'immagine oggetto tra gli oggetti della percezione e gli oggetti di fantasia³⁵, posizione che può essere descritta fenomenologicamente attraverso un confronto esterno, e che non è la stessa per ogni mezzo dell'immagine.

La novità dei nuovi media consiste proprio nella facilitazione da parte del computer della produzione di immagini che, in primo luogo, si allontanano dalla posizione intermedia appena descritta, in modi precedentemente sconosciuti e, secondo, assimilano radicalmente il modo di darsi degli oggetti di fantasia al modo in cui si presenta allo spettatore un oggetto immagine.

Con l'arrivo dei new media abbiamo, infatti, questa novità: il computer facilita la produzione di immagini che si scostano dalla via intermedia e assimilano il modo di darsi dell'oggetto immagine all'oggetto di fantasia. La novità dei new media non sono le nuove percezioni, bensì il "*nuovo potere delle immagini*": è come se venisse esteriorizzata l'immaginazione, l'immagine-oggetto diventa un oggetto di fantasia esteriorizzato.

Abbiamo ripreso il discorso fenomenologico, ma facendo un passo in più verso la contemporaneità. Non ci stiamo più fenomenologicamente approcciando ad un fantoccio in cera, ma a dei nuovi mezzi, ad una nuova realtà: ci stiamo avvicinando e stiamo entrando progressivamente nell'epoca di una cultura iperconnessa, nell'epoca del virtuale. Seguendo questo cammino stiamo progredendo verso l'iperrealismo dei new media, fino alla nascita di una "realtà iperreale", che, al contrario della statua di cera husserliana, può essere creata *ex-novo*. Ci stiamo avvicinando all'epoca della *simulazione*.

³⁵ Ivi, p. 92.

Ancora alcune precisazioni riguardo al tema dell'iperrealismo nell'epoca contemporanea (nell'era, per l'appunto, del virtuale): l'autore Krešimir Purgar definisce l'iperrealismo come uno stile artistico caratterizzato dall'assenza di discontinuità rispetto alla realtà. È proprio riguardo al tema dell'assenza di discontinuità, infatti, che si pongono i principali interrogativi. L'autore cita lo stesso Wiesing, affermando:

Lambert Wiesing inverts the fundamental question of phenomenology that reads as follows: In what way does the observer with his intentional awareness produce or, more specifically, bring into awareness the perceived object? and turns it into a question: What consequences does the reality of perception have for the observer?³⁶

Il focus diventa, dunque, il punto di vista dell'osservatore, l'impatto che ha l'oggetto percepito sul soggetto che percepisce. L'effetto artistico dell'iperrealismo per Purgar è paradossale: il dipinto così perfetto della realtà dovrebbe essere compreso come fonte di dubbio delle cose e delle apparenze che cessano di essere come le vediamo. La differenza tra le rappresentazioni pittoriche e le immagini create dal computer è che possiamo riconoscere un grande salto: l'area di transizione tra immagine e non-immagine, tra realtà e finzione³⁷. Siamo arrivati al "grande salto", al punto di svolta.

1.2.2. L'epoca della simulazione

«É in discussione il sistema di rappresentazione. L'immagine che l'uomo ha di sé stesso è virtualizzata. Non si è più davanti allo specchio, si è dentro lo schermo, che è tutta un'altra cosa».³⁸

Siamo a fine Novecento, in Francia. Un lungimirante pensatore, Jean Baudrillard, compie una puntuale analisi dell'epoca a lui contemporanea, anticipando alcuni eventi che si

³⁶ K. Purgar, *The meaning of hyperrealism today*, tr: "Lambert Wiesing inverte la questione fondamentale della fenomenologia che è: in che modo l'osservatore con la sua consapevolezza intenzionale produce o, più specificatamente, porta a consapevolezza l'oggetto percepito? E la trasforma nella domanda: quali conseguenze ha la realtà della percezione per l'osservatore?", cit. p.104.

³⁷ Cfr. K. Purgar, *The meaning of hyperrealism today*, pp. 108-109.

³⁸ C. Thibaud, *Intervista con Jean Baudrillard*, in *Cybersphere*, n. 9, tr. it. di A. Venieri e I. Spelletti *Cybersphere*, n° 9, Arci S:M:S: di Rifredi-Special!, 1998.

sarebbero verificati da lì a pochi anni: l'uso delle nuove tecnologie e di come esse avrebbero influenzato e rivoluzionato la società. Egli preannuncia la nascita di una società dominata dalla presenza della tecnologia, di una collettività oltremodo composta da individui sedotti dall'estasi comunicativa entro uno spazio iperreale³⁹. Nei luoghi della cultura come i musei, è possibile arrivare ad un punto in cui la semplice digitalizzazione e condivisione di immagini delle opere d'arte potrebbe corrispondere alla condivisione della cultura? Si può arrivare ad un punto in cui l'immagine sostituisca la realtà al punto tale che «la maggioranza silente ricerca l'immagine e non il significato»⁴⁰?

Baudrillard, a suo tempo, risponde a questo interrogativo affermando che sì, il virtuale potrà arrivare al punto di assorbire il reale. Nell'era del virtuale, per Baudrillard, la realtà scompare e diviene il suo contrario: egli criticamente afferma che ci troviamo immersi in una società del simulacro, mettendo in dubbio non solo il concetto di verità, ma anche quello di realtà. Nell'era della comunicazione virtuale i fatti scompaiono e cedono il posto ad un'apparenza che è il loro esatto contrario.

In questo modo la realtà scompare e viene sostituita dal contrario di quella che essa era: una anti-realtà, un'*iperrealtà*, la cui sola ragion d'essere è suggerirci, per contrasto, che la vera realtà esiste attorno a noi, mentre invece ci troviamo sprofondati nella sua caricatura, nella sua menzogna, nel suo contrario. Il simulacro ingannevole della realtà, la sua contraffazione virtuale, viene identificato come “la cosa reale”. Qui l'associazione tra il segno e la cosa reale è arbitraria, i segni funzionano da soli: non esprimono altro se non sé stessi. Al contrario, è la realtà a doversi uniformare ad essi. Ci troviamo nell'epoca della simulazione virtuale, copia di una realtà e di un'umanità più “perfette” di quelle originali?

Quando il reale non è più quel che era, la nostalgia assume tutto il suo significato. Mondo di verità, oggettività e autenticità seconde. [...] produzione frenetica di reale e referente, parallela e superiore alla frenesia della produzione materiale: così appare la simulazione nella fase che ci riguarda — una strategia del reale, di neoreale e iperreale, che raddoppia dovunque una strategia di dissuasione⁴¹.

³⁹ Cfr. J. Baudrillard, *Simulacres et simulations*, Paris: Galilee, 1981.

⁴⁰ J. Baudrillard, *Il delitto perfetto: la televisione ha ucciso la realtà*, Trad. di Gabriele Piana, Milano, R. Cortina, 1996.

⁴¹ J. Baudrillard, a cura di Matteo G. Brega, *Simulacri e impostura: bestie, Beaubourg, apparenze e altri oggetti*, Milano, Pgreco, 2008, p. 67.

Per l'epoca in cui scrive, nonostante si stessero già verificando alcuni eventi che avrebbero fatto pensare a ciò, Baudrillard può essere considerato una sorta di "precursore" dell'età contemporanea. Se al tempo in cui scrive, sia per quanto riguarda i "simulacri della cultura" sia per ciò che ci riporta ai new media e alle nuove tecnologie immersive, era al principio dello sviluppo, oggi assistiamo ad una crescita esponenziale di questo fenomeno.

La simulazione non è più quella di un territorio, di un'entità referenziale, di una sostanza. Oggi, essa genera attraverso i modelli di un reale senza origine né realtà: iperreale, il territorio non precede più la carta né le sopravvive. Ormai, è la carta che precede il territorio [...] qui e là sono vestigia del reale che sussistono, e non della carta, nei deserti che non sono più quelli dell'impero, ma il nostro. Il deserto del reale stesso⁴².

Per l'appunto ci troviamo nel deserto della realtà e del reale come era stato sempre conosciuto. Non parliamo più di realtà ingannevole ed illusoria come abbiamo fatto precedentemente, non si tratta più di quella rappresentazione che inganna lo spettatore, non siamo più gli osservatori ingannati dal manichino di cera, finto ma iperrealistico ed estremamente somigliante al modello a cui si riferisce. Non rappresentazione iperrealistica, dunque, ma piuttosto vera e propria sostituzione della realtà con l'apparenza, della realtà con il suo simulacro.

La *simulazione*, infatti, viene contrapposta da Baudrillard alla *rappresentazione*: mentre, come abbiamo già visto in precedenza, la rappresentazione è un'equivalenza del segno e del reale, la simulazione nega completamente il valore del segno: vi è un'equivalenza tra quest'ultima e la realtà. L'autore afferma che: "la simulazione avvolge tutto l'edificio della rappresentazione stessa come simulacro"⁴³, aggiungendo inoltre che "il passaggio dai segni che dissimulano qualcosa ai segni che dissimulano che non c'è niente, segna la svolta decisiva"⁴⁴.

Baudrillard definisce tutto ciò esattamente come "l'inizio dell'era della simulazione"⁴⁵, di una vera e propria sostituzione della realtà: "Non si tratta più di imitazione, né di

⁴² Ivi, pp. 59-60.

⁴³ Ivi, p. 66.

⁴⁴ *Ibidem*.

⁴⁵ Ivi, p. 61.

raddoppiamento, e neanche di parodia. Si tratta di una sostituzione al reale dei segni del reale”⁴⁶. Di conseguenza, secondo l’autore, più il reale avrà occasione di riprodursi, più sarà destinato alla sua morte. Più si moltiplicheranno i segni in grado di sostituire e riprodurre il reale, più esso sarà destinato a scomparire e a venire sostituito da essi.

Il risultato è, dunque, l’iperrealtà:

Quando il reale non è più quel che era, la nostalgia assume tutto il suo significato. Mondo di verità, oggettività e autenticità seconde. [...] produzione frenetica di reale e referente, parallela e superiore alla frenesia della produzione materiale: così appare la simulazione nella fase che ci riguarda — una strategia del reale, di neoreale e iperreale, che raddoppia dovunque una strategia di dissuasione⁴⁷.

Il discorso baudrillardiano va nella direzione di una sorta di *disvelamento*: ci invita a vedere nella stessa direzione, a vedere il reale attraverso i suoi doppi (i simulacri), dietro a quell’apparenza che ne è diventata completamente sostituzione. Nel suo discorso così lungimirante il filosofo francese arriva addirittura a prevedere una sorta di Grande Fratello degli anni 2000: espone la prima esperienza di TV-verità della famiglia Loud (1971), definendola un’“estetica dell’iperreale”⁴⁸: si tratta di una situazione in cui non c’è distanza, in cui tutto accade sotto gli occhi dello spettatore, non c’è copione e non ci sono attori. La famiglia vive sotto gli occhi dello spettatore *come se fosse reale*, e d’altra parte lo è, in un certo senso. Ma, come giustamente chiede Baudrillard, si tratta della verità di questa famiglia o della verità della TV? Cosa succede se non ci sono più né mezzo né distanza? Le cose che accadono a questa famiglia sarebbero state le stesse se non si fossero trovati nella “verità della TV”? Oppure è qualcosa di reale, ma distorto e condizionato dallo stesso mezzo attraverso cui viene presentata?

“La simulazione è padrona, e noi abbiamo diritto soltanto al *retro*, alla riabilitazione fantomatica e parodica di tutti i referenti perduti”⁴⁹.

L’autore francese ci parla anche dei luoghi di cultura contemporanei, del Beaubourg (come esempio per eccellenza): di quei luoghi in cui la cultura è stata stigmatizzata e

⁴⁶ Ivi, p. 61.

⁴⁷ Ivi, p. 67.

⁴⁸ Ivi, p. 92.

⁴⁹ Ivi, p. 108.

riprodotta in maniera sterile, luoghi che rappresentano questa progressiva accettazione e addirittura il radicamento della società dell'apparenza.

Lo scopo (a prescindere dai diversi campi in cui questo fenomeno avviene) è sempre lo stesso: falsificare la realtà, farne scomparire la vera immagine e sostituirla con una irreale, manipolata, da utilizzare per qualche secondo fine che non ha nulla a che fare con l'autentica informazione.

In questo senso i mezzi di comunicazione di massa, ma specialmente la televisione, operano un sistematico rovesciamento della realtà. Di conseguenza, il cittadino della società moderna viene a trovarsi nella paradossale situazione di non essere più il soggetto della propria informazione, ma l'oggetto: è la televisione che lo "pensa"; ne abbiamo la prova anche per quanto riguarda quei fatti di cronaca che, diventando eventi mediatici, finiscono per oscurare e nascondere completamente, addirittura capovolgendo nel suo contrario, l'evento reale.

Possiamo affermare, seguendo il discorso baudrillardiano, di essere sottoposti ad una sorta di "isteria del nostro tempo": la produzione e riproduzione del reale, l'iperrealismo della simulazione si traduce nell'allucinante rassomiglianza del reale a sé stesso.

Siamo giunti a tutti gli effetti nell'era della simulazione⁵⁰: il simulacro non nasconde più la verità, esso è la verità.

Se Kant ci parlava di fenomeno e noumeno (o cosa in sé), Baudrillard riprende questo concetto ma adeguandolo all'epoca della tecnica: è come se la tecnica avesse in qualche modo sostituito fino a farla scomparire la "cosa in sé". Il mondo ha iniziato pian piano a scomparire, sostituito per l'appunto dalla simulazione.

In un altro testo emblematico l'autore francese ci fa comprendere in maniera ancora più approfondita questo fenomeno: come l'immagine, la televisione, abbiano ucciso la realtà.

“Questa è la storia di un delitto: l'uccisione della realtà. E dello sterminio di un'illusione: l'illusione vitale, l'illusione radicale del mondo. Il reale non scompare nell'illusione, è l'illusione che scompare nella realtà integrale”⁵¹.

⁵⁰ Ivi, p. 20.

⁵¹ J. Baudrillard, *Il delitto perfetto: la televisione ha ucciso la realtà*, cit. p.2.

L'immagine *virtuale* ha sostituito quella *reale*. Tutto è informazione, siamo circondati costantemente da messaggi e rappresentazioni, ma se tutto è informazione sembra che niente informi più davvero. La filosofia si è sempre posta la domanda sul perché delle cose, fin dalla sua nascita indagando il reale si è posta la domanda sull'essenza delle cose. Ma in quest'epoca, per Baudrillard, la questione cambia: "La grande questione filosofica era: «Perché c'è qualcosa piuttosto che nulla?». Oggi, la vera questione è: «Perché c'è niente piuttosto che qualcosa?»⁵².

Viviamo, infatti, in un mondo in cui vi è un continuo accumularsi, una proliferazione incondizionata ed inarrestabile di: informazioni, pubblicità, immagini. Non sappiamo più, secondo l'autore, affrontare la sottrazione, la "potenza dell'assenza". Finisce anche l'illusione stessa: l'immagine coincide con il reale, essa è *il reale*. Non c'è illusione perché non ci sono immagini e referenti, il reale e il reale virtuale sono inestricabilmente connessi. La proliferazione di immagini è accelerata a tal punto che "non vi è più nulla da vedere": l'arte e i media fanno lo stesso. Come abbiamo detto si tratta della fine dell'illusione. Ma in che modo?

Ciò che si oppone alla simulazione non è dunque il reale, che ne costituisce solo un caso particolare, è l'illusione. E non c'è crisi della realtà, proprio al contrario: ci sarà sempre più realtà, poiché è prodotta e riprodotta mediante la simulazione, e non è essa stessa che un modello di simulazione. La proliferazione della realtà, come di una specie animale di cui sarebbero stati eliminati i predatori naturali, costituisce la nostra vera catastrofe. È il destino fatale di un mondo oggettivo⁵³.

Ecco emergere un concetto fondamentale: quello di *iperrealtà*. Questo concetto, anche al giorno d'oggi, può trovare un fondamento nella distruzione che è stata messa in atto dai nuovi mezzi di comunicazione riguardo la facoltà di immaginare ed illudersi: per Baudrillard, infatti, la *simulazione* non è esattamente in contrasto con realtà, bensì con l'*illusione*. Non è più possibile illudersi, non sussiste più nemmeno questa possibilità: l'iperrealtà produce l'effetto della derealizzazione e, di conseguenza, della perdita di sussistenza della realtà. La diffusione dei mezzi di comunicazione produce così un nuovo tipo di realtà: ad una potenza

⁵² Ivi, p. 3.

⁵³ Ivi, p. 9.

elevata si raggiunge la più assoluta oggettività e, infine, si arriva ad una totale anestetizzazione della sensibilità soggettiva e del giudizio critico. L'illusione era un elemento essenziale nell'esperienza soggettiva; senza di essa il mondo non può più esistere come oggetto di un io soggetto, ma come pura oggettività.

Questo processo è avvenuto in maniera talmente incontrollata da essere percepito come naturale e addirittura come una enorme opportunità: è l'uomo che usa questi mezzi tecnologici o è la tecnica che affianca e sostituisce l'uomo?

La televisione, dunque, sottotitolo dell'opera, mano agente nel delitto della realtà e sostituzione di essa con il suo simulacro, deve essere intesa sia in senso letterale ma anche come rappresentante dei diversi mezzi di comunicazione virtuali, dei nuovi media digitali in senso ampio.

Se vengono distrutte l'illusione e la visione soggettiva del mondo, viene distrutto anche quel mondo che attraverso la nostra esperienza costruiamo. La conseguenza per Baudrillard è l'emergere di un'unica illusione, che è anche la più pericolosa: la convinzione che il mondo sia unico e oggettivo, identico per ogni soggetto.

E così la profezia è realizzata: viviamo in un mondo in cui la suprema funzione del segno è quella di far scomparire la realtà e di mascherare nel contempo questa scomparsa. Oggi l'arte non fa altro. Oggi i media non fanno altro. Ecco perché sono condannati al medesimo destino⁵⁴.

Questo sistema di comunicazione totalizzante comporterebbe un'ulteriore estrema conseguenza: la scomparsa dell'Altro, il suo sterminio⁵⁵, la rimozione della capacità di un comunicare effettivo sostituita invece da un mondo di apparire, di immagini, di oggettività e simulacri. Non esiste più il problema della verità né quello della realtà: sono stati sostituiti entrambi da immagini attraverso cui, in realtà, non è possibile vedere. Siamo giunti al momento di massimo sviluppo dell'umanità o alla sua scomparsa e sostituzione?

Resta il problema di sapere se il progetto tecnico della Virtualità sia una funzione ascendente della specie umana, o un momento della sua vertiginosa scomparsa (le due cose non sono incompatibili). Non abbiamo forse inventato un

⁵⁴ Ivi, p. 5.

⁵⁵ Ivi, p. 49.

mezzo assai indiretto di radicalizzare la nostra esistenza dandole la possibilità di scomparire totalmente? Tutte le altre culture hanno lasciato delle tracce. Il nostro delitto sarebbe perfetto, poiché non lascerebbe tracce, e sarebbe irreversibile⁵⁶.

1.2.3. New media e fake news

Siamo giunti all'uccisione della realtà, baudrillardianamente, al “delitto perfetto”. Abbiamo compiuto il grande passo, siamo giunti all'epoca del simulacro, alla simulazione nell'epoca digitale.

L'inganno è compiuto, l'osservatore non è più in grado di distinguere ed orientarsi, persino l'illusione dell'osservatore husserliano è stata eliminata. Iperrealismo e mondo oggettivo: questa è la realtà presentata attraverso lo sguardo di Baudrillard nell'epoca digitale. Non abbiamo ancora considerato, però, uno dei possibili risvolti posti da questo scenario. Abbiamo infatti considerato, fino a questo punto, la desensibilizzazione e l'anestetizzazione del soggetto immerso nel mondo virtuale: la sua è un'immersione nell'epoca dominata da immagini e continui bombardamenti mediatici, ma non abbiamo ancora considerato un aspetto: c'è un ulteriore possibile risvolto, l'inganno all'apice del suo potere: al giorno d'oggi assistiamo non solo all'avverarsi dello scenario delineato da Baudrillard, ma addirittura ad un superamento di questa prospettiva. Siamo, infatti, nell'epoca in cui la realtà, oltre ad essere *iperreale* attraverso i nuovi media immersivi, viene addirittura sostituita da “una nuova realtà”, creata dagli stessi. Con l'avvento dei social media è possibile creare nuove illusioni, realtà fasulle. Abbiamo superato l'identificazione tra reale e virtuale, assistendo addirittura alla creazione di un mondo “post-verità”⁵⁷.

Con l'avvento dei social media è scomparso il *gap* che divideva mondo reale e mondo virtuale, abbiamo varcato quel confine in modo tale da poterli identificare abbastanza nettamente. Tuttavia, ciò che troviamo online è al contempo, per l'appunto, un mondo post-

⁵⁶ Ivi, p. 19.

⁵⁷ Il termine post-verità, “*post-truth*”, secondo l'Oxford Dictionary, compare per la prima volta nel 1992 sulla rivista *The Nation*, in un articolo scritto da Steve Tesich. Cfr., a riguardo, Alison Flood, ['Post-truth' named word of the year by Oxford Dictionaries.](#), su *theguardian*, 15 Novembre 2016. URL consultato il 7 agosto 2021.

verità, al cui interno sono sorte notizie deliberatamente falsificate o distorte in modo tale da orientare le decisioni individuali verso una determinata scelta (spesso in ambito politico): ecco sorgere il fenomeno delle fake news.

Fake news: un diverso modo di definire processi di disinformazione da sempre presenti, oppure nuovo fenomeno nato dall'instaurarsi di differenti meccanismi tecnologici e psicosociali? Che differenza c'è tra le fake news e la disinformazione presente da sempre nella storia?

Giuseppe Riva nel testo *Fake news: vivere e sopravvivere in un mondo post-verità* risponde a questo interrogativo affermando che il primo elemento di distinzione è la capacità di toccare la dimensione emotiva profonda del lettore. Le notizie false, infatti: “Come spiega Paolo Pagliaro “toccano l'emotività, le idee confuse e la rabbia montante di quello che è stato chiamato lo sciame digitale, cioè quel popolo anonimo che vive sulla rete””⁵⁸.

Riva prosegue affermando che, dall'analisi dei dati, emerge come la storia delle moderne fake news sia strettamente legata a quella di una specifica tecnologia: i social media. Esse utilizzano le proprietà di questa tecnologia per modificare la percezione che il soggetto ha nei confronti della realtà, facendogli credere che *siano* reali, ma anche per influenzarlo a tal punto da condividerle e diffonderle a sua volta, in maniera partecipata, coinvolta e con una ferma convinzione. L'autore ci spiega come funzioni effettivamente dal punto di vista dell'emotività questo meccanismo:

Facendo leva su un tessuto stereotipi e pregiudizi diffusi all'interno di un tessuto sociale, sfruttando emozioni facili e immediate da suscitare, quali l'ansia, il disprezzo, la rabbia e la frustrazione. La loro diffusione può contare su un uso manipolatorio dei social network e delle logiche che ne garantiscono il funzionamento.⁵⁹

L'emotività diventa una chiave, un modo di fare leva sulla maggioranza. “L'arma segreta” grazie a cui questo fenomeno riesce a persuadere e coinvolgere a tal punto i soggetti è appunto la capacità di appigliarsi alle emozioni, di sfruttarle per prenderne il comando fino

⁵⁸ G. Riva, *Fake news. Vivere e sopravvivere in un mondo post-verità*, Il Mulino, Bologna, 2018, cit. p.18.

⁵⁹ Ivi, p. 19.

in fondo, fino ad essere in grado di direzionarle. L'emozione condiziona anche la capacità razionale di assumere per vera una realtà che non lo è.

Dal punto di vista cronologico, Riva afferma che secondo diverse opinioni le fake news nascono tra il 2015 e il 2016, sullo sfondo politico delle elezioni americane.

Nello scontro tra Hilary Clinton e Donald Trump la Clinton aveva ricevuto un numero maggiore di voti, e tuttavia chi vinse le elezioni americane furono vinti da Trump. Nel 2016 nessuno poté prevedere questa vincita e, anzi, i sondaggi al tempo davano per vincitrice la Clinton. Da questo evento si iniziò a parlare della circolazione di false notizie e di come ci fosse stato un condizionamento esterno verso gli elettori per direzionare le loro scelte e far perdere progressivamente fiducia nei confronti delle istituzioni, generando una campagna mediatica di disinformazione⁶⁰. Questa strategia di disinformazione, che segue delle linee ben precise ed articolate, comprende nella sua parte più innovativa l'uso dei social media, che a sua volta spesso comprende un vero e proprio lavoro di strategia e ingegneria comunicativa e sociale, un'assoluta novità rispetto al passato.

L'effetto è quello esposto precedentemente: ci troviamo davanti ad un fenomeno in grado di impattare sui soggetti singoli e su interi gruppi sociali con un'estrema velocità e capacità di condizionamento (nelle scelte e nella ricondivisione), un fenomeno mai visto prima. Come è possibile?

La risposta si trova nello sviluppo dei media digitali e in come essi abbiano prepotentemente modificato le nostre abitudini e le nostre capacità comunicative. Al giorno d'oggi, infatti, non solo gli utenti passano la maggior parte del tempo usando internet, ma nello specifico l'uso preponderante è quello dei social network.

Si parla in questo caso di vero e proprio "processo di trasformazione sociale"⁶¹: non solo cambiano le percezioni sulla realtà in cui viviamo, ma cambia anche la nostra stessa identità.

Per quanto riguarda quest'ultimo punto, Riva ci spiega più precisamente che:

La nostra soggettività si sviluppa attraverso la consapevolezza e l'integrazione di ciò che vediamo di noi da dentro (identità personale, cioè la consapevolezza dei tratti fisici e di

⁶⁰ Ivi, pp. 21-22.

⁶¹ Ivi, p. 39.

personalità che ci caratterizzano) e di ciò che vediamo di noi da fuori (identità sociale, cioè la consapevolezza delle nostre caratteristiche in quanto membri dei diversi gruppi che compongono la nostra rete sociale)⁶².

Il soggetto, dunque, costruisce la propria identità anche in base al gruppo sociale di riferimento e alla situazione sociale in cui si trova. Ma con l'uso e lo sviluppo dei social network avviene un passaggio decisivo anche da questo punto di vista:

Fino alla nascita dei social network, le reti sociali che caratterizzavano la nostra vita quotidiana e quelle del mondo digitale erano chiaramente distinte. Con la nascita dei social network questa separazione cade generando un nuovo spazio sociale, che porta a una connessione diretta tra mondo fisico e mondo digitale, indipendentemente dal fatto che noi lo vogliamo o meno⁶³.

Nuovamente, ci troviamo di fronte all'avverarsi della profezia di Baudrillard: se un tempo, con l'iniziale comparsa di internet e della televisione, la distinzione tra mondo reale e mondo virtuale era immediata, al giorno d'oggi la situazione è completamente stravolta ("la televisione ha ucciso la realtà", il cosiddetto "delitto perfetto", che non lascia tracce). Ciò che infatti i soggetti fanno nel mondo fisico è direttamente condizionato nel mondo virtuale e, viceversa, ciò che accade nel mondo virtuale condiziona in maniera decisiva ed immediata quello che accade nel mondo fisico. Non essere parte dei social network è percepito come qualcosa di estraneo ed insolito, sembra quasi un non fare parte effettivamente del mondo reale. Paradossalmente, il distacco dalla realtà avviene per chi non si immerge completamente ed intensamente nella realtà digitale: se un tempo immergersi completamente nel virtuale era visto come una sorta di "alienazione" dalla realtà quotidiana, ora il rapporto è inverso: chi cerca di farlo viene perlopiù etichettato come "non presente fino in fondo" nel mondo reale, nelle news, negli accadimenti quotidiani.

Così come gli effetti della quotidianità e del mondo fisico sono immediatamente presenti in quello digitale, viceversa gli effetti del mondo digitale si presentano direttamente in quello reale, nella realtà "in carne ed ossa". Per questo l'autore ci parla di "interrealtà" come "nuovo spazio sociale ibrido – che mescola il mondo digitale con quello fisico"⁶⁴.

⁶² Ivi, p. 41.

⁶³ Cfr. Riva, Wiederhold e Cipresso 2016a; 2016b, in *Fake news*, p.43.

⁶⁴ G. Riva, *Fake news. Vivere e sopravvivere in un mondo post-verità*, p. 45.

Riva pone un'ulteriore interessante analisi per comprendere il funzionamento delle fake news, che si basa sulla distinzione tra fatti e fatti sociali: le fake news si basano principalmente su questi ultimi. La caratteristica principale dei fatti è l'evidenza: essi, infatti, sono immediatamente tangibili, sono qualcosa che si può cogliere attraverso l'osservazione diretta e l'esperienza⁶⁵. D'altra parte, oltre all'esperienza soggettiva in prima persona, un ulteriore modo per cogliere i fatti è attraverso la comunicazione con altri soggetti o dall'osservazione di questi. Per l'appunto in questo caso definiamo i fatti sociali: quegli eventi "la cui verità non dipende dall'evidenza, ma dall'attività della rete sociale di cui facciamo parte"⁶⁶. La caratteristica è il potere di coercizione nei confronti degli individui e il fatto che le caratteristiche vengano condizionate dalla rete sociale in cui il soggetto è inserito. Questo, secondo Riva, dipende dalla necessità della maggior parte dei soggetti di sentirsi parte di un gruppo, in quanto da questo ne dipende l'identità sociale. Egli ci espone, infatti, come anche nel prendere le decisioni spesso i soggetti si facciano condizionare dalla pressione sociale del gruppo in cui sono inseriti⁶⁷.

Queste premesse fanno emergere un elemento essenziale per rendere le fake news più credibili e persuasive: farle diventare dei fatti sociali e fare in modo che vengano supportate dal maggior numero possibile di soggetti (si parla in questo caso di "riprova sociale")⁶⁸. La persuasività di questo metodo è già stata messa in evidenza, evidenziando il ruolo fondamentale anche in campo politico di questo metodo.

Siamo giunti in questo percorso ad un punto importante. A partire dalla concezione dell'iperrealismo e dell'applicazione di quest'ultimo in campo artistico, abbiamo evidenziato come possa essere semplice ingannare le percezioni di un soggetto e renderlo vittima di inganno. Dall'uso iperrealistico della cera siamo passati all'uso dei media immersivi e dei new media digitali, facendoci guidare dall'impianto teorico di lungimiranti autori quali Wiesing e Baudrillard. Siamo, dunque, giunti all'epoca attuale: immersi in un mondo dominato in maniera preponderante dalla virtualità, in una realtà che ha superato i confini tra

⁶⁵ Ivi, pp. 52-53.

⁶⁶ Ivi, p. 53.

⁶⁷ Ivi, pp. 54-55.

⁶⁸ Ivi, p. 59.

mondo fisico e virtualità, nell'avverarsi della profezia baudrillardiana e nel suo superamento (un mondo post-verità⁶⁹ in cui è possibile ingannare in soggetto attraverso la creazione di fake news).

L'iperrealismo ha raggiunto lo stadio successivo. Il discorso teorico ci ha guidati fino a questo punto, ma i nuovi sviluppi e i nuovi usi non sono ancora terminati.

Il percorso in questa “*realtà post-verità*” continua.

⁶⁹ Si veda a riguardo il saggio di M. Ferraris *Postverità e altri enigmi*, Il Mulino, Bologna, 2017.

Capitolo 2. Deepfake: risvolti negativi nell'uso dell'intelligenza artificiale.

2.1. Introduzione all'argomento: storia del deepfake

Abbiamo visto come la somiglianza, o meglio l'eccesso di somiglianza, possa diventare un problema nel momento in cui il soggetto non riesca più a distinguere ciò che percepisce. Grazie allo sviluppo di nuove tecnologie e al progresso dell'intelligenza artificiale siamo in grado, al giorno d'oggi, di produrre immagini e video apparentemente realistici, così reali da sembrare veri, ma che in realtà non sono altro se non imitazioni e copie create *ex-novo*: immagini iperrealistiche che possono o non avere alcun modello di riferimento (ad esempio, foto di persone create dall'intelligenza artificiale che non esistono fisicamente) oppure usare foto di persone reali, in tutto e per tutto molto simili, ma ricreate e modificate in maniera completamente artificiale. Si parla in questo caso di *deepfake* (“*profondo fake*”).



Fig. 6: progetto *Synthesizing Obama*, pubblicato nel 2017.

Pensiamo ad un video del presidente degli Stati Uniti che dichiara guerra ad un altro stato o che insulta pubblicamente un altro presidente. Pensiamo ad un qualsiasi personaggio

noto e molto potente che esclama parole pericolosamente fraintendibili. Cosa causerebbe nell'opinione pubblica una falsa dichiarazione di questo tipo? Ancora, pensiamo ad una nota attrice che compare su un sito pornografico o di cui vengono divulgate pubblicamente immagini false in atti sessuali. Cosa accadrebbe se una tecnologia così avanzata fosse in grado di permettere la creazione di immagini e video falsi a tal punto? Questo avrebbe rilevanti (e in alcuni casi anche pericolose) conseguenze sulla realtà quotidiana.

L'espressione *deepfake* secondo l'enciclopedia Treccani nasce dalle due espressioni inglesi *deep learning* ("insieme di tecniche che permettono all'Intelligenza artificiale di imparare a riconoscere le forme") e *fake* ("falso, notizia falsa")⁷⁰. Questo neologismo è stato coniato nel 2017 e si riferisce ad una tecnica per la creazione dell'immagine umana basata sull'uso dell'intelligenza artificiale, usata per la combinazione di video ed immagini già esistenti con altre immagini tramite un complesso algoritmo che si avvale di reti generative avversarie⁷¹.

Il termine compare, appunto, per la prima volta nel 2017, quando un utente anonimo di Reddit sotto pseudonimo Deepfake posta diversi video pornografici di celebrità. Al tempo in cui uscirono questi video suscitarono un enorme scandalo, ma si scoprì poco tempo dopo che non si trattava di video reali, bensì creati da un'intelligenza artificiale. Furono di conseguenza cancellati, ma questo fenomeno destò molto scalpore e non fu l'unica occasione in cui si verificò. Con il passare del tempo si migliorarono alcuni *bug* nel sistema in modo tale da rendere questi video e queste immagini sempre più realistici e difficili da distinguere rispetto a quelli reali: le foto delle celebrità venivano, infatti, usate per migliorare le prestazioni dell'intelligenza artificiale⁷². Alcune di esse affermarono di non essere particolarmente toccate da questo scandalo, con la convinzione che i falsi video non avrebbero colpito la propria immagine e che sarebbero state protette dalla propria fama. Tuttavia, si destò un grande scalpore nell'opinione pubblica e iniziarono ad insorgere dei dubbi: cosa sarebbe successo se ad essere toccate fossero state donne sconosciute? Se fosse stata danneggiata la

70

[https://www.treccani.it/vocabolario/deepfake_\(Neologismi\)?fbclid=IwAR2hbOxum5hh4hmd3CkE1qL1fIKu_g8Q931nfi-eb7VQ0dKT3q323ngo6Wsc](https://www.treccani.it/vocabolario/deepfake_(Neologismi)?fbclid=IwAR2hbOxum5hh4hmd3CkE1qL1fIKu_g8Q931nfi-eb7VQ0dKT3q323ngo6Wsc) URL consultato in data 9 agosto 2021.

⁷¹ Delle GAN (create da Ian Goodfellow) si parlerà più approfonditamente nel successivo paragrafo.

⁷² Attraverso l'uso delle Gan (*Generative Adversarial Networks*).

reputazione di persone meno famose e, dunque, meno protette dalla propria fama? Se queste donne fossero state sottoposte a diffamazione attraverso pornografia involontaria o *revenge porn* non sarebbe stato possibile difendersi con tanta facilità. Di conseguenza, iniziò a delinearsi la necessità di sottoporre a legislazione il fenomeno del deepfake e di tutelare la reputazione di possibili vittime.

Nel 2017 venne anche pubblicato un articolo su *Vice* in cui si parlava per la prima volta del fenomeno “deepfake”, definendolo uno dei nuovi media che si stavano diffondendo⁷³. Si denunciava per l'appunto che come per l'uso di “Adobetool” e “Face2Face” che sono in grado di far dire qualunque cosa alle persone o registrare video con tracciamenti facciali, questo nuovo tipo di tecnologia fosse in grado di manipolare e creare video di persone reali in maniera totalmente verosimile e credibile, di mostrare dunque persone che fanno e dicono qualunque cosa che in realtà non hanno detto o fatto⁷⁴ (per l'appunto: anche facendo sesso). Nell'articolo venivano poi denunciati i nomi di tutte le celebrità che erano state sottoposte a questa falsa diffamazione (tra cui, quelle colpite, Daisy Ridley, Gal Gadot mentre ha un rapporto con il fratellastro, Emma Watson, Ariana Grande, Katy Perry, Taylor Swift e Scarlett Johansson), le cui immagini erano molto semplici da trovare attraverso internet e librerie open-source. Questo fenomeno venne definito “preoccupante” e “spaventoso” in quanto, al giorno d'oggi, la facilità con cui trovare online o attraverso Google foto e *selfies* di persone (famosi e non) è sempre maggiore, di conseguenza tutti sono sempre più esposti all'eventuale pericolo di vedere comparire la propria immagine in un tale video *fake porn* (o non). Non è necessario essere programmatori esperti per essere in grado di produrre video simili.

Nel maggio del 2018 vennero creati nuovi esempi di video fake con la stessa tecnologia: comparve un video di Trump il quale offriva al Belgio un accordo sull'argomento “cambiamenti climatici”. Si scoprì che il video era stato creato da un partito politico socialista belga che aveva postato il video su Twitter e Facebook, ma prima che venisse scoperto come falso esso provocò centinaia di commenti che esprimevano indignazione per il fatto che il

⁷³ S. Cole, AI-Assisted Fake Porn Is Here and We're All Fucked, su *Vice*, 11 Dicembre 2017. <https://www.vice.com/en/article/gvdydm/gal-gadot-fake-ai-porn> URL consultato in data 9 agosto 2021.

Presidente si fosse espresso pesando sulla politica climatica del Belgio. I creatori del video ammisero solo in seguito alle reazioni l'inautenticità del video, affermando anche la convinzione che fosse evidente la falsità a causa della scarsa qualità con cui era stato prodotto⁷⁵.



Fig. 7: video di Trump creato grazie all'uso del deepfake.

Arriviamo ad aprile 2018: il regista Jordan Peele elabora un nuovo video destinato a destare scalpore: un finto annuncio pubblico in cui Barack Obama insulta pesantemente Donald Trump, dicendo che egli è “un totale e completo imbecille”. L'intenzione dell'autore è far comprendere la pericolosità di questo strumento, le sue potenzialità ma allo stesso tempo i rischi che ne derivano. Tutto ciò per dimostrare quanto sia facile modificare filmati per far circolare falsi messaggi. Se all'inizio il video sembra un normale annuncio presidenziale, alla fine l'inganno è svelato. Tuttavia, l'attenzione ricade su quanto sia verosomigliante il video. Alla fine, infatti, il regista (che dà anche la voce al presidente nel video) lancia un messaggio pubblico: è necessario evitare che video pericolosi di personaggi politici diventino virali, trattando quindi meglio le informazioni in rete e tutelandone l'autenticità a livello legislativo.

Già nel 2017 era stato creato, attraverso l'uso delle stesse tecnologie (nello specifico Adobe After Effects e FaceApp), un video di Obama da ricercatori dell'università di

⁷⁵ Cfr. articolo scritto da Oscar Schwartz, [You thought fake news was bad? Deep fakes are where truth goes to die](#), in *The Guardian*, 12 November 2018. URL consultato in data 10 agosto 2021.

Washington (progetto “*synthesizing Obama*”). Anche in questo caso si nota la strabiliante capacità della macchina di distorcere voce ed espressioni facciali.

Giungiamo ad agosto 2018: dei ricercatori di un istituto per la ricerca informatica in Germania, tra cui Christian Teobald, producono dei video chiamati “*deep video portraits*” in cui una persona può prendere il controllo del volto in video di un’altra persona, facendole dire ciò che vuole a piacimento (come se si trattasse di una sorta di ventriloquismo). Questo ricercatore coinvolto nello studio afferma che questi “*deep video portraits*” potrebbero essere usati per un doppiaggio (con il dovuto *lip synch* permesso dall’uso di questa intelligenza artificiale) più accurato in film stranieri, per tecniche avanzate di post-produzione nei film e per l’uso di effetti speciali⁷⁶.

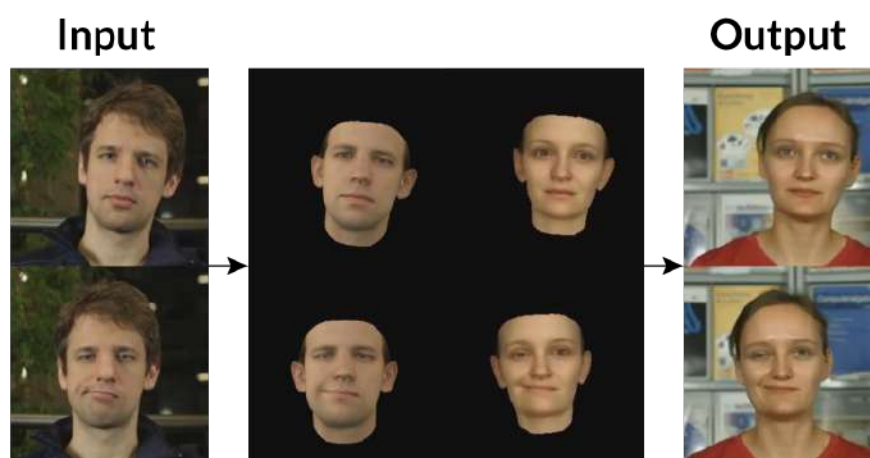


Fig. 8: Deep video portraits, Max Planck Institute for Informatics, Germania, 2018.

In un documento che accompagna la ricerca originale, gli studiosi coinvolti riconoscono l’eventuale uso improprio di questa tecnologia, in grado di produrre volti artificiali che sembrano quasi indistinguibili da quelli reali, ma nonostante questo potenziale affermano anche l’importanza che essi avrebbero potuto avere nell’industria dell’intrattenimento visivo. Grande scoperta, dunque, con un grande potenziale d’uso ma,

⁷⁶ Cfr. articolo scritto da Oscar Schwartz, [You thought fake news was bad? Deep fakes are where truth goes to die](#), in *The Guardian*, 12 November 2018. URL consultato in data 10 agosto 2021.

come tutte le grandi invenzioni, anche con molti rischi derivanti dal possibile uso improprio di questa tecnologia.

Abbiamo elencato alcuni rischi che potrebbero derivare dall'uso errato del *deepfake* (come l'utilizzo in campo politico o il già sopracitato *fake porn*). Alcuni studiosi, a questo proposito, si esprimono con preoccupazione. Hany Farid (professore della University of California di Berkeley) si esprime dicendo che i ricercatori dovrebbero fare più attenzione in quello che stanno affrontando e diffondendo, a livello sia sociale che politico: "The special effects community will love these new technologies, but outside of this world, outside of Hollywood, it is not clear to me that the positive implications outweigh the negative."⁷⁷ Lo stesso Farid, negli ultimi anni di ricerca, si è infatti occupato di creare dei metodi di rilevamento per riconoscere ed eventualmente contestare i deepfake. Sebbene sia difficile produrre software efficaci in questo campo, con l'intensificarsi della diffusione (e in certi casi del pericolo) dei deepfake, si è intensificata parallelamente la ricerca per la creazione di nuovi metodi di rilevamento. Nonostante Farid e molti altri ricercatori siano impegnati in questi studi, egli afferma che il problema non risieda solo nello sviluppo e nella velocità con cui la tecnologia sta progredendo, bensì anche nei cambiamenti e nei processi a livello sociale. La più grande preoccupazione, infatti, sta nel fatto che è sempre più facile al giorno d'oggi per la società credere ad una fake news o ad un deepfake, nonostante la verosomiglianza a volte sia carente (il falso video di Trump in Belgio, in cui la qualità era volutamente carente, ne è una dimostrazione).

Ci troviamo davanti ad uno scenario di disinformazione che può portare a due epiloghi: o la reazione sociale sarà quella di credere a qualunque cosa, oppure il suo opposto, ovvero smettere di credere a qualunque cosa. Questo fenomeno è stato chiamato "*apatia della realtà*"⁷⁸: la più grande minaccia non è semplicemente che le persone vengano ingannate, bensì che arrivino a recepire qualunque cosa come un inganno (con una conseguente sfiducia nei media e nelle istituzioni), portando progressivamente ad un annichilimento e un totale decadimento epistemico. La professoressa di diritto della Boston University Law School

⁷⁷ *Ibidem*. Trad: "La comunità dell'intrattenimento visivo adorerà queste nuove tecnologie ma, al di fuori di questo mondo, al di fuori di Hollywood, non mi è chiaro se le implicazioni positive possano superare quelle negative".

⁷⁸ *Ibidem*.

Danielle Citron, una tra le prime a preoccuparsi del potenziale negativo e dei rischi del deepfake sulla società (sia per quanto riguarda la violazione dei diritti delle donne nell'ambito del fake porn, sia come potenziale pericolo per la democrazia in campo politico), afferma: “When nothing is true then the dishonest person will thrive by saying what’s true is fake.”⁷⁹”

Tutto questo ci fa capire quanto sia importante approfondire lo studio di queste nuove tecnologie, le potenzialità, gli usi positivi ma, al tempo stesso, prevedere anche i possibili rischi e l’impatto che potrebbero avere sulla società e sugli individui, cercando di tutelarli arginando i rischi e controllandoli attraverso l’uso di software o formulando una adeguata legislazione a riguardo. Prima di tutto però, per capire adeguatamente questo fenomeno, è necessario comprenderne più dettagliatamente il meccanismo.

2.2. Gans (Generative Adversarial Networks)

Per comprendere fino in fondo ciò di cui ci stiamo occupando, verrà esposto il fenomeno del deepfake al suo esordio: si procederà parlando di Ian Goodfellow e dell’invenzione delle GANs, l’algoritmo che permette a questa intelligenza artificiale di funzionare.

I fake video possono essere creati attraverso una tecnica di apprendimento automatico chiamata “reti antagoniste generative” / “rete generativa avversaria” (o, usando il termine originale, *generative adversarial network* / *GAN*). Questa tecnica è stata elaborata nel 2014 da un informatico e ricercatore, al tempo studente, Ian Goodfellow, che ha creato i GAN come metodo per generare algoritmicamente nuovi tipi di dati da altri set di dati esistenti.

Ad esempio, una rete generativa avversaria può, a partire da migliaia di foto di Barack Obama, produrre una nuova foto che si avvicini a quelle originali, ma senza essere una copia esatta di nessuna di esse, generando quindi un ritratto completamente nuovo. Questa tecnologia può essere usata non solo nel caso di immagini e foto, ma anche per quanto riguarda video e audio. L’uso di questa tecnica di apprendimento automatico è stato limitato alla comunità scientifica di ricerca sull’intelligenza artificiale per diversi anni, fino a quando

⁷⁹ *Ibidem*, tr.: “Quando nulla è vero, la persona disonesta prospererà dicendo che ciò che è vero è falso”.

(come abbiamo visto precedentemente) nel 2017 non venne pubblicato da un utente di Reddit un fake video pornografico sotto il nome utente “deepfake”.

Nello specifico, le reti generative avversarie si basano su un gioco (inteso nel senso della teoria dei giochi) tra due modelli di apprendimento automatici implementati utilizzando reti neurali⁸⁰. Le due reti neurali vengono addestrate in maniera competitiva: una rete generativa avversaria è, infatti, composta da un modello generativo (*Generator*) e un modello discriminativo (*Discriminator*), realizzati attraverso reti neurali. Il discriminatore ha il compito di classificare correttamente i dati come reali o falsi, di darne una corretta interpretazione, mentre lo scopo del generatore è quello di produrre nuovi dati.

Il generatore cercherà di “ingannare” il discriminatore, cercando di produrre campioni che il discriminatore classificherà erroneamente come reali: il modello discriminativo dovrà cercare, dunque, di apprendere come distinguere i dati reali da quelli generati artificialmente.

Prima di addestrare la rete antagonista generativa, il discriminatore viene addestrato, in maniera separata, a riconoscere volti reali con sufficiente precisione. Il suo scopo sarà distinguere tra immagine reale e immagine artificiale, riconoscendo dunque i falsi e “battendo” il generatore. D’altra parte, il generatore prenderà come input un rumore casuale e dovrà cercare di generare immagini realistiche, cercando a sua volta di “ingannare” il discriminatore (creando volti che siano talmente realistici da essere indistinguibili rispetto a quelli reali). Nella rete generativa avversaria le immagini del generatore saranno l’input del discriminatore; quando il discriminatore riceverà l’immagine la confronterà con altre immagini ricevute in precedenza e darà un giudizio (reale/artificiale). Se il giudizio sarà errato (volto artificiale scambiato per vero), esso apprenderà da questa risposta sbagliata, migliorando quelle successive; se invece la risposta è corretta (volto artificiale riconosciuto come artificiale), questa risposta aiuterà il generatore a migliorare le successive immagini artificiali. Questo processo continuerà finché le immagini generate e i giudizi del discriminatore non raggiungeranno un punto di equilibrio. Generatore e discriminatore (reti neurali) in questo modo competono addestrandosi e “migliorandosi” a vicenda.

⁸⁰ Cfr. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, *Generative Adversarial Networks*, su *Communication of the Acm*, vol. 63, n.11, Novembre 2020, p. 140.

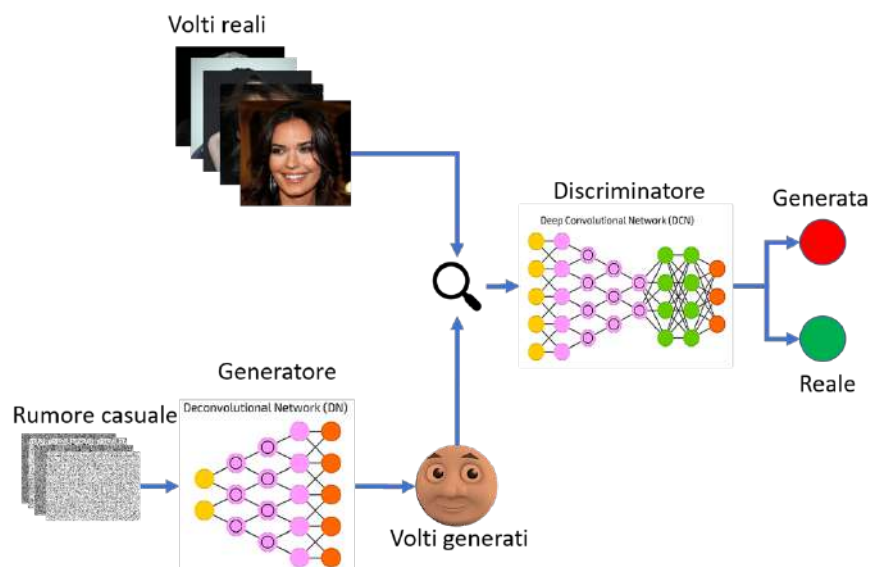


Fig. 9: GAN, schema di funzionamento nell'elaborazione di immagini.

Possiamo pensare ai GAN, in maniera più intuitiva, come ad un gioco tra falsario (generatore) e poliziotto (discriminatore): i falsari producono soldi falsi, mentre la polizia tenta di arrestare i falsari, ma facendo questo, nel frattempo, consentirà la produzione e la circolazione di denaro falso. La competizione tra falsari e polizia porterà così alla produzione di denaro falso sempre più realistico, perché i falsari avranno tempo e modo di “allenarsi” migliorando la produzione.

Arrivati a tal punto, i contraffattori, finché non verranno catturati, riusciranno a produrre falsi perfetti, in modo tale che la polizia non sarà più in grado di distinguere tra denaro reale e falso.

Un'ulteriore complicazione a questa metafora: è come se il generatore apprendesse attraverso il gradiente del discriminatore, ovvero, come se i falsari avessero una “talpa” tra la polizia che segnalasse gli specifici metodi utilizzati dalla polizia per individuare i falsi⁸¹ (di nuovo, consentendo di migliorare la produzione di denaro falso rendendolo sempre più verosimile all'originale e quindi ingannevole, difficile da distinguere).

⁸¹ Ivi, p. 141.

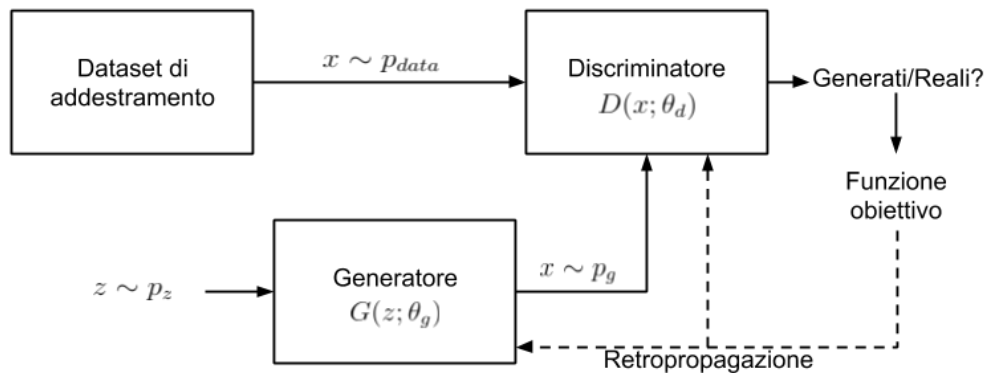


Fig. 10: schema grafico della rete generativa avversaria.

L’obiettivo di questo algoritmo, come spiegato in maniera più tecnica e specifica nello stesso documento elaborato da Goodfellow, è trovare una sorta di “*equilibrio locale di Nash*”⁸²: un punto in cui ci sia un minimo locale del costo di ogni giocatore rispetto ai parametri di quel giocatore. Con le mosse locali, nessun giocatore può ridurre ulteriormente il suo costo, assumendo che i parametri dell’altro giocatore non cambino.

L’algoritmo di allenamento più comune è quello di utilizzare un “ottimizzatore a gradiente” che esegua ripetutamente passi simultanei su entrambi i giocatori, minimizzando in modo incrementale il costo di ogni giocatore rispetto ai parametri di quel giocatore. Alla fine del processo di formazione, i GAN sono spesso in grado di produrre campioni realistici, anche per dati molto complicati contenenti immagini ad alta risoluzione.

L’unico errore reale è l’errore statistico (campionamento di una quantità finita di dati di formazione piuttosto che misurare la vera distribuzione di dati generativi sottostante) e il fallimento dell’algoritmo di apprendimento per convergere ai parametri ottimali. Molte

⁸² “The goal of a machine learning algorithm in this context is to find a *local Nash equilibrium*”, Cfr, I. Goodfellow, *Generative Adversarial Networks*, cit. p. 142.

strategie di modellazione generativa introdurrebbero queste fonti di errore e di approssimazione per l'ottimizzazione dei limiti sul costo del reale.

Risulta difficile dare ulteriori indicazioni riguardo i dettagli delle GAN, perché esse sono un'area di ricerca ancora attiva e indicazioni più specifiche possono diventare presto obsolete. Dalla loro scoperta fino al giorno d'oggi, infatti, esse hanno già raggiunto straordinari miglioramenti:

GAN PROGRESS ON FACE GENERATION

Source: Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020; AI Index, 2021



Figure 2.1.7

Fig. 11: miglioramento delle GAN nella produzione di immagini dal 2014 al 2020.

In conclusione, possiamo dire di aver appreso che: i GAN sono una sorta di modello generativo basato sulla teoria dei giochi, che hanno sempre avuto molto successo a livello pratico nel generare dati realistici, specialmente nella produzione di immagini. Non è ancora finito il lavoro di “addestramento” di queste reti neurali, ma affinché diventino una tecnica più affidabile sarà necessario progettare modelli, costi, algoritmi grazie ai quali sarà possibile trovare dei buoni equilibri di Nash in modo coerente e rapido.

Le applicazioni delle GAN sono state utilizzate in diversi campi come: sintesi di immagini a partire da altre immagini, sintesi di immagini a partire da descrizioni testuali, generazione di video in sincronizzazione labiale, restauro di video e immagini.

Possiamo elencare ora, alcuni esempi concreti del funzionamento di queste reti generative avversarie, per vedere il risultato del loro addestramento nell’elaborazione di immagini e fotografie.



Figure 3. Example results by our proposed StackGAN, GAWWN [20], and GAN-INT-CLS [22] conditioned on text descriptions from CUB test set. GAWWN and GAN-INT-CLS generate 16 images for each text description, respectively. We select the best one for each of them to compare with our StackGAN.

Fig. 12: esempio di Gan per l’uso “Text to image” (immagine presa da <https://www.spindox.it/it/blog/rete-antagonista-generativa/>)

2.2.1. Un esempio concreto dell’uso delle GAN: thispersondoesnotexist

Siamo nel 2019, Stati Uniti. Un ingegnere informatico, Philip Wang, attraverso l’uso delle GANs, insegna ad un’intelligenza artificiale a ricreare i volti delle persone, rendendoli disponibili online. L’autore si basa su una ricerca pubblicata da Nvidia, la quale spiega in che modo sia possibile creare un’infinità di ritratti non esistenti attraverso l’apprendimento

dell'intelligenza artificiale⁸³. Il sito si chiama thispersondoesnotexist.com⁸⁴ (ovvero, “questa persona non esiste”) e funziona in maniera intuitiva: basta ricaricare la pagina per riprodurre ogni volta un nuovo ritratto, in tutto e per tutto iperreale e simile a volti umani, ma con alcuni dettagli (errori a volte più, a volte meno evidenti) che dopo diverse volte in cui la pagina viene ricaricata possono iniziare a saltare all'occhio. Il sito funziona, per l'appunto, attraverso un algoritmo che ha imparato a ricreare i volti attingendo da archivi di foto che si possono trovare online (come abbiamo già visto le GANs attingono tra set di immagini o altro materiale preesistente).

Il grande impatto che ha avuto la creazione di questo sito era dovuto al fatto che esso fosse accessibile anche ad utenti comuni, dunque ad una maggioranza di persone, che sono riuscite in tal modo ad avere un'idea di quanto elevate e sviluppate possano essere le abilità di un'intelligenza artificiale. Solitamente, infatti, questo tipo di tecnologie non è accessibile al grande pubblico: non è ancora stata data sufficiente prova di quanti progressi siano stati fatti in questo campo negli ultimi dieci anni. D'altra parte, maggiore è la diffusione, maggiore diventa anche il rischio (cresce esponenzialmente man mano che crescono il potenziale e la diffusione di questi mezzi): sarà possibile tra qualche anno distinguere tra un'immagine reale e una artificiale?

Il sito attinge il materiale a partire da cui crea i volti fittizi da un numero non definito di foto reali di Flickr raccolte da archivi ad accesso libero di internet, che nel giro di una frazione di secondo vengono assemblate attraverso l'algoritmo creando ritratti altrettanto iperrealistici, dando un'apparenza di verità. Tutto questo avviene appunto cliccando il pulsante “refresh” dell'intelligenza artificiale. Come altre immagini che vengono assemblate usando degli algoritmi di intelligenza artificiale chiamati GANs, questo sito è conosciuto come generatore di “Deepfakes”. Non si tratta dell'unico esempio per farci realizzare quanto

⁸³ Cfr. a riguardo, Raffaele Angius, *Un'intelligenza artificiale ha imparato a creare dei perfetti volti di persone che non esistono*, pubblicato su *La Stampa* il 20 febbraio 2019. <https://www.lastampa.it/tecnologia/news/2019/02/20/news/un-intelligenza-artificiale-ha-imparato-a-creare-dei-perfetti-volti-di-persone-che-non-esistono-1.33682515> URL consultato in data 13 agosto 2021.

⁸⁴ Cfr. a riguardo, articolo che spiega il funzionamento tecnico del sito: Tero Karras, Samuli Laine, Timo Aila, *A Style-Based Generator Architecture for Generative Adversarial Networks*, arXiv:1812.04948v3 [cs.NE] 29 Mar 2019, consultato il 13 agosto 2021.

stia progredendo l'uso in questo campo: come questo sito, infatti, possiamo trovare esempi simili basati sui funzionamenti algoritmici delle GANs, come ad esempio: *thisAirBnBdoesnotexist*, *thesecatsdonotexist*, e *thisstartupdoesnotexist*. Anche in questi casi il funzionamento è il medesimo, con la produzione di oggetti differenti. Nel primo caso ci si presenterà la schermata di un AirBnB molto simile ad una originale (con tanto di foto, persone da contattare, descrizioni testuali), nel secondo immagini di animali (a differenza dei volti di persone umane, come in *thispersondoesnotexist*), nell'ultimo assisteremo addirittura alla creazione artificiale della pagina di una finta startup (con tanto di links, possibilità di condividere la pagina, foto di creatori e clienti fittizi, contatti ecc.), in tutto e per tutto verosimile e ricreabile allo stesso modo, attraverso il bottone refresh. Ci sono, dunque, molte applicazioni nel campo dell'I.A. per creare immagini, video e non solo, realizzandoli in modo apparentemente realistico e utilizzando archivi di immagini a volte indistinguibili da foto e video veri e propri, di soggetti reali. L'algoritmo possiede anche delle capacità di "deep learning": questo consente la capacità di correggere gli errori, confrontando le immagini con archivi originali e permettendo così di raggiungere un realismo sempre maggiore e privo di errori, una naturalezza sempre più ingannevole⁸⁵.

Facciamo l'esempio della fotografia. In questo campo, ancora molto tempo prima del machine learning e della creazione di Deepfake, le domande sul valore, il realismo e la verità sono sempre state poste. I primi risultati in campo fotografico, prima del suo sviluppo, non erano tra i migliori: le prime fotografie, infatti, riportavano delle immagini che erano capovolte al lato e non riuscivano a rappresentare adeguatamente la realtà. Inizialmente, dunque, questo mezzo rendeva visibile solo una parvenza di realtà, ma non possiamo dire che esprimesse in maniera fedele un ritratto del mondo così com'era, iperrealisticamente parlando.

La questione è stata sollevata nel momento in cui la fotografia è stata digitalizzata: da questo momento, infatti, sono stati sollevati dibattiti ("da quando il mezzo fotografico è stato digitalizzato, una definizione fissa del termine fotografia è diventata impossibile"⁸⁶). Questa

⁸⁵ Cfr. Kris Belden-Adams, *ThisTheoryDoesNotExist: Historicizing and Understanding Artificial-Intelligence-Generated, Hyper-Real, Photographic Data Visualizations*, University of Mississippi, USA, Publishing House EIDOS, 2019, p. 138.

⁸⁶ Intervista ad Andreas Gursky, 1998, Cfr. *ThisTheoryDoesNotExist*, cit. p. 141.

dichiarazione di “crisi della fotografia” ha suscitato un eco nel corso degli anni, acuendosi tra gli anni '80 e i 2000 (arrivando a dare etichette come “post-fotografia” o “condizione post-media”). Spesso tutte queste etichette si sono basate sulla nascita dei software di manipolazione digitale, che hanno fatto in modo che il mezzo perdesse il contatto diretto e il rapporto con la realtà.

Se l'uso dell'intelligenza artificiale per creare fotografie è recente, vi erano già stati in passato delle evoluzioni in questo campo (come, ad esempio, per quanto riguarda alcuni strumenti di Photoshop, che è una sorta di “precursore” di questa logica). Per quanto riguarda, invece, la fotografia delle immagini basata sull'I.A., esiste semplicemente una separazione tra la natura complicata dei ritratti di dati generati automaticamente e quella di ritratti che possono essere prodotti indipendentemente dal computer, seguendo le direttive fornite dall'uomo.

I Deepfake sono un fenomeno così nuovo da dover ancora entrare negli archivi di storia della fotografia, in ambito accademico. Tuttavia, ora come ora non ci troviamo ancora davanti a ritratti generati in modo impeccabile e devono ancora essere studiati ed approfonditi nel campo del deep learning.

Sono state fatte anche alcune ricerche per mettere in evidenza i difetti apparenti che emergono dalle fotografie prese dal sito *thispersondoesnotexist* (a partire dal 2019)⁸⁷. Il Deepfake fornisce un esempio della capacità unica di questo mezzo di rappresentare quella retorica dell'iperrealismo che però non diventa mai reale. Difficilmente è possibile fornire un'evidenza dell'identità di questo mezzo, una conferma di quanto questa “verità fotografica” sia ingannevole. Proprio per questo il fenomeno dei Deepfakes apre la strada a nuovi dibattiti in campo etico e la necessità di regolamentare e porre dei limiti all'uso dell'I.A., strumento tanto potente quanto potenzialmente rischioso: campo nuovo e che necessita urgentemente di un approfondito dibattito che include diversi campi: non solo sviluppo tecnologico ma anche campi come politica, giornalismo, programmazione informatica, fotografia e discipline umanistiche. Vedremo più approfonditamente ora perché

⁸⁷ Cfr. Kris Belden-Adams, *This Theory Does Not Exist: Historicizing and Understanding Artificial-Intelligence-Generated, Hyper-Real, Photographic Data Visualizations*, University of Mississippi, USA, Publishing House EIDOS, 2019, pp. 143-147.

è necessario questo dibattito e, dunque, come si configurano più dettagliatamente le criticità e i rischi nell'uso di questo strumento.



Fig. 13: esempi di foto generate artificialmente attraverso l'uso delle GANs, *thispersondoesnotexist*

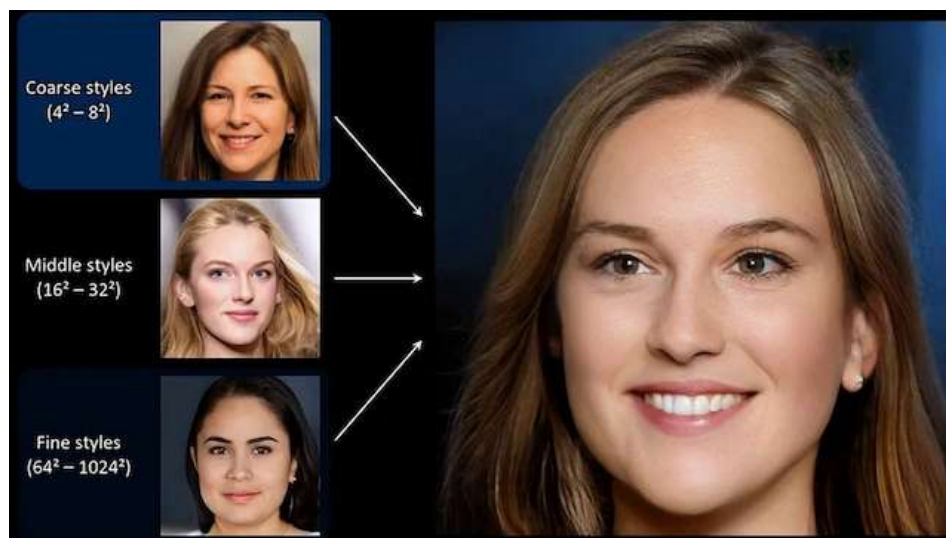


Fig. 14: esempio del funzionamento di *thispersondoesnotexist* (immagine presa da <https://www.firstpost.com/tag/this-person-does-not-exist>)

2.3. Problemi e criticità

Abbiamo già messo in evidenza alcune criticità per quanto riguarda l'uso dei Deepfake nell'introduzione sulla loro storia e il riguardante sviluppo a livello cronologico. Alcuni esperti hanno preso apertamente posizione contro l'uso di queste tecnologie, considerate pericolose sotto più punti di vista e in diversi campi, se non regolamentate e limitate da un'adeguata legislazione. Nel paragrafo precedente, infatti, sono stati nominati esperti di giurisprudenza, come Danielle Citron, e di informatica, come Hany Farid, i quali hanno effettuato opportune ricerche e cercato dei metodi per tutelare i soggetti colpiti dall'uso errato di questi nuovi mezzi digitali sia a livello legislativo, di dibattito oppure di vero e proprio, sia in campo tecnologico, nello sviluppo di software per riconoscere l'autenticità dei deepfake.

Nell'articolo *The Emergence of Deepfake Technology* di M. Westerlund i deepfake vengono addirittura definiti “una grave minaccia per la nostra società, il sistema politico e le imprese”⁸⁸ per più motivi. Tra questi emergono il mettere sotto pressione i giornalisti che già lottano per smascherare le fake news, la minaccia per la sicurezza nazionale (con la diffusione della propaganda e l'interferire nelle elezioni), l'ostacolare la fiducia dei cittadini nei confronti dell'autorità e delle informazioni da parte delle autorità, i problemi di gestione della sicurezza informatica per persone ed organizzazioni.

Anche nel caso delle fake news, come già stato messo in evidenza, il giornalismo ha dovuto subire un calo della fiducia da parte dei cittadini (maggiore è la diffusione, maggiore sarà la sfiducia a credere anche a notizie autentiche, fenomeno che è stato chiamato “apatia del reale”). Nel caso dei deepfake la minaccia è ancora maggiore, duplicata: essi rappresentano infatti una minaccia più delle fake news in quanto più difficili da individuare e più efficaci a livello di manipolazione (lo spettatore fatica ancor più a distinguere un prodotto vero da uno falso).

⁸⁸ Cfr. Westerlund M., *The emergence of deep fake technology: a review*, Technology Innovation Management Review, 9(11) : 40-53.

La comunità dell'intelligence è preoccupata del fatto che i falsi video possano essere usati per minacciare la sicurezza nazionale con propaganda politica o interruzione delle campagne elettorali⁸⁹. Soprattutto in un caso come quello delle elezioni (che abbiamo analizzato anche in relazione alle fake news), far dire a qualcuno false dichiarazioni in un video virale può diventare un'arma potente per la disinformazione e la manipolazione nelle scelte: i video manipolati possono agire in maniera rilevante sulle decisioni degli elettori.

Di conseguenza, traspare come l'aspetto più pericoloso dei deepfake non sia solo la disinformazione, ma tutto ciò che questa porta: il costante contatto dei cittadini li porterà a credere che la maggior parte delle informazioni e dei video non sia attendibile, arrivando ad epiloghi come la già nominata "apatia della realtà" o ad una "apocalisse dell'informazione"⁹⁰. La sfiducia nelle istituzioni e le teorie di cospirazione vengono alimentate da tutto questo: la minaccia maggiore non sarà che le persone vengano ingannate, bensì che arrivino a percepire qualunque cosa come un inganno.

Un'altra minaccia emerge nei problemi di sicurezza informatica: in campo aziendale ha la necessità di proteggersi da eventuali frodi virali. Anche in questo campo l'uso del deepfake potrebbe essere mirato ad influenzare l'opinione pubblica, mostrando ad esempio falsi video di amministratori che affermano false dichiarazioni (magari razziste, omofobe) per manipolare mercato e far perdere in azioni. Anche nel caso del già citato fake porn si potrebbero far circolare video per sabotare qualcuno, ricattarlo, screditarlo (e, come abbiamo già messo in evidenza, nei casi in cui non si tratti di celebrità, facilmente dichiarabili come falsi, sarebbe più difficile difendersi e dichiarare apertamente l'inganno). Sempre in ambito aziendale un falso video potrebbe essere utilizzato per ricevere informazioni private, confidenziali, non trasmissibili. È stato, inoltre, scoperto di recente un sito web per deepfake con lo scopo di estorcere criptovalute⁹¹.

Abbiamo visto, dunque, il lato ludico dei deepfake, che è rappresentato dalla maggior parte di quelli che circolano sulle maggiori piattaforme come Facebook e Youtube, creati per intrattenimento o talvolta anche a scopo artistico, ma esiste anche il lato oscuro di questo

⁸⁹ *Ibidem*.

⁹¹ Cfr. Westerlund M., *The emergence of deep fake technology: a review*, Technology Innovation Management Review, 9(11) : 40-53.

fenomeno, che è quello della diffusione di video di fake porn per vendetta o ricatto, o la manipolazione in campo politico ed elettorale.

Ad essere stati colpiti dalla manipolazione digitale di video non sono stati solo i soggetti politici, ma anche personaggi famosi non in ambito politico. Ad esempio, nel giugno 2019 venne pubblicato un video di due artisti inglesi con Mark Zuckerberg, CEO di Facebook, in cui egli affermava di aver creato un'organizzazione capace di prendere il controllo di miliardi di dati di utenti in modo tale da manipolarne il futuro⁹². Ovviamente il video era falso, ma è l'ennesima riprova di come la tecnologia abbia compiuto passi enormi e lo sviluppo del deepfake possa rappresentare un pericolo nella manipolazione di dati, in diversi campi.

2.3.1. Applicazioni a livello politico e fake news

In ambito politico sono già stati citati in precedenza i due famosi esempi di deepfake che riguardano i presidenti americani Obama e Trump, ma non sono gli unici casi in cui i deepfake e la manipolazione di video hanno condizionato l'ambito politico, prendendo di mira personaggi di spicco della scena politica americana e non solo. Vi è, infatti, il caso di Nancy Pelosi: la politica americana nel maggio del 2019 fu schernita attraverso un video in cui si presentava ubriaca durante un discorso pubblico. Il video divenne addirittura virale, venendo condiviso anche dallo stesso Donald Trump. Si scoprì successivamente che si trattava di un video manipolato.

Una volta capita l'enorme diffusione di questo fenomeno, il Senato statunitense, attraverso alcuni politici del partito americano, cercò di avvisare il Governo stesso sul rischio di condizionare le elezioni presidenziali stesse: cosa sarebbe successo se un video fosse diventato virale il giorno prima delle elezioni, per poi venire dichiarato come falso solo successivamente?

⁹² Cfr. articolo di Arcangelo Rociola, *Il finto video in cui Zuckerberg dice di avere il controllo delle nostre vite*, pubblicato su *AGI-Agenzia Italia*, il 12 giugno 2019, https://www.agi.it/economia/zuckerberg_facebook_deep_fake-5647497/news/2019-06-12/ URL consultato in data 17 agosto 2021.

Oltre agli esempi già citati di deepfake in campo politico, possiamo citarne un altro di significativo impatto: nel 2018, in Africa centrale, Ali Bongo (presidente del Gabon) si era allontanato dal paese ed era scomparso dalla scena pubblica per ricevere delle cure mediche. Gli abitanti del paese stavano diventando sempre più sospettosi riguardo alle condizioni di salute del presidente, il governo non rispose ed alcuni dissero che era morto. Seguì un'ondata di disinformazione e l'annuncio che il presidente avesse avuto un ictus, ma fosse sopravvissuto. Non vennero però fatte apparizioni pubbliche, venne rilasciato tempo dopo un video che lasciava ampio spazio alle preoccupazioni dei cittadini sulla salute del presidente. Seguì un colpo di stato, a causa della stranezza del video come prova che qualcosa, nel presidente, non andava. Bruno Ben Moubamba, politico del Gabon, sostenne la falsità del video, affermando che si trattasse di un deepfake.

Questo fatto ci fa capire come non solo nei paesi occidentali, ma soprattutto nei paesi in via di sviluppo si possa presentare questo pericolo (in cui vi sono governi più fragili e i cittadini possiedono una minore informazione a livello di cultura digitale)⁹³.

La politica italiana non è rimasta immune⁹⁴: nel 2019 circolò un video dell'ex premier Matteo Renzi che, dopo la scissione dal Partito Democratico, sbeffeggiava alcuni membri della scena politica italiana come l'attuale presidente del Consiglio Conte, il segretario del Pd Zingaretti, il leader del M5S Di Maio e addirittura il presidente della Repubblica Mattarella. I contenuti del video erano così verosimili da far risultare perfettamente l'inganno: parlata quasi perfetta e accento fiorentino molto simile a quello originale dell'ex premier lasciavano poco spazio alla percezione di essere vittime di un inganno. Le reazioni si susseguirono, infatti, sui social, poiché data l'incredibile somiglianza il video venne preso per vero. Tuttavia, esso venne subito smentito dal protagonista Renzi, che lo accolse con una

⁹³ Cfr. Articolo di Ali Breland, *The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink*, pubblicato su *Mother Jones*, 15 marzo 2019, <https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/> URL consultato in data 21 agosto 2021.

⁹⁴ Cfr. articolo di Massimo Gramellini, *Profondo Falso*, in *Corriere della Sera*, 26 settembre 2019. https://www.corriere.it/caffe-gramellini/19_settembre_26/profondo-falso-e1d251c8-dfcf-11e9-aa5f-fbca0c81b7e9.shtml URL consultato il 18 agosto 2021.

risata dichiarandolo semplicemente “Una perfetta imitazione”. D’altra parte, nonostante l’uso nel caso citato a scopo ludico, all’introduzione di questo fenomeno in Italia non seguì un’adeguata informazione sui potenziali rischi di questa tecnologia. Il Senato della Repubblica italiana si esprime sull’argomento in un discorso di dicembre 2019 parlando appunto della minaccia del deepfake e come affrontarla. All’interno di questo discorso, infatti, pur riconoscendo gli enormi progressi e i vantaggi portati da Internet e dallo sviluppo tecnologico, si evidenzia però anche la necessità di riconoscere le minacce e i pericoli che si annidano nella realtà virtuale, basati su un “uso distorto” di questo mezzo. L’enorme quantità di dati che si trovano in rete può essere usata sia per scopi legali ed innovativi, ma anche per manipolare e distorcere quella che è la realtà, per trasmettere messaggi ingannevoli e per contraffare dati, con un impatto sociale di portata notevole. Nel documento si legge:

I deepfake, cui è dedicato appunto questo convegno, altro non sono se non il prodotto finale della contaminazione tra Intelligenza Artificiale e dati sensibili utilizzata a fini illeciti. Una nuova e assai temibile minaccia che, nel prossimo futuro, rischia di trasformare l’ecosistema digitale in un mondo nel quale riuscire a distinguere ciò che è vero da ciò che è falso risulterà sempre più arduo. Manipolare al computer, partendo da semplici fotografie e video, i volti, la voce e i movimenti delle persone per creare narrazioni visive molto realistiche con un’aderenza totale delle parole al labiale: le possibili conseguenze di un uso illegittimo della tecnologia deepfake sono angosciose, se pensiamo alla possibilità - già in effetti sperimentata - di creare falsi video in cui politici o personaggi pubblici fanno delle affermazioni in grado di cagionare conseguenze di un certo peso.⁹⁵

Questa tecnica viene inoltre definita “subdola”, in quanto in grado di ingannare e difficile da smascherare, in grado di trasformarsi in un’arma pericolosa e capace di veicolare informazioni sbagliate, condizionando l’opinione pubblica. Nonostante le fake news siano sempre circolate, se applicate ai video e fatte circolare attraverso questo mezzo, diventa ancora più pericoloso per la maggioranza delle persone distinguerne la veridicità, in quanto il video è considerato un mezzo, normalmente, che garantisce una prova autorevole di fatti.

Anche in questo caso emerge il concetto di “collasso della realtà”, affermando che in un’epoca già definita della “post-verità” e già condizionata dalla circolazione delle fake

⁹⁵ Cfr. *La minaccia del deep fake: come affrontarla in Italia?* Discorso pronunciato nella Sala Koch di Palazzo Madama il 9 dicembre 2019, http://www.senato.it/4171?atto_presidente=4701 URL consultato in data 18 agosto 2021.

news, la diffusione di deepfake perfettamente credibili e fatti circolare in maniera rapida e spesso incontrollata attraverso la rete, potrebbe portare ad un irreversibile “collasso della realtà”⁹⁶. Questo fenomeno viene dichiarato pericoloso per persone ed enti che possono essere pubblicamente screditati da nemici e malintenzionati, ma anche per il potenziale criminale e la caduta della corretta informazione digitale, con potenziali gravi conseguenze per la resistenza dei sistemi democratici. Si rende quindi imminente la necessità di regolamentare dal punto di vista legislativo, normativo e per collaborare sull’alfabetizzazione informatica, dunque in campo educativo soprattutto verso i giovani, per promuovere un uso responsabile del web.

2.3.2. Fake porn

Un ulteriore ambito in cui si è discusso approfonditamente ed è emersa la necessità di affrontare l’uso del deepfake e quello del fake porn, ovvero della creazione di deepfakes sessuali non consensuali.

I primi video comparsi in questo ambito in rete furono quelli postati da un utente Reddit alla fine del 2017, in modo anonimo. Questi clip erano stati creati dall’utente stesso, autonomamente da casa, cercando nel web migliaia di foto di celebrità, tra i più grandi nomi di Hollywood, e addestrando gli algoritmi dell’Intelligenza Artificiale per permettere la sostituzione di volti di pornstar con volti di celebrità. I risultati erano stati così convincenti e spaventosi da attirare migliaia di utenti nella pagina “Deepfake” di Reddit, per esaminare personalmente i video. Inoltre, quando venne rilasciata un’app per creare questi video molto più facilmente anche per utenti non esperti di informatica e programmazione, emersero altre decine di nuovi video e, secondo lo sviluppatore dell’app, essa venne scaricata più di 100.000 volte dalla sua creazione (ovvero da inizio febbraio di quell’anno). Da quel momento molte piattaforme (tra cui Reddit) hanno vietato la pubblicazione e la distribuzione di Deepfake, ma non sono mancati altri luoghi in cui gli utenti potessero scambiarsi questi video.

⁹⁶ *Ibidem.*

Dopo questo episodio diverse altre celebrità sono state colpite, tra cui nel 2011 l'attrice Scarlett Johansson, le cui foto sono state rubate e postate online da un hacker, condannato poi a 10 anni di carcere. L'attrice dichiarò poi in seguito a questo episodio che, sebbene la creazione di video e immagini pornografiche false fosse già in atto da decenni, l'aumento dei video deepfake rappresentava una vera e propria minaccia per le donne, che potevano essere umiliate dentro e fuori dai riflettori⁹⁷.

Nell'articolo *Politics and porn: how news media characterizes problems presented by deepfakes* la questione è stata affrontata sotto due punti di vista: le risposte delle aziende di social media con questi contenuti e i danni che questi comportano per quanto riguarda le donne⁹⁸. Gran parte dei contenuti di questo tipo di deepfake si trovano su siti come Pornhub, Reddit e Tumblr, tutti gli altri invece hanno annunciato il loro impegno per rimuoverli e bandirli dai loro siti. Con la divulgazione di questi video è emersa la consapevolezza di quanto sia un problema centrale nell'ambito della tecnologia dei deepfake, ma il punto centrale è far emergere ai lettori le risposte delle compagnie di social media manager. Riassumendo in generale, possiamo dire che ci sono stati tre tipologie di danni principali nelle donne colpite da questo fenomeno: angoscia emotiva e psicologica, perdita di autonomia del proprio corpo e della propria reputazione, connessione tra deepfake e altri possibili crimini. Mentre c'è già stata una discussione sui danni emotivi e psicologici causati dai deepfake, spiegando che questo tipo di fenomeni può lasciare il soggetto in uno stato di angoscia ed umiliazione, non si è parlato molto dei casi di cybermisoginia, ovvero la distribuzione di immagini intime non consensuali. Per esempio, alcune vittime affermano che il periodo seguito alla distribuzione di immagini intime non consensuali le ha portate a vivere in uno stato di angoscia e preoccupazione costante. La maggior parte di questi fake video sono stati creati senza il consenso delle donne presenti, ma bisogna sottolineare che le donne prese di mira non erano soltanto quelle i cui volti sono stati inseriti, ma anche i corpi presenti che sono stati usati in maniera non consensuale.

⁹⁷ Cfr. articolo pubblicato da Drew Harwell, *Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*, in *The Washington Post*, 31 dicembre 2018, URL visitato in data 18 agosto 2021.

⁹⁸ Chandell E. Gosse, Jacquelyn Burkell, *Politics and porn: how news media characterizes problems presented by deepfakes*, in *Critical Studies in Media Communication*, FIMS Publications, 2020.

La maggior parte del focus è stata rivolta ai fake video di celebrità, come nei casi già citati di Emma Watson e Gal Gadot: in questi casi la preoccupazione è rivolta alla preoccupazione per la loro immagine pubblica e la reputazione a livello di fama internazionale. Vi era infine la preoccupazione che i deepfake potessero essere usati per crimini come diffamazione, molestie, ricatto ed estorsione. Il deepfake è stato così messo alla pari di altri crimini di abusi online, principalmente di quelli riguardanti l'immagine (scattare foto intime senza consenso, divulgazione). L'abuso online è diventato centrale nel dibattito, la legge ha dovuto stare al passo con lo sviluppo tecnologico e in particolare per quanto riguarda gli abusi sull'immagine. Il deepfake nell'ambito della post-produzione di immagini pornografiche basate su persone reali è un serio ambito di dibattito che necessita degli opportuni divieti e legislazioni. L'oggettivazione delle donne e l'intento misogino in questi casi rivelano uno scenario degradante, sia che si tratti di celebrità, sia che si tratti di persone meno conosciute. Ci troviamo davanti ad un fenomeno e a delle possibili conseguenze che non vanno ignorate in nessun modo.

2.4. Possibili soluzioni

Nel corso di questa ricerca è emerso come sia necessario regolamentare, sotto diversi punti di vista, l'uso dei deepfake. Analizzando la storia di questi nuovi media digitali e il loro utilizzo in campi il cui uso può diventare pericoloso, è stata messa in luce ripetutamente la necessità di individuare la falsità di alcuni video in circolazione (vedasi in particolare l'ambito politico e quello del fake porn).

Cercando di riassumere i diversi articoli esaminati possiamo affermare che ci siano principalmente quattro modi per combattere deepfakes⁹⁹:

- 1) Attraverso la legislazione e la regolamentazione in materia;
- 2) Con politiche aziendali adeguate,

⁹⁹ Cfr. Westerlund M., *The emergence of deep fake technology: a review*, Technology Innovation Management Review, 9(11) : 40-53.

- 3) Attraverso l'educazione e, nello specifico, l'istruzione e la formazione soprattutto delle fasce più giovani (che devono essere alla base di un'adeguata formazione) così come l'alfabetizzazione digitale;
- 4) Con una tecnologia anti-deepfake (che comprenda il rilevamento del falso, riesca a fornire un'autenticazione dei contenuti e una prevenzione attraverso la distinzione tra ciò che è vero e ciò che è falso).

Come abbiamo visto il crescente sviluppo delle tecnologie di Intelligenza Artificiale richiede la formulazione di nuovi tipi di leggi e normative. Ad esempio, andrà regolamentata la questione della privacy, del copyright (soprattutto nei casi in cui vengano usate immagini o materiali appartenenti ad altre persone esistenti, anche se riformulati attraverso l'I.A.).

La maggior parte delle possibili soluzioni elaborate provengono da centri di ricerca privati e da studi di ricerca all'interno delle università. Un esempio è il centro Faculty di Londra, impegnato nella creazione di un'ampia biblioteca e nell'elaborazione di software automatici che possano verificare la veridicità di questi video. Non solo, ma è in corso un'approfondita ricerca anche per quanto riguarda i rischi a lungo termine nell'uso del deepfake e del deep learning. Anche Microsoft ha annunciato il lancio di uno strumento in grado di generare una sorta di percentuale di autenticità per qualunque video venga inserito, per riuscire a prevenire la minaccia del deepfake nell'ambito delle elezioni presidenziali. D'altra parte, questo *tool* non pare essere ancora del tutto affidabile, sembra che ci siano ancora diverse lacune. Il rischio appare, dunque, ancora molto forte. Si sta discutendo un sistema di etichettatura da parte dei social in cui vengano verificati i contenuti multimediali presenti sui social network; purtroppo, però, fino al momento in cui questi sistemi non diventeranno effettivamente funzionanti, l'affidabilità di queste informazioni nel web resterà precaria. Di conseguenza i deepfake rimarranno una minaccia per la libertà d'informazione e la capacità di informarsi adeguatamente e consapevolmente, sia nell'effettuare scelte politiche, sia nella ricerca di altro materiale (per non parlare della già citata diffamazione e dei danni alla reputazione verificatisi nell'ambito del fake porn).

Abbiamo infatti visto come la divulgazione in ambito non satirico e amatoriale possa essere seriamente dannosa e avere conseguenze preoccupanti per i soggetti coinvolti, con dinamiche insidiose e talvolta pericolose.

A livello tecnico sono stati elaborati dei software (ancora in corso di sviluppo) per identificare i deepfake usando strumenti automatici. Ne è un esempio lo strumento sviluppato dagli informatici dell'università di Buffalo, che analizzando i riflessi di luce negli occhi dei soggetti riesce a smascherare il 94% delle falsificazioni¹⁰⁰. In questo caso la stessa I.A. con cui i deepfake vengono creati può essere utile per smascherarli.

Shu Hu e gli altri collaboratori propongono un metodo di rilevamento basato sull'analisi del riflesso della luce nelle pupille del soggetto ritratto e generato dalla GAN. Si tratta di un rilevamento fisiologico, che funziona in base ad un semplice fatto: anche se l'intelligenza artificiale è in grado di creare volti iperrealistici, in tutto e per tutto simili a soggetti reali, non è tuttavia in grado di creare dei riflessi oculari adeguatamente realistici. In un volto fake ci sono, infatti, differenze per quanto riguarda: forma, estensione, colore, riflessi tra occhio destro e sinistro. A questo proposito Siwei Lyu, uno dei principali ricercatori coinvolti nello studio, afferma che: «La cornea è quasi come una semisfera perfetta ed è molto riflettente. Pertanto, qualsiasi emissione di luce che giunga all'occhio avrà una sua immagine sulla cornea»¹⁰¹.

Le immagini generate da una GAN per la maggior parte non sono in grado di riportare in maniera adeguata una somiglianza di riflessi degli occhi (che di solito nelle foto reali hanno forma e colore simili). Nelle immagini artificiali, invece, i riflessi sono disomogenei, incoerenti, probabilmente per la grande quantità di immagini assemblate tra loro. Questo strumento “sfrutta” questa incoerenza: riesce infatti ad analizzare il volto e a confrontare queste differenze di forma e luce, riuscendo a determinare se quel volto corrisponda ad un volto reale oppure ad uno creato artificialmente. Sono state utilizzate a questo scopo immagini di Flickr e del sito [thispersondoesnotexist](http://thispersondoesnotexist.com). Nell'analisi il sistema riuscirà a generare un punteggio per determinare la somiglianza: più il punteggio sarà alto, più il volto sarà verosimilmente reale, al contrario più il punteggio sarà basso più il volto sarà molto probabilmente creato da un'intelligenza artificiale. In questo modo già nel 94% dei casi si sono riusciti a smascherare una gran quantità di deepfake.

¹⁰⁰ Shu Hu, Yuezun Li, and Siwei Lyu. (2020). Exposing GAN-generated Faces Using Inconsistent Corneal Specular Highlights. arXiv:2009.11924 [cs.CV].

¹⁰¹ Bankead M., How to spot deepfakes? Look at light reflection in the eyes, University at Buffalo, 10 marzo 2021, <http://www.buffalo.edu/news/releases/2021/03/010.html> URL consultato in data 20 agosto 2021.

Ci sono tuttavia dei limiti: il primo è che questo sistema non funziona correttamente in qualsiasi caso, ma solo nei casi in cui ci sia abbastanza luce, tale da generare un riflesso. È necessario poi che nella foto ci siano entrambi gli occhi in vista, ben evidenti, altrimenti la rilevazione non potrà funzionare. Infine, c'è il rischio che questo errore possa essere corretto man mano che si usa lo strumento attraverso l'I.A., portando alla fine del processo all'inganno.

Anche se questo strumento è ancora in fase di sviluppo, è uno dei primi in grado di aiutarci a rilevare i deepfake che circolano in maniera sempre più incontrollata all'interno della rete. Nonostante ci si trovi infatti in una fase di miglioramento e di ricerca, grazie a strumenti simili saremo in grado di difenderci adeguatamente dai fake che dilagano in misura sempre maggiore (e che con l'avanzare del tempo saranno sempre più sofisticati).

L'impegno di questi ricercatori nello sviluppo di tali algoritmi è uno sforzo di sensibilizzazione che si può notare nella necessità generale di sviluppare e produrre strumenti adeguati al rilevamento di contenuti fake. Tra questi strumenti possiamo nominare anche la piattaforma online open-source *Deepfake-o-meter*¹⁰².

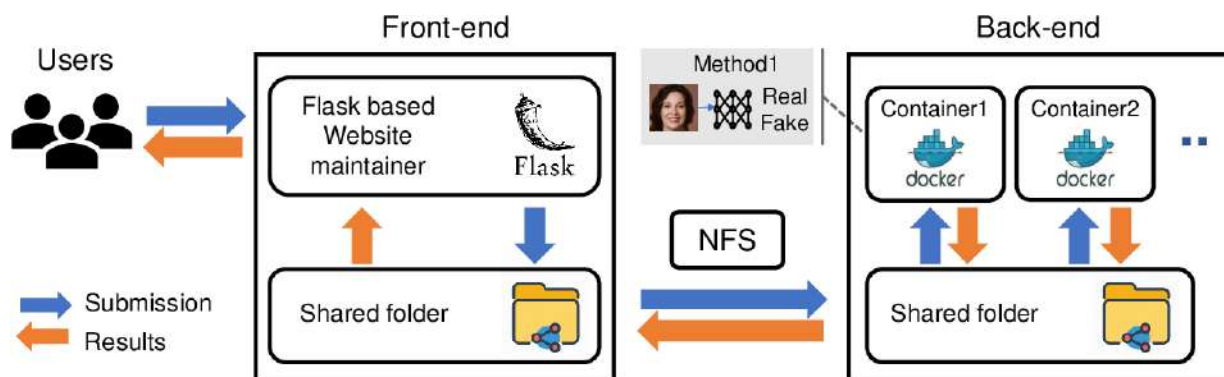


Fig. 15: una panoramica dell'architettura della piattaforma "Deepfake-o-meter"

¹⁰²Yuezun Li, Cong Zhang, Pu Sun, Honggang Qi, Siwei Lyu, *Deepfake-o-meter: an open platform for Deepfake Detection*, Ocean University of China, 2021, arXiv preprint arXiv:2103.02018.

Questa piattaforma integra metodi di rilevamento dei deepfake all'avanguardia, coniugando un'interfaccia appropriata per gli utenti. Nel documento in cui si spiega il funzionamento di questo strumento, gli autori affermano che, per l'appunto, lasciato incontrollato il Deepfake può aumentare il pericolo della disinformazione e far diminuire progressivamente la fiducia nei media digitali. Le crescenti preoccupazioni per i Deepfakes hanno portato anche una sensibilizzazione verso il rilevamento di questi ultimi: a questo fine è stata sviluppata una piattaforma di rilevamento online con degli scopi precisi¹⁰³, tra cui:

- Aiutare gli sviluppatori di algoritmi di rilevamento deepfake con un'architettura API;
- Aiutare i ricercatori con una piattaforma in grado di valutare e confrontare più algoritmi con lo stesso input;
- Aiutare gli utenti fornendo un portale in grado di utilizzare più algoritmi di rilevamento all'avanguardia.

Anche questo strumento è ancora in fase di sviluppo, due di questi algoritmi sono tra i migliori esecutori della *Global Deepfake Detention Challenge*. Questa piattaforma rappresenta dunque un grande traguardo, che tuttavia è ancora in fase di elaborazione e miglioramento, pur rappresentando di per sé lo sforzo e la sensibilizzazione riguardo questo argomento.

Anche in Italia si sta avviando un processo di consapevolezza, attraverso la ricerca in ambito accademico e lo sviluppo di strumenti adeguati. A questo proposito le Università di Trento e di Firenze hanno collaborato da ottobre 2020 ad un progetto chiamato *Unchained*¹⁰⁴ (*Uncovering media manipulation chains through container and content detectable traces*), con lo scopo di rilevare i contenuti fake presenti in rete, manipolati e potenzialmente diffamatori. Il progetto è finanziato dall'agenzia governativa del Dipartimento della Difesa degli Stati Uniti. Mentre l'Università di Firenze si occupa dell'analisi del formato, il team di Trento si impegna nell'analisi del contenuto, sfruttando le competenze maturate nell'analisi

¹⁰³ *Ibidem*.

¹⁰⁴ Cfr. articolo *Le Università di Trento e Firenze contro il deepfake. Rilevare in rete contenuti multimediali manipolati e diffamatori*, redazione ANSA, 27 ottobre 2020, https://www.ansa.it/toscana/notizie/2020/10/27/le-universita-di-trento-e-di-firenze-contro-il-deep-fake_e5f7aea9-a0ca-4472-9b16-9170ee5310a8.html URL visitato in data 31 agosto 2021.

del viso umano e del segnale. La professoressa del Dipartimento di Ingegneria e Scienza dell'informazione dell'Università di Trento, Giulia Boato, afferma a riguardo:

Il nostro lavoro consiste nello sviluppare algoritmi, basati sia su analisi statistiche sia su paradigmi di deep-learning, adatti a scandagliare e ripercorrere tutta la catena del dato multimediale. L'investigazione forense ha bisogno della ricostruzione complessiva. Solo così possiamo dare un supporto ai servizi di intelligence, per la polizia postale e per tutti gli attori preposti a tracciare contenuti falsi e malevoli¹⁰⁵.

Alessandro Piva, docente del Dipartimento di Ingegneria dell'informazione, che lavora nell'ambito dell'autenticazione di contenuti digitali manipolati da oltre dieci anni, afferma come all'inizio della sua ricerca le manipolazioni riguardassero unicamente le immagini e non i video. Con lo sviluppo della tecnologia, l'aumento e il miglioramento degli strumenti per la manipolazione, diventa sempre più difficile difendersi. Il professore dichiara a riguardo:

Oggi mediante algoritmi di intelligenza artificiale è possibile creare in poco tempo manipolazioni complesse sia di immagini e di video, che diventano ben presto virali. L'identificazione di manipolazioni diventa quindi sempre più complessa. A Firenze ci stiamo concentrando sullo studio dell'integrità di video, in particolare quelli scambiati nell'ambiente social¹⁰⁶.

Appare di certo necessario, finché non sarà possibile verificare attraverso adeguati software la veridicità dei video, adottare un differente approccio: più consapevole, più critico. Allo stesso tempo però, adottando questo approccio al suo estremo e dunque mettendo in discussione qualunque cosa, si potrebbe arrivare alla totale diffidenza verso tutto ciò che ci si presenta online (e non solo), generando una cultura del sospetto. Per evitare di cadere in questa aporia, come già affermato precedentemente, è necessario dare un'adeguata educazione ai nuovi media digitali (soprattutto ai giovani) in modo tale da avere degli

¹⁰⁵ *Ibidem*.

¹⁰⁶ Cit. articolo *Deepfake: l'Università di Firenze in prima linea contro i video falsificati*, pubblicato su InToscana il 28 ottobre 2020, <https://www.intoscana.it/it/articolo/deepfake-universita-firenze/> URL visitato in data 31 agosto 2021.

strumenti per essere in grado di discernere e, quantomeno, provare a distinguere il vero dal falso.

Tutto ciò sarà necessario, se non si vorrà cadere nell'illusione e nell'inganno totale, al pari, se non peggio, della scena originaria della dama husserliana. L'iperrealismo ha raggiunto la sua massima evoluzione.

documenti non è una fotografia originale, bensì un ritratto generato digitalmente e artificialmente, a computer, attraverso l'uso di diversi software e tecnologie 3D. Il ritratto appare come una fotografia, eppure non lo è. Pietro Conte in *Mockumentality: from hyperfaces to deepfakes* afferma che “esso non è ontologicamente una fotografia, eppure appare fenomenologicamente come tale” (di nuovo, emerge il contrasto tra “essere” ed “essere percepito”)¹⁰⁷. Il ritratto iperrealistico, dunque, crea problema a livello di implicazioni sociopolitiche (non può, infatti, venire considerato un documento valido), eppure a tutti gli effetti a livello artistico il lavoro creativo e il processo alla base ha una sua valenza, una commistione tra genio e originalità che caratterizza tutti i lavori artistici)¹⁰⁸. L'iperrealismo crea in questo caso ancora dei problemi, principalmente a livello legale (si tratta pur sempre di un documento d'identità, in tutto e per tutto simile e dunque confondibile con l'originale), ma a livello artistico il processo usato desta curiosità: cosa succederebbe se l'uso delle GANs e dell'intelligenza artificiale, approfondito precedentemente in relazione ai deepfake e all'uso iperrealistico, venisse invece applicato al campo artistico e creativo? Si potrebbe, in questo caso, aprire una nuova strada.

Riassumendo quanto è già stato detto precedentemente e chiarendolo in maniera più dettagliata: per quanto riguarda il funzionamento di una rete neurale qualsiasi, possiamo dire in maniera semplicistica che questa ha la funzione di calcolare dei valori (pesi) che moltiplicati per i valori di *input* producono i valori attesi in *output*. Dunque, nel caso delle immagini, poniamo che l'*input* sia un'immagine e l'*output* sia un'altra immagine, e che lo scopo della rete neurale sia imparare a trasformare la prima nella seconda. La prima immagine è vista dalla rete come una serie di valori numerici, così come la seconda (ricordiamo, infatti, senza farci fuorviare dalla “trappola umanizzante” che le immagini vengono ridotte a numeri: il confronto fatto dalla GAN è sempre numerico, anche se il risultato finale a noi sembra visivo). Dunque, la rete deve trovare dei valori che moltiplicati

¹⁰⁷ Conte P., *Mockumentality: From Hyperfaces to Deepfakes* in WORLD LITERATURE STUDIES, vol. 11, pp. 11-25 (ISSN 1337-9690), p. 17.

¹⁰⁸ Cfr. a riguardo, Ivi P. 19. Cit: *CMI* has, of course, several socio-political implications, but it first concerns aesthetics, for it challenges the traditional opposition between the *mechanical* process of imprint taking (which generates copies that can serve as documents) and the *creative* process of hand modeling (which is an essential part of the “genius” and “originality” that we use to ascribe to all works of art).

per la prima matrice producano la seconda. Per farlo la rete assegna in principio dei valori casuali, che necessariamente, essendo casuali, produrranno un risultato molto diverso dall'immagine desiderata. Ma, nell'architettura delle GANs, avendo un *training set* di riferimento (ovvero l'insieme delle immagini desiderate) possiamo calcolare una funzione che misuri l'errore di stima prodotto dai pesi casuali. Questo errore viene usato per correggere gli iniziali valori casuali in modo da ridurre l'errore (tecnicamente è un problema di ottimizzazione per cui vogliamo minimizzare la funzione dell'errore). Grazie a queste continue correzioni e aggiustamenti dei pesi alla fine la rete stima i parametri più efficaci e in questo senso impara a produrre l'immagine desiderata.

Ma, come abbiamo visto, ciò richiede un *dataset* iniziale (*training set*) su cui la rete impara. Ora, come possiamo fare a produrre lo stesso risultato senza usare un *training set*? L'idea è di usare due modelli, un generatore (che essenzialmente funziona come la rete generica descritta prima) e un discriminatore. Il discriminatore è un classificatore addestrato semplicemente a riconoscere se un'immagine è vera o falsa.

Il generatore in questo schema lavora in modo non supervisionato, ovvero non ha il *training set*. Quindi inizialmente genera dei valori casuali, ma come è possibile misurare l'errore non avendo un'immagine di riferimento? L'idea è fare riferimento al discriminatore: se quest'ultimo riconosce l'immagine come falsa (e più in generale fintanto che il discriminatore ha vita facile a distinguere il falso dal vero) allora il generatore ha un errore e può basarsi sulla stima del discriminatore per calcolare l'errore. A quel punto ci applica più o meno lo stesso procedimento della rete generica e si correggono i parametri fino a produrre immagini per le quali il discriminatore viene ingannato sistematicamente (ovvero immagini non facilmente distinguibili dal vero).

Dunque, possiamo dire che il feedback che il discriminatore invia al generatore non consiste in un semplice 0 / 1 (reale/fake), ma in una *scala* di valori che indica la probabilità che l'immagine sottoposta dal generator sia reale o *fake*. Quando il valore di queste probabilità cresce anche di poco, il generatore apprende di essere migliorato, di trovarsi più vicino al risultato desiderato¹⁰⁹.

¹⁰⁹ Si ringrazia per la spiegazione e il chiarimento sul funzionamento delle reti neurali e delle GANs il professor Alfio Ferrara.

Questo per chiarire ciò che avevamo visto in precedenza riguardo il funzionamento delle GANs e, nello specifico, la produzione di immagini iperrealistiche. Ma poniamo un ulteriore interrogativo e facciamo un passo oltre: se come set di allenamento delle GANs anziché foto, ritratti e immagini realistiche ponessimo immagini di quadri e dipinti cosa succederebbe? Seguendo questo ragionamento, le GANs sarebbero portate a produrre sicuramente un prodotto nuovo, ma non ci sarebbe una vera e propria “spinta creativa”, in quanto l’unico intento sarebbe “ingannare” il discriminatore ponendo dati simili al set di addestramento (in questo caso opere). Ma se, restando in campo artistico, una persona in carne ed ossa nell’intento di creare un’opera d’arte ne creasse una in tutto e per tutto simile ad un’altra già esistente, non potremmo mai considerare quest’ultima vera arte: sarebbe solo una copia identica ad un’altra già esistente: dunque un plagio, una truffa, un’imitazione.

Possiamo dire che, cercando di darne una definizione¹¹⁰, l’arte deve sicuramente riprendere qualcosa di già conosciuto (basarsi su dei fondamenti, appresi attraverso lo studio delle tecniche e della storia dell’arte), ma una parte fondamentale deve essere quel carattere di originalità, di sorpresa. Le GANs, viste fino a questo punto, hanno una funzione più imitativa ed emulativa che creativa, una funzione che abbiamo spesso definito, soprattutto in relazione ai deepfake e alle *fake news*, *ingannevole*. Ma c’è spazio, nell’uso di questi strumenti, per la creatività?

3.1.2. Le Creative Adversarial Networks: dall’emulazione alla creatività

Abbiamo notato fino ad ora l’incredibile progresso e gli sviluppi avvenuti nel campo dell’intelligenza artificiale. Resta aperta una questione che suscita molti interrogativi e provoca un nuovo dibattito: è possibile per l’uomo insegnare alla macchina e al computer a pensare in modo creativo? Se tanti progressi sono stati fatti e se l’uomo è riuscito ad insegnare

¹¹⁰ Si veda, a riguardo, la definizione di arte espressa da Daniel E Berlyne in *Aesthetics and psychobiology*, volume 336. JSTOR, 1971.

alla macchina a creare nuovi prodotti iperrealistici, è possibile invece, allo stesso modo, insegnare alla macchina come realizzare prodotti artistici, oppure la creatività è una caratteristica prettamente umana? Ci si presenta una nuova sfida, che suscita diverse controversie.

Giugno 2017: viene pubblicato un documento¹¹¹ in cui diversi ricercatori del dipartimento di informatica della Rutgers University of New York e del dipartimento di storia dell'arte del College of Charleston, introducono l'idea di un nuovo strumento, le Creative Adversarial Networks (CANs). Le CANs sono, in maniera riassuntiva, delle GANs in grado di “pensare in maniera creativa”, o meglio, di produrre delle immagini artificiali che non sono, come abbiamo visto precedentemente, dei prodotti iperrealistici confondibili con la realtà, bensì dei prodotti paragonabili a vere e proprie opere d'arte. Ma come funzionano nello specifico queste reti neurali?

Le CANs sono architettate in maniera molto simile alle GANs, con la differenza che vi è stata aggiunta una parte che permette al Generatore di produrre immagini “pensando” in maniera creativa, con un risultato dunque differente rispetto a quelli visti in precedenza. Basandosi su un processo simile a quello sopra descritto, anche in questo caso il discriminatore cercherà di imparare a classificare le immagini come reali o false, ma classificandole stavolta anche secondo diversi stili artistici (come rinascimento, cubismo, astrattismo, surrealismo, espressionismo ecc.). Il generatore non solo cercherà di ingannare il discriminatore, come nell'architettura delle GANs, ma cercherà anche di rendere a quest'ultimo difficile il classificare le immagini prodotte secondo uno degli stili.

Nel documento redatto dai ricercatori, si esprime l'obiettivo dello studio: “The goal of this paper is to investigate a computational creative system for art generation without involving a human artist in the creative process, but nevertheless involving human creative products in the learning process”¹¹². L'obiettivo è, infatti, quello di creare un algoritmo che

¹¹¹ Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, Marian Mazzone, *CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms*, arXiv:1706.07068v1 [cs.AI] 21 giugno 2017.

¹¹² Ivi, cit. p. 2. Tr: “L'obiettivo di questo articolo è quello di indagare un sistema creativo computazionale per la generazione di arte senza coinvolgere un artista umano nel processo creativo, ma comunque coinvolgendo prodotti creativi umani nel processo di apprendimento”.

abbia anche una funzione estetica: un sistema interattivo in cui il computer abbia uno spazio per la creatività e l'uomo assuma la funzione dello spettatore, che tuttavia dà le indicazioni per avviare e guidare il processo. Nel caso delle GANs (spiegate in maniera più dettagliata nel capitolo precedente) lo scopo era quello dell'emulazione: non c'era motivo di generare qualcosa di creativo. Il *training set* (ovvero le immagini su cui il generatore doveva basarsi per creare immagini artificialmente) era composto da immagini reali, dunque il risultato finale doveva essere quello di immagini iperrealistiche e confondibili con la realtà, che ingannassero lo spettatore. Come appurato, tutto questo poteva essere fatto anche con dipinti, quadri, immagini artistiche, ma sarebbe stato ugualmente un processo di emulazione. Con le CANs, dunque, è possibile passare da un processo di emulazione ad uno più creativo: in che modo? Vediamolo in maniera più tecnica.

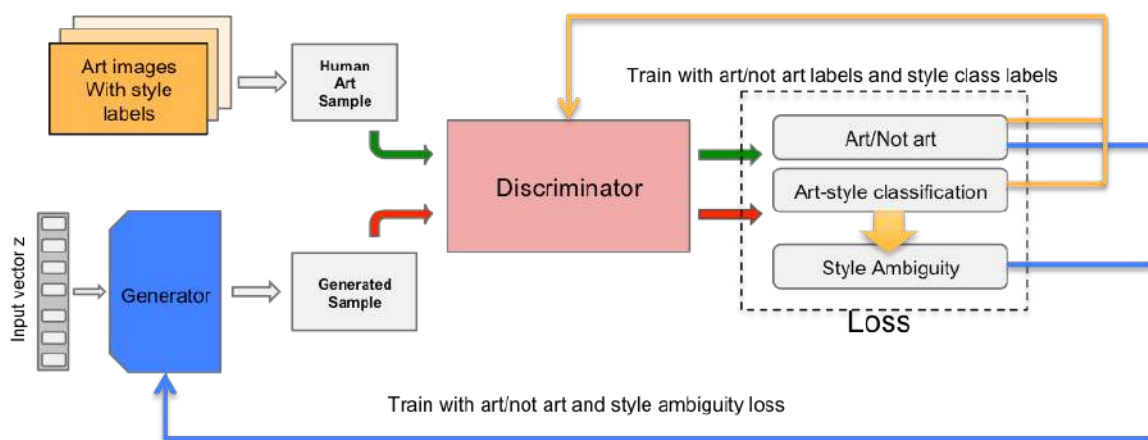


Fig. 17: schema tecnico dell'architettura delle CANs.

Possiamo vedere in maniera più chiara dallo schema proposto il funzionamento e la differenza rispetto al caso precedente: in questo caso il generatore riceverà due segnali (anziché uno solo) dal discriminatore, per poter raggiungere tre punti (per poter equilibrare il sistema):

- 1) elaborare nuove immagini,

- 2) non elaborare opere troppo “innovative”: l’immagine generata non può essere troppo lontana dalla distribuzione altrimenti produrrà un’eccitazione eccessiva (attivando un sistema di avversione)¹¹³,
- 3) il lavoro generato dovrebbe aumentare l’ambiguità stilistica, rendendo difficile la collocazione dell’immagine in uno degli stili precedentemente definiti.

Dunque, come possiamo vedere, la rete generativa avversaria è composta da generatore e discriminatore, come nel caso delle GANs. Il discriminatore ha la possibilità di accedere ad un’ampia gamma di opere artistiche, appartenenti a diversi stili pittorici. Le immagini vengono usate per poter discernere tra i vari differenti stili. Il generatore, invece, non ha la possibilità di accedere ad alcuna opera. Come nelle GANs il generatore inizia a creare a partire da un input casuale, ricevendo però dal discriminatore questa volta due diversi segnali come risposta al lavoro generato:

- 1) l’immagine creata è arte/ non è arte: questo segnale permetterà al generatore di regolare la produzione delle successive immagini e, dunque, di allenarsi. La risposta verrà data dal discriminatore ugualmente in base al set di immagini artistiche che è stato sottoposto. Essendo il discriminatore allenato in questo modo, la risposta verrà data in base all’arte conosciuta (che è stata sottoposta). Il generatore, allenato in questo modo, cercherà di produrre immagini sempre più simili all’arte.

- 2) l’immagine è classificabile in uno degli stili predefiniti. Questo segnale definisce quanto bene il discriminatore può classificare l’arte, negli stili conosciuti. Se le immagini elaborate dal generatore sono facilmente classificabili in uno di questi stili il generatore è riuscito ad ingannare il discriminante, facendogli credere che l’arte prodotta fosse reale. Il generatore “creativo” deve essere in grado da un lato di generare un’immagine che possa ingannare il discriminatore (facendogli credere che si tratti di arte), dall’altro di confonderlo sullo stile.

Ci troviamo davanti a segnali contraddittori. Se da un lato un segnale spinge in direzione di opere che il discriminatore possa accettare come arte, dall’altro se riesce in

¹¹³ Cfr. a riguardo la definizione di “*arousal*” di D.E. Berlyne, spiegata nel documento *CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms*, p. 3.

questo scopo, il secondo segnale sarà in grado di classificare secondo gli stili predefiniti. Il secondo segnale, invece, spingerà verso la produzione di lavori stilisticamente ambigui. Di conseguenza, i due segnali insieme dovrebbero portare il generatore ad esplorare nuove zone dello spazio creativo, per portarlo a raggiungere l'obiettivo di avvicinarsi alla definizione di arte (primo obiettivo), ma anche ad allontanarsi dagli stili artistici standard (ambiguità). La rete neurale andrà ovviamente allenata (come già visto precedentemente nel caso delle GANs).



Fig. 18: esempi di immagini artistiche prodotte da CANs.

Abbiamo visto in maniera più tecnica e dettagliata il funzionamento di queste reti neurali e la loro differenza: l'aggiunta che permette di differenziarsi rispetto al caso precedente delle GAN. Tuttavia, resta necessario soffermarsi ancora un attimo per poter capire il perché l'aggiunta della classificazione in stili artistici permetta al generatore di pensare in maniera creativa. Bisogna, infatti, puntualizzare una definizione concreta di creatività in riferimento ad una macchina.

La creatività viene percepita dagli spettatori quando un pezzo d'arte rappresenta unicità: esso deve essere originale, ma al tempo stesso riconducibile a qualche categoria, a qualcosa di conosciuto. Come leggiamo nell'articolo quando si parla di *arousal*¹¹⁴: nell'osservazione di un'opera da parte dello spettatore, nel caso in cui essa abbia troppo poco potenziale di eccitazione verrà considerata dallo spettatore noiosa e, d'altra parte, nel caso in cui essa dia vita ad una "sovraccitazione" attiverà un sistema di avversione (e, dunque, la risposta sarà negativa).

Nell'architettura del sistema neurale sopra illustrata, il generatore verrà premiato quando riuscirà a generare immagini che il discriminatore non sia in grado di classificare facilmente in uno degli stili artistici conosciuti. Di conseguenza, esso sarà costretto a generare immagini diverse, uniche, più creative. Il generatore dovrà comunque ingannare il discriminatore in modo tale da pensare che le immagini siano reali, quindi non generando immagini esageratamente fantasiose che non si possono ricondurre a qualcos'altro. Ecco la simulazione di come comunemente vediamo la creatività nell'arte.

Riassumendo: una *Creative Adversarial Network* è basata su una *Generative Adversarial Network*, con l'aggiunta di casualità nell'output della rete neurale generativa. L'output può essere addestrato: troppa casualità renderà il tutto poco accattivante esteticamente. Questo genererà nuovi risultati, con un tentativo di imitazione della creatività umana e la rottura rispetto a modelli preesistenti.

Come spiegato nel documento dell'università di Rutgers che spiega il funzionamento delle CANs, questo innovativo sistema è stato allenato usando un ampio set di immagini prese da collezioni d'arte dal quindicesimo al ventunesimo secolo. I risultati sono stati valutati sulla base di opinioni di spettatori umani che, spesso, hanno confuso l'arte generata artificialmente con l'arte generata da altri esseri umani, valutando addirittura in alcune occasioni l'arte generata dalle reti neurali di un livello superiore.

Vengono precisate inoltre le caratteristiche che deve avere il sistema proposto: capacità di produrre opere nuove (immaginazione), capacità di generare opere di qualità

¹¹⁴ Ivi, pp. 3-4.

(abilità) e la capacità di valutare ciò che è stato creato¹¹⁵. In questo caso, quindi, l'interazione tra i due segnali del discriminatore forzerà il sistema e il generatore ad esplorare nuovi spazi creativi, restando vicino al confine dell'arte (venire riconosciuta come opera artistica), ma al tempo stesso discostandosi dagli stili predefiniti (originalità). La novità è l'apprendimento attraverso la storia dell'arte: questo sistema sarà in grado di apprendere attraverso l'esposizione a nuovi stili e nuove immagini, adattando le nuove immagini generate a ciò che è stato appreso nell'esposizione ad immagini e stili diversi tra loro¹¹⁶. Ovviamente, in quanto macchina e meccanismo automatizzato, la rete neurale non è in grado di capire fino in fondo la disciplina artistica e il concetto di arte, tuttavia è in grado di apprendere la storia e di prendere come esempio i modelli proposti.

Siamo giunti ad un nuovo stadio, quello della creatività, e abbiamo parzialmente risposto ad un interrogativo, ovvero: possono le macchine essere creative? Possono generare arte? Se da un lato questa nuova strada nel progresso delle GANs e dell'intelligenza artificiale apre nuove possibilità e nuovi margini di esplorazione (più positivi rispetto al campo di applicazione del *deepfake* precedentemente visto, con lo scopo di ingannare lo spettatore), la questione della creatività da parte delle macchine crea ancora un ampio dibattito e lascia spazio alla discussione. Rimane ancora tanto da indagare e il campo della ricerca è ancora ai suoi esordi, tuttavia tra le diverse controversie, alcuni fatti ci dimostrano che l'arte e la digitalizzazione non viaggiano su binari completamente distanti, anzi. Le reti neurali sono già state applicati a diversi campi creativi e hanno dato vita ad alcune opere artistiche. Sarà opportuno capire quali possano essere i campi artistici in cui questi strumenti possono essere applicati.

3.1.3. I diversi campi artistici in cui sono state impiegate le Creative Adversarial Networks

¹¹⁵ Ivi, cit. p. 20. Si veda, a riguardo, Simon Colton. *Creativity versus the perception of creativity in computational systems*, in *AAAI spring symposium: creative intelligent systems*, volume 8, 2008.

¹¹⁶ Ivi, pp. 20-21.

Siamo giunti ad un nuovo, sorprendente risultato: le GANs offrono la possibilità ai computer e alle macchine di essere creativi. Ci sono diversi campi in cui possono essere applicate, tra cui: la creazione di immagini artistiche, il riempimento di parti mancanti di immagini, la conversione del testo in immagine (*Text-to-Image Synthesis*), la composizione di musica.

Oltre alla creazione di immagini a partire da altre immagini, vi è infatti la possibilità di usare questo strumento per convertire un testo in immagine: spesso le opere che ne risultano sono anche molto accurate e si basano su una descrizione testuale, anziché su set di immagini preesistenti. In questo caso si può “addestrare” una GAN affinché impari a scrivere. Questo metodo di creazione delle immagini, chiamato per l’appunto *Text-to-image generation* (o *synthesis*), viene realizzato attraverso delle reti neurali chiamate vmCAN, ovvero *visual-memory Creative Adversarial Networks* e funziona generando immagini a seconda di alcune frasi narrative corrispondenti, dunque fondendo una memoria di conoscenza visiva esterna e ad una fusione multimodale di immagini. La generazione di immagini attraverso descrizioni linguistiche è un campo di ricerca considerato tuttora molto attuale, applicabile a diversi ambiti creativi come l’editing di immagini o l’elaborazione di videogiochi. Questo modello di Creative Adversarial Network per la generazione di immagini complesse del mondo è tuttora in via di sviluppo¹¹⁷ e dimostra come l’uso delle GANs possa essere applicato in maniera creativa a campi innovativi e diversi da quello unicamente “ingannevole” dei *deepfake*.

Restando nel campo della creatività e cercando di fornire altri esempi di sperimentazione di questi algoritmi, è opportuno indicare come anche il campo musicale sia stato indagato attraverso l’uso delle GANs: nel saggio *Alla ricerca della creatività: le GAN come paradigma dell’autonomia nel software per la composizione musicale*¹¹⁸, Caterina Moruzzi illustra l’impatto dell’intelligenza artificiale sulla natura delle opere musicali. Nel saggio l’autrice, a partire da una considerazione sul concetto di creatività e cercando di darne una definizione, espone come nel campo delle arti visive l’uso delle GAN abbia dato i suoi

¹¹⁷ Si veda, a riguardo, S. Zhang, H. Dong, W. Hu, Y. Guo, C. Wu, D. Xie, and F. Wu. *Text-to-image synthesis via visual-memory creative adversarial network*, in *PCM*, 2018.

¹¹⁸ C. Moruzzi, *Alla ricerca della creatività: le GAN come paradigma dell’autonomia nel software per la composizione musicale*, in A. Barale, *Be my gan*, JacaBook, Milano, 2020.

frutti. Tuttavia, anche in campo musicale i risultati potrebbero essere sorprendenti (nonostante la generazione di immagini e video sia stata sperimentata maggiormente e in un lasso di tempo più ampio rispetto al caso della musica).

L'autrice espone in primo luogo la difficoltà nel definire il concetto di creatività in relazione alla tecnologia: quest'ultimo, infatti, sembra avere un diretto collegamento con il soggetto umano che ne è l'autore. Nonostante ciò, abbiamo visto come altre definizioni di creatività includano principalmente i concetti di sorpresa e di novità, pur basandosi su dati preesistenti. Nel saggio il concetto di creatività viene arricchito dalle caratteristiche di autonomia e capacità di auto-valutazione. Nonostante la qualità dei prodotti musicali prodotti da GAN non abbia raggiunto ancora risultati sorprendenti, l'autrice mette in luce come “la qualità esteticamente inferiore dei risultati prodotti attualmente dalle GAN nel campo della generazione musicale non possa essere utilizzata come argomento contro la loro creatività”¹¹⁹.

Prima di porsi la domanda sulla tecnologia, e più nello specifico sulla possibilità per i computer di creare musica, è necessario porsi innanzitutto la domanda sulla creatività: vi è la reale possibilità, per le macchine, di essere creative? Si cerca, dunque, di dare una risposta definendo la creatività come “processo autonomo di risoluzione dei problemi”¹²⁰. Nel saggio viene, inoltre, aggiunta una definizione di Margaret Boden:

La creatività è la capacità di inventare idee o creazioni nuove, sorprendenti e dotate di valore. [...] un'attività creativa produce qualcosa che non è mai stato fatto prima. Inoltre, la novità che è frutto di un atto creativo deve essere necessariamente accompagnata da una sensazione di sorpresa. Infine, questa idea o creazione nuova e sorprendente deve essere “dotata di valore”¹²¹.

Vi sono però dei limiti nella definizione proposta, soprattutto per quanto riguarda la definizione di valore: se così fosse, nel caso specifico della musica la macchina perderebbe la possibilità di essere creativa a causa della ridotta qualità del risultato e, dunque, della mancanza di valore attribuita all'opera. Per arginare questo problema, C. Moruzzi collega più esaurientemente la definizione di creatività a quella di autonomia nella risoluzione dei

¹¹⁹ Ivi, cit. p. 148.

¹²⁰ Ivi, cit. p. 149.

¹²¹ Ivi, cit. p. 150.

problemi. Vedendo il concetto sotto quest'ottica, l'uso di questi strumenti e l'interazione dell'uomo con la tecnologia può diventare un proficuo terreno di sviluppo: la tecnologia non risulterebbe “nemica ingannatrice” o sostituta dell'uomo, bensì un modo per accrescere ed eventualmente migliorare le capacità della mente umana. Da questo punto di vista la creatività dell'intelligenza artificiale offre spunti di sviluppo molto interessanti, che possono suscitare l'attenzione di artisti, informatici e filosofi, unendo diverse discipline. Da questo punto di vista, dunque, si può attribuire la caratteristica della creatività alle macchine: non solo, l'interazione uomo-macchina, uomo-tecnologia potrebbe migliorare le capacità dell'uomo e dare vita a nuovi sviluppi, ampliandone le potenzialità.

Questo per quanto riguarda, ad un livello più generale, il discorso sulla creatività e sulla possibilità di creare arte attraverso l'uso dell'intelligenza artificiale.

Parlando nello specifico delle GAN, l'autrice precisa come questi algoritmi si presentino sottoforma di diverse tipologie, in grado di elaborare dati principalmente sulla base del *training set*. Se nel campo delle immagini e dei video questi algoritmi hanno suscitato un particolare consenso, dal punto di vista della composizione musicale il processo di produzione deve ancora conoscere alcuni progressi. Tuttavia, riprendendo la definizione di creatività e collegandola al concetto di autonomia, appare evidente come il processo delle GAN possa essere considerato a tutti gli effetti creativo: “l'interazione tra *generator* e *discriminator* che ha luogo nella struttura delle GAN conferisce al processo di creazione un considerevole grado di autonomia dal *training set*”¹²².

Ma se l'auto-valutazione viene considerata una delle caratteristiche principali della creatività, in che modo le GAN possono essere considerate autonome e in grado di auto-valutarsi? Secondo l'autrice, questo avviene grazie all'interazione, per l'appunto, tra generatore e discriminatore e al loro *feedback* con il *training set*. Si può, infatti, paragonare questo procedimento a quello più umano che si basa su tentativi ed errori. Nel saggio si mette in luce come il processo sia qualcosa di fondamentale per la considerazione della creatività, molto più del risultato: non importa se il risultato ottenuto non sia di estrema ed elevata qualità, ciò che conta più di tutto è il processo che ha portato a quel risultato. E, sotto questo

¹²² Ivi, cit. p. 154.

punto di vista, le macchine e le GAN nello specifico possono essere classificate come dotate di creatività.

Il ciclo di *feedback* – che costituisce un elemento vitale dell’architettura delle GAN – si può interpretare come un processo di auto-valutazione. [...] particolarmente rilevante è il fatto che questo meccanismo di *feedback* agisca all’interno delle GAN e non presupponga attori esterni. Tradizionalmente, la computazione evolutiva interattiva si serve di giudici umani per la valutazione nei casi in cui una funzione di *fitness* è ignota o risulta troppo complessa da determinare [...]. In tali casi, il *feedback* proviene dall’esterno e non può quindi essere considerato un’auto-valutazione. Nelle GAN, per contro, il *feedback* proviene dall’interno¹²³.

La caratteristica appena messa in evidenza è fondamentale: rispetto ad altri modelli computazionali e di intelligenza artificiale spicca la novità introdotta dalle GAN, ciò che le differenzia e le rende più autonome (e di conseguenza, in grado di essere creative): è il tipo di apprendimento che avviene nel confronto tra le due reti neurali e il riferimento al *training set*. Paragonando questo meccanismo alla mente umana, il *training set* può essere pensato come una sorta di “modello di riferimento”, una raccolta di dati e di materiale a cui la macchina fa riferimento così come l’uomo, nella realtà, fa riferimento ad una serie di regole e di istruzioni per mettere in atto processi creativi.

La caratteristica più sorprendente ed innovativa, tuttavia, è proprio questa “competizione” che avviene tra il generatore e il discriminatore. È come se le due reti neurali fossero contrapposte e in competizione, in una sorta di gara in cui una delle due fosse destinata a vincere e l’altra a perdere, ma in realtà non “vincessero” né “perdessero” mai l’una sull’altra: come abbiamo già visto, infatti, il processo funziona quando si raggiunge l’equilibrio tra le due. Quando le due reti trovano l’equilibrio, ecco nascere i risultati migliori. E, tuttavia, seppur nate e destinate a competere tra loro, esse imparano e si migliorano a vicenda: imparano facendo, imparano dalle vittorie e dagli errori. Contrapponendosi si migliorano a vicenda e migliorano le proprie prestazioni, producendo così un risultato ogni volta migliore. Vi è una dipendenza ambivalente.

Questo processo può, in qualche modo, ricordare l’essere umano: indubbiamente i processi computazionali e algoritmici sono ancora molto lontani da quelli della mente umana,

¹²³ Ivi, cit. p. 158.

ma in qualche modo la sorprendente capacità di auto-valutazione e di apprendimento delle GAN ci può far confrontare con un grande salto in avanti nel campo dell'intelligenza artificiale, aprendo nuovi e fino ad ora sconosciuti orizzonti.

Nel caso delle Creative Adversarial Networks, abbiamo visto come oltre alle due reti neurali antagoniste e a confronto con un *training set*, vi sia anche la possibilità di discostarsi da quest'ultimo, creando nuovi stili. Anche in questo caso, l'intervento umano è pressoché assente: il sistema algoritmico riesce a deviare dagli stili proposti dal *training set* autonomamente e creando, almeno in prima istanza, dipinti che appaiono astratti. Ci stiamo spingendo sempre di più nella direzione di un sistema autonomo rispetto ai dati esterni e sempre più simile a quello che potremmo a tutti gli effetti definire un processo creativo.

Nel campo della musica, che è quello di cui ci stiamo occupando attualmente, come precisa C. Moruzzi, non sono ancora stati raggiunti risultati qualitativamente elevati (cosa che invece è stata raggiunta maggiormente nel campo delle arti visive – vedremo in seguito come anche le case d'asta si siano interessate sempre di più all'arte digitale e all'arte prodotta da *Gan artists*) e vi sono ad oggi ancora diversi limiti e diverse critiche a riguardo¹²⁴.

Principalmente, i limiti e le critiche riguardano l'addestramento delle reti neurali nel campo della produzione musicale e i risultati apparentemente inferiori rispetto ad altri sistemi computazionali e algoritmici. Proprio ciò che precedentemente abbiamo classificato come caratteristica principale delle GAN, ciò che le rende propriamente creative, ovvero l'interazione tra generatore e discriminatore, è anche ciò che causa maggiori difficoltà a livello di addestramento del sistema. Abbiamo infatti visto che l'equilibrio è ciò che porta le reti a funzionare adeguatamente e a produrre i migliori risultati. D'altra parte, questo equilibrio non è semplice da raggiungere, anzi, spesso risulta instabile.

Una possibile soluzione¹²⁵ in campo musicale è stata trovata nell'affiancare le GAN ad altri modelli, per migliorare ulteriormente i risultati che non sono ancora definibili come “qualitativamente riusciti”. Non ci occuperemo nel dettaglio del funzionamento algoritmico nella produzione di musica e nemmeno di individuare in maniera dettagliata questi programmi e modelli informatici. Questi esempi sono necessari, invece, per mostrare come

¹²⁴ Cfr. Ivi, p. 161.

¹²⁵ Cfr. Ivi, p. 162.

lo sviluppo delle GAN abbia subito un incremento esponenziale nell'arco di pochi anni e come, inoltre, il processo di questi algoritmi stia raggiungendo la caratteristica della creatività. Di nuovo, in contrapposizione ai *deepfake* e alle *fake news*, vediamo lo schiudersi di enormi potenzialità e diversi campi di applicazione.

In relazione a questo discorso e alle potenzialità di questi mezzi in campo artistico possiamo dire, riprendendo le parole dell'autrice che:

La qualità inferiore della musica prodotta dalle GAN rispetto ad altri algoritmi non riduca il livello di creatività che si può ascrivere alle GAN stesse. [...] la creatività va misurata in funzione del processo, e non dalla qualità del prodotto. [...] la qualità di un prodotto non dipende dalla creatività del processo che ha condotto alla sua creazione. Se una studentessa d'arte realizza un disegno la cui qualità è inferiore a quella di un disegno realizzato da un artista di professione, ciò non chiama in causa la creatività degli studenti d'arte, ma semmai l'inesperienza della studentessa. Analogamente, non possiamo giudicare la creatività di un algoritmo sulla base della qualità del suo output¹²⁶.

Si stanno aprendo molti spunti di riflessione e molte strade. Se la musica è ancora un campo inesplorato, quello delle arti visive, invece, ha dato modo di indagare le capacità artistiche e creative delle GANs in maniera qualitativamente superiore. Esistono, infatti, opere artistiche create proprio attraverso l'uso di questi algoritmi e diversi artisti (artisti generativi) che ne hanno fatto uso. Tutto ciò nel modo dell'arte non solo è stato sperimentato, ma ha avuto un suo riconoscimento.

Christie's e Sotheby's, due delle più importanti ed autorevoli case d'asta al mondo, negli ultimi hanno accolto sempre più artisti che fanno uso di intelligenza artificiale e dato spazio all'utilizzo di strumenti digitali e alla produzione di opere d'arte attraverso questi ultimi (alcuni di questi artisti verranno approfonditi nel successivo paragrafo). Le case d'asta si sono progressivamente adattate alla nuova realtà digitale, lo testimonia il fatto che dal 3 al 10 giugno Sotheby's abbia programmato sul proprio sito la piattaforma *Natively Digital: A Curated Nft Sale*: una vendita di opere di alcuni tra i più importanti artisti del mondo digitale (tra cui Anna Ridler, artista generativa). Lo scopo di queste iniziative, come afferma Sebastian Fahey, Sotheby's Managing Director for Europe, è quello di "ridurre il *gap* tra

¹²⁶ Ivi, cit. p. 163.

l'arte fisica e quella digitale"¹²⁷. Egli aggiunge, inoltre, che: "La competenza curatoriale ci aiuta a comporre una vendita che si basi sul talento degli artisti in un modo non così lontano da una scelta fisica"¹²⁸. Questo permette di comprendere come lo sviluppo e l'esplorazione del mondo digitale siano in continua crescita, fatti confermati dalla vendita e dall'uso di piattaforme online per la commercializzazione di queste opere: una novità ed un rinnovamento per quanto riguarda il mercato dell'arte.

Anche la casa d'asta Christie's conferma l'interesse verso questo nuovo mondo, a maggior ragione essendo stata la prima a vendere un'opera creata dall'intelligenza artificiale¹²⁹. Mariolina Bassetti (Chairman Post-War & Contemporary Art Continental Europe e Chairman Italy) conferma il boom del mercato nell'ambito degli Nft per tutto il 2020 e l'aumento dei prezzi di queste opere d'arte, dichiarando:

Siamo sempre stati all'avanguardia nell'arte e nella tecnologia: in collaborazione con Artory, primo database orientato al mondo dell'arte e leader del settore, Christie's è stata la prima casa d'aste internazionale a registrare vendite di opere su Blockchain, nel 2018 dalla collezione Ebsworth. Inoltre, è stata la prima casa d'aste a vendere un'opera d'arte creata dall'intelligenza artificiale/AI venduta per 432.500 dollari¹³⁰.

Tutto questo ci dimostra quanto questa disciplina abbia raggiunto notevoli sviluppi e quanto ancora ci sia da indagare. Non solo il campo artistico schiude nuove possibilità ancora inesplorate, ma esistono già tuttora artisti ed opere d'arte che hanno fatto uso di queste tecnologie e case d'asta tra le più importanti al mondo che le hanno vendute.

Giunti a questo punto, è opportuno indicare ed approfondire quali siano alcuni tra questi artisti più importanti nel suddetto campo, che ci guideranno e condurranno fino alla fine del nostro percorso. Conosciamo, dunque, più da vicino gli artisti delle GANs.

¹²⁷ Cfr. M. Moro, *Speciale Nft. Le case d'asta*, pubblicato su *Il giornale dell'arte* il 21 giugno 2021, <https://www.ilgiornaledellarte.com/articoli/speciale-nft-le-case-d-asta/136368.html?fbclid=IwAR0oB7cU-ULGOi7aeCUBj10wkyq4sE2gyxWze8HC7y-ri2P72EucozZYrxQ> URL visitato in data 8 settembre 2021.

¹²⁸ *Ibidem*.

¹²⁹ *Ibidem*. In questo caso si parla dell'opera *Portrait of Edward de Belamy*, creata dal collettivo *Obvious* e venduta nel 2018, che verrà trattata in maniera più approfondita nel prossimo paragrafo.

¹³⁰ *Ibidem*.

3.2. *Gan Artists*: gli artisti e l'uso delle Generative Adversarial Networks

3.2.1. *Obvious*

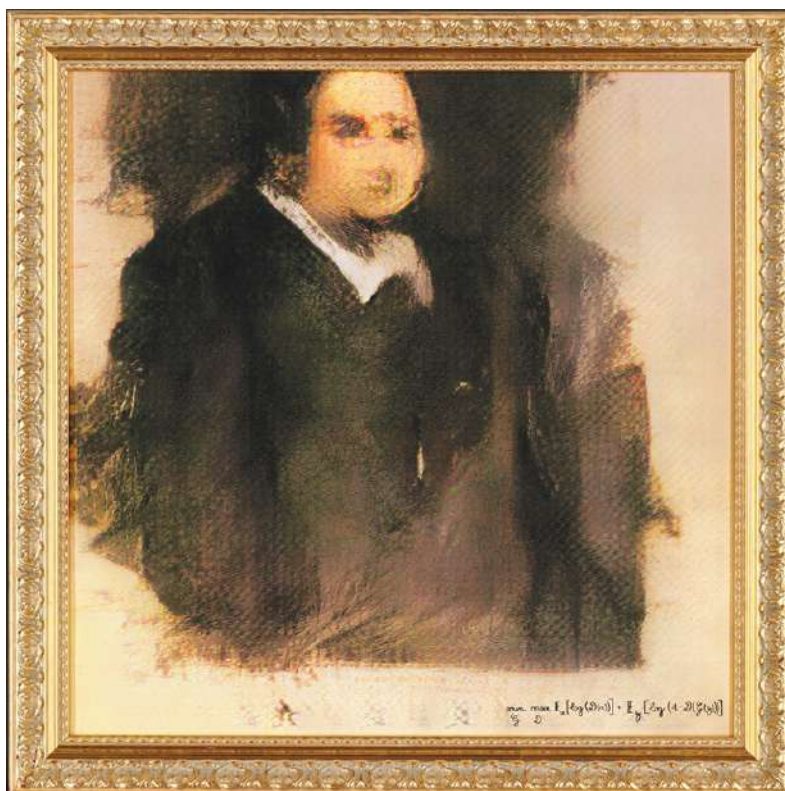


Fig. 19: *Edmond de Belamy, Obvious, 2018. Algoritmi GAN, stampa inkjet su tela, 70x70 cm.*

Gennaio 2018: un ritratto, parte della serie *La Famille de Belamy*, viene acquistato da un collezionista per 10.000 euro. L'evento suscita scalpore, ma non è l'unico. Qualche mese dopo, nel giugno dello stesso anno, un'altra opera, la cui base d'asta oscillava tra i 7 e i 10 mila dollari, viene battuta all'asta da Christie's, a New York, per la somma finale di 432.500 dollari. Si tratta dell'opera *Edmond de Belamy*, facente parte della serie *La Famille de Belamy*. Gli autori invece sono dei giovani francesi, Hugo Caselles-Dupré, Pierre Fautrel e Gauthier Vernier, ricercatori guidati da un comune interesse per il crescente valore dell'AI e del machine learning in ambito artistico, che insieme formano il collettivo *Obvious*.

In questi termini, parrebbe una normale vendita ad opera di una casa d'asta, ma non lo è. Per quale motivo? In questo quadro, in basso a sinistra vi è la firma dell'autore e stavolta non si tratta unicamente dei giovani artisti francesi, bensì di un algoritmo: “**min G max D Ex[log(D(x))] + Ez[log(1-D(G(z)))]**”. È la prima vendita nel mondo dell'arte di un'opera creata dall'intelligenza artificiale. La formula corrisponde alla sezione centrale dell'algoritmo di una GAN e il nome stesso del quadro, *Edmond de Belamy*, è un tributo all'inventore delle GANs Ian Goodfellow (infatti, in inglese il cognome si può letteralmente tradurre con “buon amico” o “buon compagno”, in francese *bel ami*).

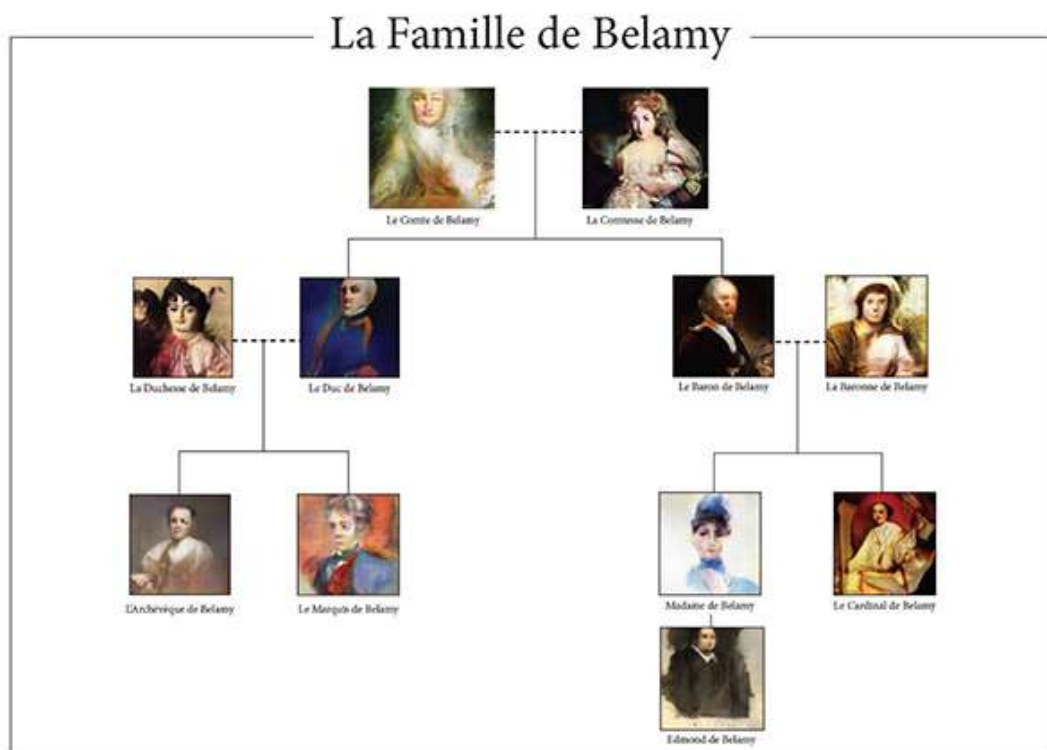


Fig. 20: *La Famille de Belamy*, Obvious, 2018.

Entrambi i quadri citati fanno parte di una serie di “ritratti di famiglia”: 11 ritratti di persone immaginarie creati archiviando 15mila dipinti artistici realizzati tra il XIV e il XX secolo e rielaborati attraverso l'algoritmo, che ha partendo dal set iniziale è riuscito appunto a dare forma a queste opere. I ritratti sviluppati dall'I.A. sono stati incorniciati in cornici d'oro, diventando vere e proprie tele, opere d'arte. Essi, inoltre, sono stati collegati attraverso

un albero genealogico che, come descritto dagli artisti stessi¹³¹ ha diversi significati in base al modo in cui lo si guarda: sul piano verticale riporta i differenti periodi della storia dell'arte, mentre su quello orizzontale le varie ramificazioni rappresentano lo sviluppo dell'intelligenza artificiale nella società. “Alla base di questa serie di ritratti vi è la volontà di offrire un modo nuovo di guardare al nostro retaggio artistico e culturale”¹³². L'attenzione, di nuovo, ricade sul processo e su come nello specifico il processo creativo delle GAN sia fondamentale in questo campo:

Più ancora che nei risultati, tuttavia, l'aspetto più affascinante di questo nuovo modo di fare arte sta nel processo in sé. Lo strumento assume un ruolo più decisivo e centrale nel processo di creazione. Gli algoritmi possono ancora essere paragonati a dei pennelli, ai quali tuttavia viene demandata la fatica del processo creativo, attraverso le schede grafiche ad alta efficienza utilizzate per l'elaborazione¹³³.

Nonostante la caratteristica principale e l'originalità di questo processo sia dovuta al fatto che sono le GAN le protagonisti principali nella creazione dell'opera, ciò non significa d'altra parte che l'artista non dia il suo contributo inventivo a questo processo: si tratta semplicemente di un diverso tipo di creatività, di una nuova forma di processo artistico, in cui la maggior parte del lavoro viene delegato allo strumento, ma l'uomo resta ugualmente parte fondamentale nel raggiungimento del risultato finale. L'artista, infatti, non solo deve programmare adeguatamente l'algoritmo, la scelta del mezzo e far sì che la programmazione segua il giusto processo, ma deve anche scegliere il tema e avere l'ispirazione. Vediamo, quindi, come già accennato nel caso della musica, quanto vi sia un'interazione tra l'uomo e la tecnologia e non una sostituzione totalizzante della macchina rispetto all'uomo.

Gli stessi artisti autori dell'opera affermano come il dibattito sulla creatività delle macchine fosse stato centrale per l'ispirazione e lo sviluppo dell'opera: nonostante per il collettivo francese le GAN rappresentassero uno strumento molto potente, tutto ciò non era sottratto ad equivoci. Nella società, infatti, esistono tuttora fraintendimenti riguardo l'intelligenza artificiale dovuti all'influenza della fantascienza: spesso, infatti, questo

¹³¹ Cfr. *Obvious, La Famille de Belamy* e *i Sogni elettrici di Ukiyo: reinterpretazioni e accelerazioni*, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020, p. 167.

¹³² Ivi, cit. p. 168.

¹³³ Ivi, cit. p. 168.

strumento viene frainteso con la visione distorta fantascientifica di qualcosa dotato di coscienza ed intenzionalità, che non è tuttavia ciò con cui stiamo trattando. Di conseguenza, gli artisti si sono chiesti come fosse possibile trovare un mezzo che fosse abbastanza universale e raggiungesse il maggior numero di persone possibile, per adottare una sorta di linguaggio universale in grado di raggiungere una pluralità di persone proveniente da differenti origini, mentalità, opinioni: questo mezzo era quello artistico. Chiaro, espressivo, tangibile, diretto. Gli artisti francesi spiegano:

Il messaggio doveva essere universalmente comprensibile, ed era necessario che risultasse quanto più semplice possibile. Per raggiungere tale obiettivo abbiamo deciso di lavorare sull'arte così come la concepiva la nostra mentalità occidentale carica di pregiudizi: e cioè sulla ritrattistica classica¹³⁴.

Da qui la raccolta di dipinti, selezionati in base al risultato che si voleva raggiungere, con l'obiettivo principale di dimostrare la potenza degli algoritmi. La chiave di lettura, per rendere ancora più evidente il processo alla base della creazione e per rendere lo spettatore in grado di interpretare il tutto autonomamente, sta nell'algoritmo posto come firma (come si può vedere più chiaramente nel ritratto precedentemente esposto *Edmond de Belamy*). Questo processo creativo è, dunque, simile a qualsiasi altro processo adottato da un artista: a partire da un messaggio e da un'immagine mentale, si sceglie lo strumento più adatto per trasmetterlo. La novità principale sta nello strumento in sé e nel fatto che questi algoritmi portano alla luce il concetto di novità: questi algoritmi, pur ricevendo input identici, riescono a generare di volta in volta diversi dipinti, dando sempre risultati differenti. Ed è proprio in ciò che la creatività umana si rispecchia: come l'uomo difficilmente creerà la stessa identica opera per due volte, ugualmente la macchina.

Le GAN sono strumenti affascinanti che non comprendiamo appieno. Non siamo ancora riusciti a capire le ragioni che si celano dietro alla creazione di queste immagini. Tutto ciò che sappiamo è che questi strumenti sono catalizzatori estremamente potenti della nostra creatività. Ci consentono di creare sulla base di un'idea e di esempi, sostituendo la conoscenza delle tecniche tradizionali della pittura o del disegno con quella dell'apprendimento automatico e della programmazione¹³⁵.

¹³⁴ Ivi, cit. p. 170.

¹³⁵ Ivi, cit. p. 173.

Dunque, grazie all'arte ecco come il dibattito sull'intelligenza artificiale è stato in questo caso portato su un piano diverso rispetto agli equivoci dovuti alla fantascienza. Come abbiamo precisato, un effettivo segnale di accettazione è stato reso ufficiale attraverso il riconoscimento delle case d'aste, in particolare dopo la vendita dell'opera *Edmond de Belamy* attraverso Christie's, ad un prezzo molto più elevato rispetto a quello stimato. Con le parole degli autori:

A partire da quel momento, l'atteggiamento del mondo dell'arte nei confronti di questa nuova tecnologia è drasticamente mutato. Il fatto che una delle istituzioni più prestigiose del mondo dell'arte contemporanea abbia scelto di saggiare la risposta del mercato a questo tipo di opere [...] e il fatto che tale risposta sia stata così sensazionale, ha mutato la visione complessiva dell'intelligenza artificiale da parte del mercato dell'arte¹³⁶.

Ecco prendere forma una versione più realistica e meno fantascientifica dell'intelligenza artificiale come strumento che racchiude molte potenzialità e che può essere indirizzato dall'uomo per diversi scopi (positivi o negativi, con fini di intrattenimento e creativi piuttosto che ingannevoli e potenzialmente pericolosi). Sicuramente, si è giunti ad una reazione da parte della società, ad una visione più realistica che fantascientifica e ad una consapevolezza di trovarsi di fronte ad una vera e propria rivoluzione tecnologica: senza dubbio una grande esperienza.

Come tutte le grandi rivoluzioni, lo strumento non è stato esente da critiche: alcuni hanno considerato l'intelligenza artificiale “uno strumento pericoloso che rischia di annientare gli artisti”¹³⁷ o “qualcosa che contraddice la natura umana”¹³⁸. Inizialmente addirittura il fatto che le macchine potessero partecipare al processo creativo non era stato propriamente accettato da molti artisti, con la convinzione che questo tipo di arte non dovesse essere promossa, o addirittura che non fosse effettivamente arte (forse, sempre per la manchevole ed erronea comprensione degli strumenti utilizzati).

¹³⁶ Ivi, cit. p. 176.

¹³⁷ Ivi, cit. p. 177.

¹³⁸ Cfr. Ivi, p. 178.

Certo è che anche dal momento in cui il mercato dell'arte ha riconosciuto l'opera, in un modo o nell'altro il dibattito non è mancato. E qualcosa di importante e rivoluzionario fa parlare di sé, nel bene o nel male.

Un'ulteriore fonte di critica venne possiamo dire dall'interno della Gan art stessa: come nel mondo dell'arte più "ordinario" e "canonico" anche in questo caso vi possono essere una sorta di "gelosie" e "competizioni". Nel momento in cui vi fu l'asta e la vendita dell'opera da parte di Christie's, l'artista generativo Mario Klingemann criticò la mancanza di *background* artistico del collettivo *Obvious* e il fatto di aver usato un algoritmo già esistente (usato, infatti, precedentemente da un altro artista, Robbie Barrat¹³⁹). Pur riconoscendo i meriti a Barrat e ringraziandolo dimostrandosi riconoscenti, *Obvious* in quell'occasione rivendicò l'attenzione ricevuta da Christie's come potenzialmente utile per l'impatto mediatico e per l'effetto che avrebbe avuto sull'intero panorama artistico dell'arte generativa. Una fama, dunque, in qualche modo condivisa¹⁴⁰.

$$\min_G \max_D \mathbb{E}_x [\log(D(x))] + \mathbb{E}_z [\log(1 - D(G(z)))]$$

Fig. 21: l'algoritmo usato da *Obvious*, posto come firma in basso a destra alle opere de *La Famille de Belamy*.

3.2.2. Mario Klingemann

¹³⁹ Si vedrà in maniera più approfondita nei paragrafi successivi la controversia con l'artista Robbie Barrat per quanto riguarda l'uso dell'algoritmo nella creazione dell'opera *Edmond de Belamy*.

¹⁴⁰ Per altre opere di *Obvious* si veda: <https://obvious-art.com>.



Fig. 22: *Memories of Passersby I*, Mario Klingemann, 2018.

Marzo 2018, Londra. La casa d'asta Sotheby's vende un'altra opera appartenente alla GAN art, creata attraverso l'uso dell'intelligenza artificiale. Stavolta l'artista è il tedesco Mario Klingemann, uno dei pionieri nel campo dell'intelligenza artificiale e delle reti neurali, che lavora nel campo dell'apprendimento informatico e del mondo dell'arte da più di 20 anni.

Si tratta dell'opera *Memories of Passersby I* e viene acquistata per la cifra di 40 mila sterline. La vera differenza in questo caso, rispetto alle opere citate precedentemente, è la concezione dell'opera d'arte: la vera opera non è, infatti, quella che viene prodotta dall'algorithm, bensì l'algorithm in sé e l'artista è tale proprio in quanto ideatore dell'algorithm¹⁴¹.

Nello specifico per quanto riguarda l'opera, si tratta di un'installazione in cui il sistema informatico è contenuto in una console realizzata a mano in legno di castagno, che dall'aspetto ricorda una sorta di antica radio, ma al suo interno ospita il “cervello”

¹⁴¹ Cfr. IArte, *Memories of Passerby: l'algorithm che sogna volti*, pubblicato su *IA arte, Nuove frontiere tra arte e Intelligenza Artificiale*, il 27 maggio 2019, <http://iarte.altervista.org/memories-of-passerby-lalgorithmo-volti/> URL visitato in data 9 settembre 2021.

dell'intelligenza artificiale. Questa console contenente il meccanismo, che elabora e produce continuamente i volti, è collegata attraverso due fili a due schermi di grandi dimensioni. In tempo reale, in un flusso costante e ininterrotto, le reti neurali producono sugli schermi i volti, a partire da un training set in cui sono stati raccolti dipinti che vanno dal XVII al XIX secolo. L'algoritmo, basandosi sul set di immagini, produce attraverso il generatore i ritratti e migliora continuamente attraverso il lavoro del discriminatore.

La questione interessante nel caso di quest'opera è che i ritratti vengono prodotti in tempo reale davanti agli occhi dello spettatore, seppur per pochi istanti. Inoltre, i ritratti creati attraverso l'uso delle reti neurali sono unici ed originali, continuamente diversi e generanti un continuo effetto di sorpresa.

In maniera diversa dalle altre opere di arte generativa, la caratteristica peculiare in questo caso sta nel fatto che l'opera non contiene un database: il "cervello" dell'I.A. crea in continuazione ritratti inediti, non si tratta di qualcosa di prestabilito ma dell'interpretazione che la macchina dà rispetto al suo stesso output: le immagini, così, non si ripetono mai, le GAN producono una sequenza infinita. Siamo dunque "un passo avanti" nella storia dell'intelligenza artificiale, proprio perché l'opera rappresenta "un soggetto creativo indipendente"¹⁴².

L'autore ha addestrato la rete neurale attraverso i dipinti, in seguito il processo di apprendimento è stato accelerato tramite un'applicazione che permette alla macchina di stabilire delle preferenze estetiche. I ritratti diventano così "inquietanti interpretazioni del volto umano"¹⁴³, di una bellezza molto particolare, a tratti piacevole, a tratti disturbante.

Possiamo tuttavia dire che:

Talvolta le immagini si fondono in disposizioni astratte di pixel, prodotte dallo sforzo della macchina inteso a creare un nuovo ritratto. Allo spettatore, *Memories of Passersby I* offre un'esperienza ipnotica, un'occasione di osservare un'intelligenza artificiale mentre "pensa" in tempo reale e di ammirare ritratti assolutamente unici, che non vengono registrati e non si ripetono mai¹⁴⁴.

¹⁴² Cfr. M. Klingemann, *Memories of Passersby I*, 2018, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020, cit. p. 78.

¹⁴³ *Ibidem*.

¹⁴⁴ *Ibidem*.



Fig. 23: *Uncanny Mirror*, Mario Klingemann, 2018.

Noi stessi, la nostra immagine. Per quanto il mondo fuori possa essere fonte di dubbio (non sappiamo mai chi abbiamo davanti fino in fondo, non abbiamo la certezza totalizzante di ciò che è fuori di noi), se c'è qualcosa di certo, di cui abbiamo una ferma consapevolezza, siamo noi stessi. Ci guardiamo in uno specchio e vediamo la nostra immagine riflessa, sappiamo riconoscerci, questo avviene fin dalla prima infanzia, fin dai primi mesi di vita. Ma cosa succederebbe se guardandoci allo specchio non vedessimo più la nostra immagine, ma qualcos'altro di simile e al tempo stesso diverso? Non sarebbe forse qualcosa di estremamente perturbante?

Questo è ciò che accade con un'altra opera di Mario Klingemann: *Uncanny Mirror* (2018). Anche in questo caso, l'opera funziona attraverso l'uso dell'intelligenza artificiale. In questo caso però, non solo lo spettatore vede in tempo reale il realizzarsi dell'opera, ma addirittura interagisce con essa, diventa egli stesso parte integrante dell'installazione. Il sistema di reti neurali stavolta apprende in maniera interattiva: mentre lo spettatore guarda l'opera, l'I.A. analizza i suoi parametri biometrici facciali, la postura, i gesti e crea ritratti in maniera quasi istantanea, usando il materiale appreso. La macchina dà come risultato il suo modo di vedere lo spettatore (che non è

ovviamente iperrealistico, essendo una rielaborazione di diverse immagini che la macchina “vede”).

In questo caso, per l’autore: “Il pubblico costituisce «un’interessante banca dati» i cui input implicano un margine di rischio e di imprevedibilità. La sua installazione apprende costantemente, assimilando i dati di tutti coloro che guardano all’interno di questo inedito specchio”¹⁴⁵.

Dunque, stavolta, i materiali delle GAN non sono più i dipinti di secoli precedenti, il training set non è costituito da opere di artisti rielaborate dalla macchina, bensì il training set è costituito dai volti delle persone e degli spettatori in tempo reale. La macchina osserva, apprende, rielabora e ogni volta dà vita ad un volto nuovo che ha qualcosa dei volti precedenti. Interazione uomo-macchina: lo spettatore stesso è parte integrante del costituirsi e del continuo realizzarsi dell’opera. Il risultato, tuttavia, non sempre è “rassicurante”, come suggerisce il titolo stesso dell’opera.

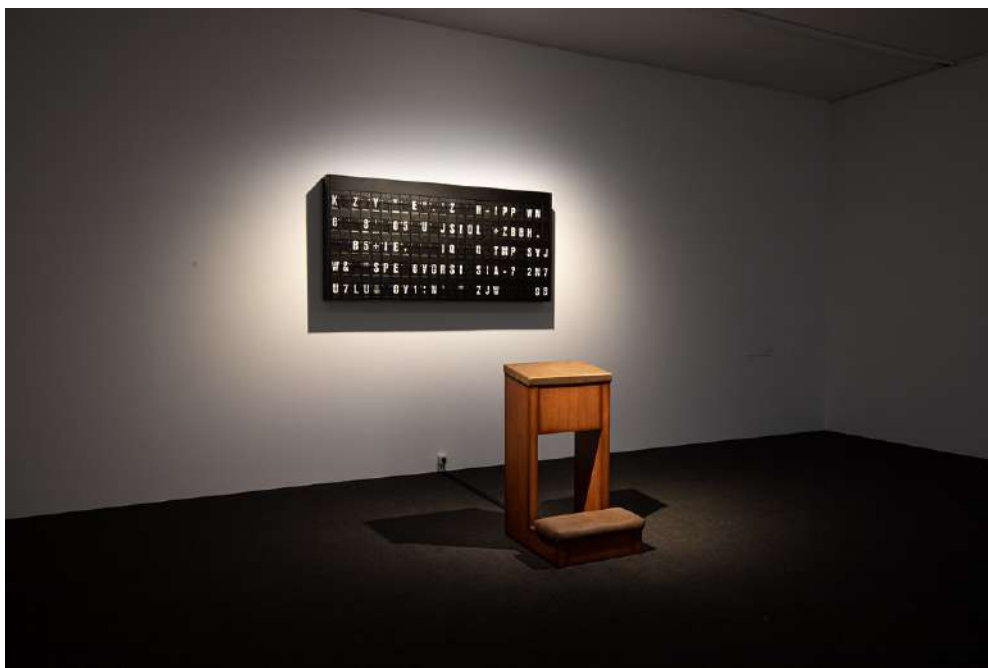


Fig. 24: *Appropriate response*, Mario Klingemann, 2020.

¹⁴⁵ Ivi, cit. p. 97.

“Words are probably the most powerful tools available to humankind. Words can make people do things, can change their lives”¹⁴⁶.

Un inginocchiatoio, un display con una sequenza di lettere che, come nel caso precedente, interagisce con lo spettatore, ma stavolta non più attraverso immagini, bensì attraverso frasi e aforismi. Anche nel caso di *Appropriate Response* l’opera è interattiva e riflette sul significato del rapporto tra l’uomo e l’intelligenza artificiale.

Le lettere cambiano di continuo e una volta che lo spettatore si inginocchia sull’apposita postazione, esse danno vita ad una frase o aforisma. L’opera interagisce in qualche modo e “risponde” allo spettatore con delle frasi sul display, ma l’altra caratteristica peculiare, oltre all’interazione tra lo spettatore e la macchina, è che nessuno spettatore assisterà alla stessa frase. Ogni frase è, infatti, diversa: ad ogni spettatore un aforisma, una “perla di saggezza”: anche se brevi le frasi possono essere d’aiuto, d’ispirazione, ogni frase porta un suo significato (che verrà interpretato dallo spettatore), è proprio questo il potere delle parole. Il fatto che esse possano essere fonte d’ispirazione.

Klingemann spiega: “On the one hand we fear AI but we also have hopes that it might help us to solve some problems. That balance between hope and fear is closely related to religious experience, so I felt that kneeling was very fitting”¹⁴⁷.

Per questo la scelta dell’inginocchiatoio, per dare una sorta di “legame sacro”, di ritualità, di continuità tra presente e passato.

Paura verso l’intelligenza artificiale e al tempo stesso speranza, possibilità di interagire con essa: questo racchiude l’opera di Klingemann. Lo spettatore è il protagonista diretto, fornendo significato ad ogni frase che compare.

In questo caso la rete neurale è stata addestrata dall’autore attraverso l’uso di enciclopedie, poesie, libri di ricette e citazioni online; le lettere vengono poi disposte

¹⁴⁶ Cfr. Mario Klingemann su Artsy.net, <https://www.artsy.net/artwork/mario-klingemann-appropriate-response-1> URL visitato in data 10 settembre 2021. Tr.: "Le parole sono probabilmente gli strumenti più potenti a disposizione dell'umanità. Le parole possono far fare alle persone cose, possono cambiare la loro vita".

¹⁴⁷ *Ibidem*, Tr.: “Da un lato temiamo l'IA, ma speriamo anche che possa aiutarci a risolvere alcuni problemi”, spiega Klingemann. “Quell'equilibrio tra speranza e paura è strettamente legato all'esperienza religiosa, così ho sentito che inginocchiarmi era molto appropriato”.

dalla rete neurale attraverso un display *split flap*. Grazie a quest'opera nel 2020 Klingemann ha vinto il premio Menzione d'Onore al *Prix Ars Electronica*¹⁴⁸.

3.2.3. Anna Ridler

“La GAN conosce il mio stile. Chiunque vi possa accedere può accedere anche a un pezzetto di me [...]. L'intelligenza artificiale è in grado di copiare e di suggerire, ma in ultima analisi tutto inizia e finisce con me”¹⁴⁹.

È così che l'artista londinese Anna Ridler, vincitrice nel 2018 del programma europeo di *media art* e del *Dare Art Prize* nel 2019, definisce il proprio lavoro. L'affermata artista, che ha già esposto in diverse gallerie e musei (tra cui *Tate Modern*, *V&A Museum* ed *Ars Electronica*), lavora utilizzando l'Intelligenza Artificiale per creare vere e proprie narrazioni. Per fare ciò, al pari degli altri *Gan artists*, vengono utilizzati dati preesistenti (*training set*) e un algoritmo attraverso cui l'intelligenza artificiale lavora. Il processo è simile ai casi presentati precedentemente: la GAN viene addestrata su una serie di immagini (*training set*) la quale a sua volta viene utilizzata per creare una versione realistica di un'immagine, che potrebbe verosimilmente assomigliare al *training set* in questione.

Fino a qui nulla di strano: è la stessa attività delle due reti antagoniste generative, che lavorano tra loro e migliorano attraverso la competizione per generare immagini. Ma la novità, in questo caso, sta nella scelta del *training set* e delle immagini attraverso cui vengono addestrate le reti neurali:

Occorre cautela nel creare un *training set*. Se le dimensioni sono eccessive, se le immagini sono oltremodo numerose, i risultati saranno perfetti – e andranno perdute le imperfezioni e le particolarità che rendono interessante il mezzo espressivo [...] se invece il *training set* è troppo ridotto, non conterrà informazioni sufficienti e risulterà confuso, non producendo alcun risultato o limitandosi a ripetere all'infinito una o due variazioni del *training set* stesso¹⁵⁰.

¹⁴⁸ Per altre opere di Klingemann si veda: <https://quasimondo.com>.

¹⁴⁹ Cfr. A. Ridler, *Set di dati e decadenza: Fall of the House of Usher*, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020. Cit. pp. 126-127.

¹⁵⁰ Ivi, cit. p. 113.

Per l'artista, dunque, è importante il processo di generazione stesso e, soprattutto, il fatto che vi possa essere una traccia di quel processo nel risultato finale. Se le imperfezioni rappresentano quella traccia, allora le imperfezioni dovranno rimanere presenti nell'immagine generata. Nonostante la tecnologia punti spesso al realismo più estremo e nonostante l'obiettivo sia spesso quello di ridurre al minimo gli errori presenti nell'immagine risultante, a volte sono proprio quegli errori a rappresentare una traccia del processo seguito dall'artista.

Nel lavoro e nella riflessione di Anna Ridler si cela una domanda: siamo assolutamente certi che la perfezione e il realismo (o l'iperrealismo) siano il risultato più auspicabile? La mancanza di errori o di imperfezioni in un'opera creata attraverso l'uso di mezzi digitali è necessariamente indice della qualità di questo lavoro? Per l'artista la risposta è negativa: le imperfezioni sono proprio la traccia dell'umanità presente nell'opera, dell'unione tra il lavoro dell'uomo e quello della tecnologia. È, infatti, anche attraverso la presenza di errori che l'uomo resta, in qualche modo, presente nell'artificialità.



Fig. 25: *Fall of the House of Usher*, Anna Ridler, 2017.

Nell'opera *Fall of the House of Usher* vediamo il realizzarsi di questo impianto teorico: inizialmente la Ridler inserisce un ritratto creato manualmente nella GAN, a sua volta creata da lei. Nota, tuttavia, che la GAN non si limita a ridisegnare l'opera nello stesso identico modo in cui l'aveva disegnata l'artista (in maniera, dunque, iperrealistica), bensì aggiungendo un elemento in più (in questo caso uno scialle). Vediamo, appunto, come gli errori rendano più evidente la natura artificiale dell'opera e siano frutto di sovrapposizioni generate dalla GAN stessa. Ecco le tracce del processo creativo di cui si parlava precedentemente.

Ma vi è un'ulteriore novità che è opportuno sottolineare: in questo caso il *training set* è composto da disegni, creati manualmente dall'artista. La GAN non viene, infatti, addestrata attraverso immagini o foto iperrealistiche, bensì attraverso i disegni che l'artista stessa ha creato.

L'ispirazione di quest'opera nasce da un racconto di Edgar Allan Poe del 1839, basato sulla storia di una casata decaduta. Il racconto, che fu precedentemente rappresentato attraverso l'animazione digitale, viene rappresentato dalla Ridler attraverso dei disegni. Questi disegni creati manualmente formano il *training set*, che a sua volta sarà la base su cui la GAN lavorerà. Il risultato sarà una commistione tra arte e tecnologia, che darà vita ad un'opera del tutto originale.

L'artista spiega la genesi dell'opera affermando: “Per realizzare la mia animazione ho preparato un training set di 200 disegni ispirandomi alla versione del 1929 de *La caduta degli Usher*, in modo da insegnare alla GAN a disegnare il mio stile”¹⁵¹.

Dunque, possiamo notare come in questo caso il disegno sia una parte fondamentale per il processo di apprendimento della macchina, che richiede diverso tempo, ma lascia la fondamentale traccia dell'artista nel risultato finale. In questo modo, inoltre, vi è un tentativo di avere una sorta di controllo sull'opera: pur lasciando autonomia alla macchina nel rielaborare il materiale di partenza, all'autrice sembra che usando questa tecnica vi sia un controllo maggiore sul risultato finale. Tuttavia, quella del controllo è solo un'illusione: il

¹⁵¹ Ivi, cit. p. 118.

risultato finale dei fotogrammi sarà sempre imprevedibile. La Ridler evidenzia come gli errori presenti nei disegni iniziali vengano amplificati dalla rete neurale, facendo realizzare anche a chi in primo luogo li ha prodotti, qualcosa di cui non era nemmeno totalmente consapevole (come, ad esempio, il fatto che gli occhi e le sopracciglia dei soggetti vengano disegnati in maniera simile dall'artista e vengano di conseguenza confusi durante la rielaborazione della macchina, che confonde i due elementi)¹⁵².

Di conseguenza, seppur vi sia una sorta di “volontà di controllo” da parte dell'artista sul risultato, esso appare alla fine del processo più incontrollato che mai: “Il risultato è opera mia e al tempo stesso non lo è – è riconoscibile come mio, ma non è qualcosa che sarei stata in grado di realizzare da sola”¹⁵³.

Sorgerebbe spontaneo chiedersi dove risieda a questo punto la creatività dell'opera, nell'artista che disegna e dà il materiale di input iniziale, o nella macchina che elabora il risultato finale? Forse la peculiarità sta proprio in questo: nelle potenzialità che schiude la commistione tra i due. Il modo che ha la GAN di creare l'immagine è sicuramente diverso da quello dell'artista in carne ed ossa, la sua azione non si può definire propriamente quella di disegnare, anche se a tutti gli effetti il risultato potrebbe dare quest'idea. Eppure, il legame tra uomo e tecnologia in questo caso risulta molto più intenso rispetto a quello che si instaurerebbe in un semplice addestramento della rete neurale attraverso un set preesistente di immagini: questo accade proprio perché la GAN riporta lo stile dell'artista e, come afferma la Ridler “chiunque vi possa accedere, può accedere anche a un pezzetto di me”¹⁵⁴.

Vediamo come, adottando il punto di vista dell'autrice londinese, il lavoro dell'intelligenza artificiale non sia solo qualcosa di freddo e meccanico. Nell'opera vi è davvero “un pezzetto di lei”: in questo tipo di arte digitale non vi è la volontà di rendere quel mondo reale, spesso così imperfetto e confuso, qualcosa completamente privo di imperfezioni, meccanico, sterile e artificiale (per l'appunto, “robotico”): a partire dall' algoritmo e dal lavoro delle GAN si introduce anche l'umanità. Ecco come l'intelligenza

¹⁵² Cfr. Ivi, p. 119.

¹⁵³ Ivi, cit. p. 126.

¹⁵⁴ *Ibidem*.

artificiale arrivi così a conoscere un nuovo sviluppo: riportando con sé, in maniera diretta, la mano e il tocco di chi la guida.

Questa “umanità” della macchina caratterizza anche altri lavori dell’autrice, come ad esempio l’opera *Alice and Bob* (2017).

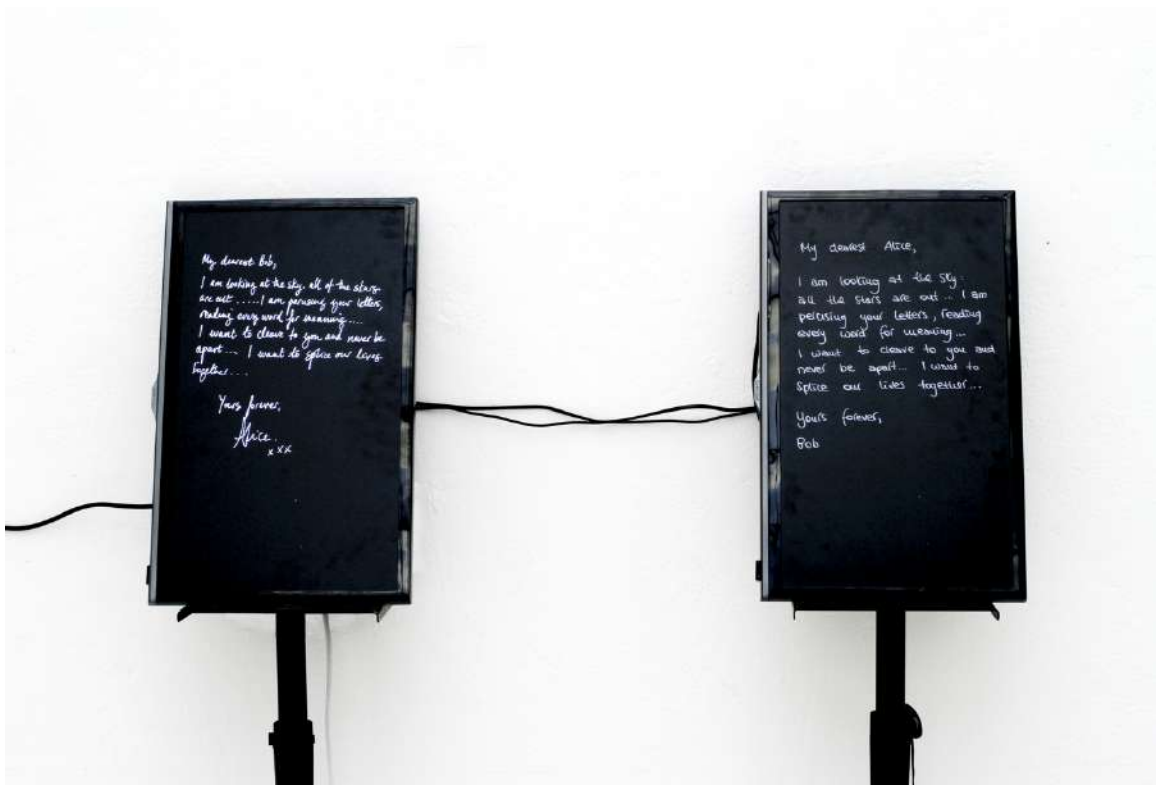


Fig. 26: *Alice and Bob*, Anna Ridler, 2017.

Anche in quest’opera vediamo l’intrecciarsi di quella dimensione umana di cui parlavamo precedentemente, che caratterizza lo stile della Ridler.

L’opera¹⁵⁵ è basata sul potenziale dei fotoni ed è composta da due schermi che utilizzano i dati di un computer quantistico. L’installazione, attraverso una serie di lettere d’amore mostrate sui due schermi, racconta nel corso di tre atti la relazione tra due personaggi fittizi, Alice e Bob, complicata da alcune apparizioni di un terzo personaggio, chiamato Eve.

¹⁵⁵ Cfr. Anna Ridler, *Alice and Bob*, 2017, <http://annaridler.com/quantum-computing-art> URL visitato in data 12 settembre 2021.

Le lettere scritte prendono ispirazione (parole e frasi) da estratti e titoli di alcuni articoli scientifici quantistici, apparendo sugli schermi come delle vere e proprie frasi scritte a mano, con tutta l'intimità che potrebbe avere una vera corrispondenza tra due personaggi reali. Ancora una volta quelli che potrebbero essere sterili algoritmi si trasformano in questo modo in poesia e vengono ricontestualizzati attraverso l'arte in qualcosa di prettamente umano, come una relazione sentimentale tra due persone.

Possiamo notare nuovamente il sorprendente intreccio tra scienza, arte e letteratura e la capacità dell'artista di rendere, in qualche modo, l'umanità nella tecnologia, la presenza dell'uomo attraverso l'intelligenza artificiale.

Anna Ridler in tutte le sue opere raccoglie informazioni per creare delle connessioni e raccontare, attraverso l'arte e la tecnologia, delle narrazioni. Si creano in questo modo vere e proprie storie che vengono presentate allo spettatore attraverso l'ausilio delle GAN e dell'intelligenza artificiale, siano esse storie fittizie o reali. Ad esempio, per quanto riguarda narrazioni più realistiche e attinenti al mondo attuale, l'artista ha indagato anche i tabù della società, più precisamente il controllo che le grandi aziende hanno su ciò che dovrebbe qualificarsi come un termine tabù, chiedendosi se queste decisioni riflettano la cultura *mainstream*. Il metodo è stata la ricerca di quali parole non fossero riconosciute dalla correzione automatica di un telefono e indagando così come funzioni la correzione automatica su alcuni sistemi operativi (in *No Replacements Found*, 2015)¹⁵⁶. In questo modo i risultati dell'autocorrezione possono rivelare quelle parole considerate ancora tabù e, di conseguenza, l'implicito conservatorismo di alcune grandi aziende. Oppure ancora, per restare in tema di attualità, analizzando ed utilizzando i dati di *Wikileaks* è stata indagata dall'artista la posizione delle donne in Arabia Saudita (in *Wikileaks: Women*, 2016)¹⁵⁷¹⁵⁸.

¹⁵⁶ Cfr. Anna Ridler, *No Replacements Found*, 2015.

<http://annaridler.com/no-replacements-found> URL visitato in data 12 settembre 2021.

¹⁵⁷ Cfr. Anna Ridler, *Leaked Documents: Wikileaks*, 2015-2017.

<http://annaridler.com/leaked-documents>, video on Vimeo <https://vimeo.com/165036571> URL visitato in data 12 settembre 2021.

¹⁵⁸ Per altre opere di Ridler si veda: <http://annaridler.com/works>.

3.2.4. Robbie Barrat



Fig. 27: *Infinite Skulls*, Robbie Barrat/Ronald Barrot, 2019, olio su tela, stampa UV su plexiglass.

Febbraio 2019, Parigi. Al pubblico francese viene presentata una mostra, ma non si tratta di una qualsiasi mostra d'arte. Cosa differenzia quest'esposizione dalle altre? Un particolare: le opere sono prodotte dalla collaborazione di due persone e, più nello specifico, di due persone e una macchina. Si tratta di un pittore francese già noto, Ronan Barrot, e di un giovane artista prodigio americano che sperimenta già da qualche anno l'uso dell'intelligenza artificiale, Robbie Barrat. Da questa collaborazione nascono gli *Infinite Skulls*¹⁵⁹. Ma, nello specifico, qual è la genesi di quest'opera, così particolare?

Tutto nasce in principio dall'abitudine del pittore Barrot di dipingere con gli avanzi di colore, una volta terminati i suoi quadri, alcuni teschi. Una particolare abitudine che per l'artista "consacra" la fine del processo creativo e rappresenta una sorta di "pulizia", anche

¹⁵⁹ Cfr. a riguardo Jason Bailey, *AI Artist Robbie Barrat And Painter Ronan Barrot Collaborate On "Infinite Skulls"*, pubblicato su *Artnome* il 6 febbraio 2019. <https://www.artnome.com/news/2019/1/22/ai-artist-robbie-barrat-and-painter-ronan-barrot-collaborate-on-infinite-skulls> URL visitato in data 12 settembre 2021.

materiale, dello strumento. Adottando questa abitudine nel corso degli anni, il francese arriva a realizzare numerosi teschi, generando migliaia di immagini. Da qui “l’incrociarsi di destini” che darà vita alla collaborazione con Barrat. Ecco nascere così l’innovativa idea che darà vita ad un’opera del tutto nuova: perché non usare queste migliaia di immagini come *training set* di una GAN? L’artista americano, che già aveva sperimentato l’uso di queste tecnologie, vede i dipinti come una possibilità di sperimentare ulteriormente la creatività delle macchine.

Nello specifico questo lavoro si divide in due fasi di elaborazione, chiamate dagli autori rispettivamente Epoca 1 ed Epoca 2. Nella prima fase il *training set* è composto dai teschi di Barrot, che risultano molto simili tra loro, con simili forme, dimensioni, orientamento. Tuttavia, in questa fase, avendo i teschi dimensioni e forme quasi identiche, la GAN riproduce in maniera abbastanza verosimile i teschi del *training set*, dando come risultato un’opera sicuramente buona, ma non esattamente creativa ed originale, essendo tutt’al più una copia dei teschi disegnati da Barrot. Si passa, così, all’Epoca 2.

Nella seconda fase Barrat riprende ugualmente i teschi del pittore francese, ma modificandoli: ne cambia dimensioni e orientamento, rotazione e prospettiva. In questo modo il numero aumenta, passa da 500 a 1700. Ecco l’emergere della creatività e dell’originalità del lavoro, ecco l’emergere del “punto di vista” e della prospettiva della macchina.

Afferma Barrat a riguardo: “Utilizzo lo stesso modello, ma il numero dei teschi contenuti nel *training set* passa da 500 a 1700. E i risultati sono veramente ottimi. Vengono fuori immagini davvero strane, che non ci si aspetterebbe. Si capisce che sono teschi, ma non risultano familiari”¹⁶⁰.

L’artista dà sfogo così alla propria creatività nel modificare le immagini di partenza, presenta alla macchina i teschi rovesciati o stirati. La macchina a sua volta diventa creativa dando vita ad immagini simili alle originali, ma rielaborate sotto una diversa prospettiva. Il teschio, tradizionalmente, è “il simbolo per eccellenza della *vanitas* della vita e dell’arte stessa”¹⁶¹, ma in questo caso anche il significato di questo simbolo cambia, perché dalla prospettiva della macchina non viene visto come l’uomo tendenzialmente lo vede: ciò che

¹⁶⁰ Cfr. A. Barale, *A hideaway in the wastelands: nuove sfide filosofiche dell’AI art*, in A. Barale, *Be my Gan*, Jacobo, Milano, 2020. Cit. p. 217.

¹⁶¹ *Ibidem*.

per la macchina è familiare non lo è per l'uomo. Non più semplici teschi, non più simbolo di vanità della vita, bensì elemento quasi di gioco. Non si parla più di *vanitas* della vita o di “morte dell'arte”, ma di sviluppo infinito delle capacità e delle possibilità dell'arte: infinito come infiniti sono i teschi prodotti dalla GAN.

Di nuovo vediamo, grazie alle sperimentazioni di questi artisti, come sia possibile la collaborazione tra intelligenza artificiale e tocco artistico della mano e della mente umana: con gli *Infinite Skulls* collaborano un pittore, un artista che si avvale dell'uso di tecnologie digitali e una GAN. Quest'ultima rende l'uomo partecipe di uno “sguardo” diverso, di un “vedere” che, pur non riuscendo ad avere una vera e propria comprensione di ciò che le si sottopone (comprensione tipica dell'essere umano), riesce tuttavia ad accedere ad un livello più complesso, un livello a cui l'uomo non riuscirebbe, invece, ad accedere. In questo modo grazie all'interazione tra potenzialità della macchina e potenzialità dell'uomo si schiudono nuove possibilità:

Cambiano le regole del vedere. Il riconoscimento di una GAN (se di “riconoscimento” si può parlare) non è il riconoscimento umano. [...] Proprio questo, però, rende particolarmente interessante il loro processo di elaborazione, in quanto traduzione – estremamente incompleta e problematica – di una lingua nell'altra. Attraverso lo sguardo “altro” della GAN sembrano prefigurarsi nuove modalità del vedere *per noi*¹⁶².

Si tratta quindi di una collaborazione continua, di uno scambio continuo tra macchina e uomo. Barrot crea i teschi, Barrat li modifica, la GAN genera a sua volta altre immagini che possono essere, alla fine, modificate ulteriormente dall'artista. Così l'artista affida la sua opera a qualcosa su cui non ha il controllo: le sue infinite trasformazioni.

Questo accade anche in altre opere di Barrat, come *Landscape Series* (2018). Stavolta il *training set* scelto dall'artista non viene creato manualmente, ma si compone di opere scelte da librerie *opensource* come Wikiart (la creazione di questa serie avviene, tra l'altro, prima dell'incontro con Barrot e della mostra in collaborazione a Parigi). In maniera simile a quella vista precedentemente, i paesaggi creati in primo luogo dalla GAN sono fin troppo realistici, ma nel progredire dell'addestramento delle reti neurali, ecco schiudersi l'originalità: i

¹⁶² Ivi, cit. p. 222.

paesaggi si modificano ed emergono nuove prospettive. Compaiono così linee diverse delle montagne, tronchi fluidi, colori più scuri.

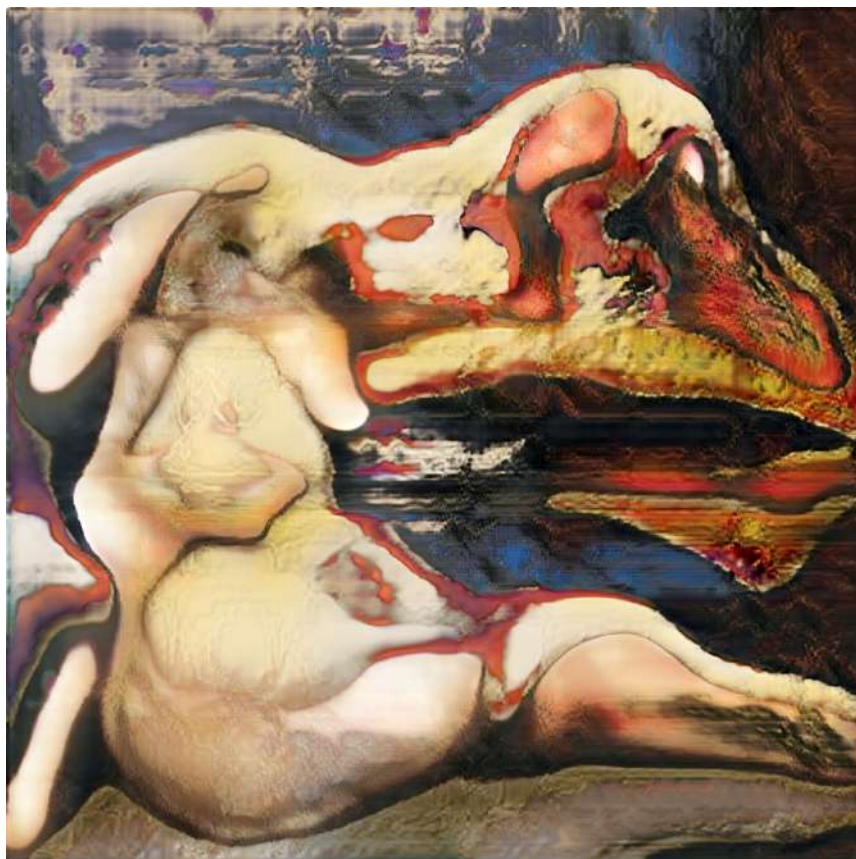


Fig. 28: *AI Generated Nude Portrait #1*, Robbie Barrat, 2018, immagine digitale prodotta da GAN.

In maniera simile ai *Landscapes*, nel 2018 Barrat crea dei dipinti generati da una GAN usando come *training set* immagini prese ugualmente da Wikiart, scegliendo però di prediligere i ritratti, i nudi di persone: nascono così gli *AI Generated Nude Portraits*.

La rete neurale viene addestrata da Barrat attraverso questo *training set*, ma in questo caso i risultati sono tutt'altro che realistici: la macchina, non riuscendo ad apprendere le caratteristiche umane principali presenti nei ritratti e a rielaborarli, dunque, adeguatamente, cade in una sorta di “minimo locale” in cui le figure risultanti sono simili a macchie di carne generiche. Sembra che, di nuovo, la macchina ci sottoponga la sua prospettiva, questa volta su come vede le persone.

Il lavoro venne postato per la prima volta nell'aprile 2018 sulla piattaforma *SuperRare*: Barrat venne selezionato, diventando uno dei primi *Gan artists* ad essere pubblicato da questa piattaforma nello specifico e, più in generale, come uno dei primi esponenti di questo tipo di arte. Barrat all'epoca stava già sperimentando le reti neurali e, nello specifico, si stava specializzando sui ritratti del XIX secolo.

Come possiamo vedere in Fig. 27, nella prospettiva della macchina quelli che in origine erano ritratti di corpi umani perdono la loro "umanità" e diventano quasi figure aliene, difficilmente distinguibili nel risultato finale. Stavolta a partire dal training set, le figure vengono quasi totalmente distorte, rimangono macchie di colore indistinte, lasciando luogo a qualcosa che, se fosse umano, porterebbe con sé sicuramente qualcosa di sinistro.

In seguito al *AI Generated Nude Portrait #1* vennero elaborate altre opere di stampo simile, alcune delle quali esposte anche da Christie's in occasione del primo Tech Summit a Londra, il 17 luglio 2018¹⁶³. In quell'occasione *SuperRare* venne invitato ad aggiungere delle "borse regalo" per gli oltre 300 partecipanti e, con l'intento di proporre qualcosa di originale ed innovativo, che includesse l'emergente arte digitale, decise di collaborare con alcuni artisti di questo campo, tra cui, per l'appunto, Robbie Barrat. In occasione di questo evento l'artista creò gli *AI Generated Nude Portrait #7*. Le opere di Barrat vennero incluse nelle borse regalo da dare ai partecipanti, ma ai tempi l'arte digitale e la cryptoart non erano ancora adeguatamente valorizzate: non vi era la consapevolezza dell'importanza che avrebbe acquisito questo fenomeno nel mondo dell'arte da lì a pochi anni, culturalmente ed economicamente parlando.

Poco più di tre mesi dopo questo evento, infatti, il nome dell'artista (ai tempi ancora sconosciuto e di giovane età – aveva appena 18 anni) sarebbe diventato conosciuto internazionalmente attraverso il nome del collettivo Obvious: l'algoritmo creato da Barrat sarebbe stato riutilizzato dal trio francese per la prima opera creata dall'intelligenza artificiale venduta da Christie's a 432.000 dollari. In quell'occasione¹⁶⁴ Barrat non venne menzionato

¹⁶³ Cfr. a riguardo: *The Lost Robbies*, pubblicato su *SuperRare Editorial*, il 5 settembre 2020. <https://editorial.superrare.com/2020/09/05/the-lost-robbies/> URL visitato in data 14 settembre 2021.

¹⁶⁴ Per approfondimenti circa questa controversia, cfr. Tom Simonite, *How a Teenager's Code Spawned a \$432,500 Piece of Art*, pubblicato su *Wired*, il 20 novembre 2018, <https://www.wired.com/story/teenagers-code-spawned-dollar-432500-piece-of-art/> URL consultato in data 14 settembre 2021.

e non ricevette alcun contributo dalla vendita dell'opera, ma il suo nome sarebbe comunque diventato conosciuto in tutto il mondo di lì a poco¹⁶⁵.

3.2.5. Mauro Martino



Fig. 29: Alcuni esempi di immagini generate da *AI Portraits Ars*, Mauro Martino, 2019.

Abbiamo già messo in luce in precedenza come l'uso delle GAN per quanto riguarda la creazione di opere d'arte possa assumere una dimensione interattiva: abbiamo infatti evidenziato come spesso alcuni artisti ricerchino nelle loro installazioni una sorta di interazione e di contatto con lo spettatore.

È questo il caso di *AI Portraits Ars*, creata dal designer italiano Mauro Martino (direttore presso il *Visual Artificial Intelligence Lab di Ibm research* e docente presso la *Northeastern University of Boston*).

¹⁶⁵ Per altri lavori di Robbie Barrat: <https://aiartists.org/robbie-barrat>, <https://robbiebarrat.github.io/>.

Il funzionamento è innovativo e unisce le più recenti tecnologie digitali all'arte, attraverso l'uso delle GAN. Si tratta in questo caso di un'applicazione: l'utente ha in primo luogo la possibilità di scattarsi un *selfie*, in seguito l'immagine della persona in questione viene rielaborata attraverso l'applicazione e viene trasformata in un'opera: viene cioè modificata basandosi su dipinti già esistenti di alcuni tra i più grandi maestri dell'arte (passando attraverso stili ed epoche differenti, come ad esempio da Van Gogh, a Rembrandt a Warhol).

Anche in questo caso il sistema diventa quasi un pittore umano in carne ed ossa: vi è sempre un'originalità, i tratti e le linee sono verosimili ad opere dipinte e non elaborate artificialmente, i colori sono anch'essi particolari ed originali.

L'applicazione, che crea queste immagini, è stata costruita attraverso le GAN: la doppia rete neurale agisce sulle foto, in primo luogo scomponendole in diverse componenti¹⁶⁶ (occhi, labbra, capigliatura, espressione e vari connotati che caratterizzano il soggetto). In secondo luogo, dopo la scomposizione, avviene la ricomposizione: a partire dal *training set* che fornisce i dati su cui basarsi (in questo caso, per l'appunto, i celebri dipinti di artisti di differenti stili ed epoche), gli elementi scomposti vengono ricomposti sulla base dei dipinti appartenenti al *training set*. Il risultato finale avrà quindi in sé qualcosa dell'originale, manterrà i tratti principali riconducendo inconfutabilmente alla persona che ha scattato la foto in principio, ma attraverso il lavoro delle reti neurali il "ritratto" originale verrà trasformato in qualcosa di completamente nuovo, basandosi su stili esistenti e appartenenti a differenti periodi storici. Innovazione e storia si fondono così inevitabilmente tra loro.

Possiamo identificare due "fasi di sviluppo" di questa applicazione: durante la prima (periodo 2018)¹⁶⁷ le immagini utilizzate per allenare la rete neurale erano foto di celebrità e personaggi noti, come ad esempio: Elon Musk, Marie Curie, Mark Zuckerberg, Donald Trump, Hillary Clinton e Nicholas Cage. I risultati in alcuni casi erano incredibilmente

¹⁶⁶ Cfr. Maria Francesca Fortunato, *Fatevi fare un ritratto (vero) dall'AI. Si ispira ai grandi maestri, ma inventa come un pittore*, pubblicato su *Italian.Tech* il 16 luglio 2019. https://www.italian.tech/2019/07/16/news/fatevi_fare_un_ritratto_vero_dall_intelligenza_artificiale-302647695/ URL visitato in data 16 settembre 2021.

¹⁶⁷ Cfr. Dan Robitzski, *Here's What AI Thinks A Bunch of Famous People Look Like*, pubblicato su *Futurism* il 18 agosto 2018. <https://www.fastcompany.com/90376689/what-you-look-like-as-an-renaissance-painting-according-to-ai> URL visitato in data 20 settembre 2021.

accurati, in quanto l’I.A. riusciva a riconoscere il volto e a riprodurlo in maniera alquanto fedele, in altri casi invece le immagini generate artificialmente assumevano un aspetto confuso, a tratti quasi inquietante, generando volti che sembravano più una sorta di “incubo” che una riproduzione fedele o una rielaborazione creativa della foto. Ad esempio, nella Fig. 29 possiamo vedere una foto originale a sinistra di Mark Zuckerberg e una riprodotta artificialmente a partire dalla prima. Come vediamo compare senza motivo al posto del sorriso iniziale una sorta di “smorfia” con la lingua. Si può trattare di una rielaborazione creativa dell’immagine di partenza? Forse, ma il risultato non è esattamente piacevole al primo sguardo. In questa fase, però, l’intelligenza artificiale è stata allenata su volti già noti e, come sappiamo, le GAN migliorano e diventano sempre più sofisticate attraverso l’allenamento, l’addestramento iniziale e il continuo lavoro. A questa “prima fase” ne segue infatti una seconda.



Fig. 30: *AI Portraits*, Mauro Martino, 2018.

Nel 2019 nasce un nuovo progetto del designer italiano, simile al primo, ma più elaborato. Gli *AI Portraits* vengono stavolta fatti lavorare su volti di persone comuni, che possono caricare una propria foto sul sito dell’applicazione e farsi “dipingere” dalla rete neurale, che segue un vero e proprio stile artistico. È in questa fase che l’algoritmo diventa creativo, diventa quasi un artista in grado di disegnare e dipingere i volti che vede.

L'algoritmo ideato da M. Martino in questo periodo ha la capacità di creare a partire da un rumore (from scratch)¹⁶⁸. Nascono gli Ai Portraits Ars: l'utente che carica la propria foto può vederla riprodotta nello stile dei grandi pittori rinascimentali. L'immagine, in seguito, viene direttamente eliminata dal server per poter tutelare la privacy degli utenti. Anche in questo caso possono esserci degli errori, ad esempio Martino afferma che in uno dei primi esperimenti la macchina aveva confuso in proprio volto durante dell'elaborazione, dando come risultato l'immagine di una donna che poteva sembrare "uscita da una vecchia edizione di un romanzo di Jane Austen"¹⁶⁹. Ci si chiede, dunque, come mai la GAN decida di disegnare un volto in base ad un determinato stile e seguendo una certa linea pittorica, mentre in un altro caso faccia scelte differenti. L'ipotesi di Martino è che in alcuni casi, ad esempio, se lo sfondo di riferimento è chiaro la rete neurale può scegliere di disegnare con matita e inchiostro (prendendo ispirazione da altri ritratti fatti in questo modo). Queste sono alcune ipotesi, ma a volte i risultati della GAN sono sorprendenti non solo per il pubblico, ma addirittura per colui che l'ha creata. Questo ci fa capire quanto questo strumento dia spazio a risultati sorprendenti e quanto ancora ci sia da indagare a riguardo. Ad ogni passo può esserci una scoperta nuova, con ogni "addestramento" si raggiunge un livello di sofisticatezza maggiore.

In un'intervista¹⁷⁰ riguardante il tema dell'Intelligenza Artificiale e nello specifico la possibilità di applicarlo al mondo dell'arte e del design, aumentando la produttività e la creatività, Mauro Martino afferma: "I migliori risultati sono probabilmente nel campo visivo, per questo i designers dovrebbero essere i primi a utilizzare questa tecnologia. Grazie ai modelli GAN (Generative Adversarial Network) possiamo giocare con le forme e i colori fondendoli e sfumando i confini tra diverse classi di oggetti. L'AI è la nuova musa del designer"¹⁷¹.

¹⁶⁸ Cfr. Katharine Schwab, *What you look like as a Renaissance painting, according to AI*, pubblicato su *FastCompany* il 16 luglio 2019. <https://www.fastcompany.com/90376689/what-you-look-like-as-an-renaissance-painting-according-to-ai> URL visitato in data 20 settembre 2021.

¹⁶⁹ *Ibidem*.

¹⁷⁰ Cfr. Intervista a Mauro Martino, pubblicata su *Brera Design District*, 2021. <https://www.breradesigndistrict.it/people-and-stories/mauro-martino/> URL visitato in data 16 settembre 2021.

¹⁷¹ Per altri lavori di Mauro Martino: <https://www.mamartino.com/projects.html>.

3.2.6. Chris Peters



Fig. 31: Chris Peters, *AI Paintings*, 2019.

In maniera simile a quella vista precedentemente per quanto riguarda l’artista Robbie Barrat, un altro artista nel marzo del 2019 riesce a creare, utilizzando immagini di dipinti già esistenti di paesaggi (appartenenti al diciannovesimo secolo), delle immagini ritraenti paesaggi del tutto nuovi. Si tratta di Chris Peters, il quale durante un’esibizione privata del suddetto anno, espone alcuni dipinti creati attraverso l’uso delle GAN¹⁷².

Nell’immagine (Fig. 30) appaiono due opere: quella a destra è un dipinto ad olio creato manualmente dall’artista, mentre quella a sinistra è stata generata, traendo ispirazione dal dipinto ad olio, dall’intelligenza artificiale¹⁷³, da una GAN elaborata anch’essa da Peters.

L’artista, infatti, definisce la GAN una sorta di “musa dell’intelligenza artificiale”: si tratta di *Tensordream*, con cui egli, possiamo dire, riesce a tutti gli effetti a collaborare. È a tutti gli effetti un’interazione tra il lavoro manuale e quello artificiale, un “lavorare insieme”

¹⁷² Cfr. Srishti Deoras, *How GAN Was The True Artist In 2019*, pubblicato su *Analytics India Magazine*, il 10 dicembre 2019, <https://analyticsindiamag.com/how-gan-was-the-true-artist-in-2019/> URL visitato in data 18 settembre 2021.

¹⁷³ Cfr. *When the artist isn't there*, pubblicato su *Higher Nationals*, il 23 luglio 2019, <https://www.highernationals.com/blog/when-artist-isnt-there> URL visitato in data 18 settembre 2021.

creando dei dipinti ad olio, che vengono poi rielaborati e riprodotti artificialmente dalla macchina: nuovamente, si presenta così la possibilità della collaborazione tra uomo e macchina.

Tensordream ha origine da una rete neurale, che tuttavia viene modificata dall'artista per assimilare la complessità dei paesaggi. Per diversi giorni la rete studia la composizione e i colori di migliaia di paesaggi, per arrivare infine a comprenderne la forma e a proporre poi a sua volta delle composizioni e a produrre da sé dei “dipinti”.

La maggior parte dei dipinti forniti nel *training set* viene selezionata da Peters stesso, che sceglie quelli che più rappresentano il tonalismo americano. Nonostante le immagini che vengono in prima battuta create dalla macchina abbiano un aspetto strano, non del tutto verosimile a paesaggi realistici, bisogna mettere in luce come la macchina inizi il processo “da zero”, con la sola capacità di autoapprendere.

A partire dai dipinti generati dall'intelligenza artificiale, inizialmente confusi e rappresentanti forme quasi “aliene” il compito finale di Peters era modificare le immagini e renderle più verosimili, una sorta di “concretizzazione”, dunque, dell'idea della macchina. Seppur non partendo da un compito facile, l'artista riuscì infine a capire la logica che la maggior parte delle volte la macchina adottava, capì che per l'appunto quel risultato era dovuto alla mancanza di conoscenza da parte della macchina del mondo reale.

Creazione dell'opera da parte della macchina, concretizzazione dell'artista: una commistione di punti di vista per un lavoro in continua simbiosi, che stimola e arricchisce le potenzialità umane. È così che l'artista concepisce questo tipo di lavoro: ecco il motivo per cui *Tensodream* viene definita proprio una sorta di sua “musa dell'intelligenza artificiale”¹⁷⁴.

“I learned in art school it can take hours and hours of careful observation before your mind quiets down to the point you can really see and understand something,” [...] “I wanted to get inside the AI's head, to achieve some understanding of what it was trying to do. I was able to, but only after days and days of looking at them while painting them”. [...] “If I just printed out the image, I would not understand 1/100th of what is there compared to standing for hours and hours and days and days painting”¹⁷⁵.

¹⁷⁴ Cfr. <https://tensordream.ai>, URL visitato in data 18 settembre 2021.

¹⁷⁵ Cfr. J. Vincent, *A never-ending stream of AI art goes up for auction*, pubblicato su *The Verge*, il 5 marzo 2019, <https://www.theverge.com/2019/3/5/18251267/ai-art-gans-mario-klingemann-auction-sothebys-technology>. URL visitato in data 18 settembre 2021. Tr.: “Ho imparato nella scuola d'arte che ci possono volere ore e ore di attenta osservazione prima che la tua mente si calmi fino al punto in cui puoi davvero vedere e capire qualcosa” [...] “Volevo entrare nella testa dell'IA, per raggiungere una certa comprensione di ciò che

3.2.7. Tamiko Thiel



Fig. 32: Tamiko Thiel, *Lend Me Your Face, Go Fake Yourself!*, Metronom, 2021.

Abbiamo esposto, in precedenza¹⁷⁶, il funzionamento dei deepfake e la loro relazione con le *fake news*: l’aspetto, dunque, più “ingannevole” e possiamo dire il “risvolto negativo” che questa tecnologia può avere in relazione all’uomo. Ma se il deepfake in relazione alle *fake news* può avere un impatto negativo, esiste la possibilità di rivalutare questo mezzo alla luce di un uso più “positivo”? In relazione all’arte e rivalutando questo fenomeno in maniera creativa, possono i deepfake interagire con lo spettatore senza metterlo però “in pericolo”, senza dunque che egli venga volutamente ingannato?

stava cercando di fare. Ero in grado di, ma solo dopo giorni e giorni di osservazione mentre li dipingevo”. [...] “Se avessi solo stampato l’immagine, non avrei capito 1/100 di ciò che era presente, rispetto allo stare in piedi per ore e ore e giorni e giorni a dipingere”.

¹⁷⁶ Cfr. Capitolo 2, *Deepfake: risvolti negativi nell’uso dell’intelligenza artificiale*.

È quello che ci propone l'artista Tamiko Thiel, con la sua installazione *Lend me your face*: si tratta della stessa tecnologia di intelligenza artificiale dei deepfake usata in questo caso per animare singole foto di volti con espressioni facciali che vengono prese da video pubblici di personaggi noti e celebrità (tra i vari esempi: Greta Thunberg – che vediamo in Fig. 31 – Queen Elizabeth II, il Primo Ministro inglese e altri personaggi di spicco in campo politico).

L'installazione, che in questo caso non ha lo scopo di ingannare, bensì di trasmettere un messaggio allo spettatore (che anche in questo caso, come già visto con altri artisti, partecipa attivamente all'opera) ci fa riflettere su quanto la tecnologia dei deepfake possa modificare quella parte che è per l'essere umano al tempo stesso la più intima e la più pubblica di sé: il volto. In questo caso le emozioni espresse possono essere manipolate e portate al di fuori del controllo sia di chi partecipa e interagisce con l'opera (lo spettatore), sia di chi "presta" il suo volto (le celebrità).

L'uso della tecnologia è simile a qualcosa che abbiamo già mostrato precedentemente (si veda il paragrafo sui deepfake o, per quanto riguarda le opere d'arte e la rielaborazione delle foto dello spettatore da parte della macchina, gli AI Portraits Ars di Mauro Martino), ma in questo caso ciò che cattura l'attenzione è il messaggio che si vuole trasmettere e la riflessione che è auspicabile incentivare in chi fruisce di quest'opera: i video risultanti sottolineano il crescente sviluppo di queste tecnologie ed evidenziano quello che è da un lato il desiderio delle persone di esplorarle e di indagarne l'uso, impegnandosi e interagendovi, dall'altro le problematiche e le paure che si celano sotto a questo desiderio¹⁷⁷.

In occasione dell'esposizione di questa installazione nella galleria *The Photographers Gallery*, l'artista in un'intervista¹⁷⁸ con la curatrice Sarah Cook, spiega il significato della sua scelta e l'impatto che possono avere i deepfake e l'uso di queste tecnologie:

An emotion I want to play with is the exhilaration of giving up your individuality to become part of a larger, more powerful "body". Of course, this is not just a Big Brother/Fascist attraction, or the euphoria of being with the winning team, etc. This is why I included Greta

¹⁷⁷ Cfr. *Lend me your Face! - Tamiko Thiel*, on *The Photographers Gallery*, 2021. <https://thephotographersgallery.org.uk/whats-on/lend-me-your-face-tamiko-thiel> URL visitato in data 17 settembre 2021.

¹⁷⁸ Cfr. S. Cook, *Maintaining Composure: An Interview with Tamiko Thiel*, pubblicato su *Unthinking Photography*, gennaio 2021. <https://unthinking.photography/articles/maintaining-composure-an-interview-with-tamiko-thiel> URL visitato in data 17 settembre 2021.

Thunberg: she's become a figurehead for a movement that I myself very strongly believe in, value and participate in. I believe in and repeat her words, and feel she speaks for us, fellow believers. So in *Lend Me Your Face!* (the title comes from Shakespeare's play **Julius Caesar**, where Mark Antony says, "Friends, Romans, countrymen, lend me your ears!") you're becoming part of someone else's chorus, for good or for evil – and each person might have a different idea of who is good, and who evil!¹⁷⁹

L'artista, inoltre, spiega come spesso le venga chiesto dagli spettatori dell'opera se possono scegliere chi essere, mentre la sua risposta è no, proprio in quanto:

That's the whole point: we're putting our faces out there on social media, they're being harvested, we're not being asked for permission at all, we're not even being told that it's happening. In *Lend Me Your Face!*, within the constraints of German data privacy laws, I'm trying to give you as much of that experience as possible¹⁸⁰.

È inoltre fondamentale che lo spettatore provi, davanti alla visione del proprio volto applicato ad un video con volti pubblici di altri personaggi di spicco, una sensazione di iniziale spaesamento, di familiarità e al tempo stesso di estraneità davanti a ciò che vede. Un volto familiare e allo stesso tempo “perturbante”, un volto che riconosce ma con una voce di qualcun altro, qualcuno che non viene scelto personalmente. Nella stessa intervista l'artista spiega che questo *uncanny moment* è specificatamente voluto e ricercato, si cerca di suscitare quella sensazione di perturbante in chi fruisce dell'opera, proprio per portarlo a riflettere su ciò che sta vedendo. E sarebbe impossibile non farlo, specialmente quando si tratta del nostro volto.

Abbiamo così un ulteriore esempio di come l'arte, in maniera immediata e spesso sorprendente, possa colpire lo spettatore veicolando un messaggio e portandolo a riflettere, proprio coinvolgendolo in prima persona, su temi importanti riguardo l'attualità e lo sviluppo di queste tecnologie. Di nuovo, arte come “via positiva” per l'interazione con lo spettatore,

¹⁷⁹ *Ibidem*, Tr: “Un'emozione con cui voglio giocare è l'euforia di rinunciare alla propria individualità per diventare parte di un “corpo” più grande e potente. Certo, non si tratta di un Grande Fratello o di un'attrazione fascista, o dell'euforia di essere parte di una squadra vincente. Ecco perchè ho incluso Greta Thunberg: è diventata una figura di spicco per un movimento in cui io stessa credo fortemente, apprezzo e partecipo. Credo e ripeto le sue parole, e sento che parla per noi, compagni credenti. Così in *Lend Me Your Face!* (il titolo deriva dalla commedia di Shakespeare Giulio Cesare, dove Marco Antonio dice: "Amici, Romani, compatrioti, prestami le tue orecchie!") stai diventando parte del coro di qualcun altro, per il bene o per il male - e ogni persona potrebbe avere un'idea diversa di chi è il bene e chi il male!”

¹⁸⁰ *Ibidem*, Tr: Questo è il punto: stiamo mettendo le nostre facce là fuori sui social media, vengono raccolte, non ci viene chiesto il permesso a tutti, non ci viene nemmeno detto che sta accadendo. In *Lend Me Your Face!*, entro i limiti delle leggi tedesche sulla privacy dei dati, sto cercando di darvi quanta più esperienza possibile.

arti visuali come metodo immediato di colpire chi assiste (e, in questo caso, partecipa addirittura) all'opera¹⁸¹.

3.2.8. Memo Akten

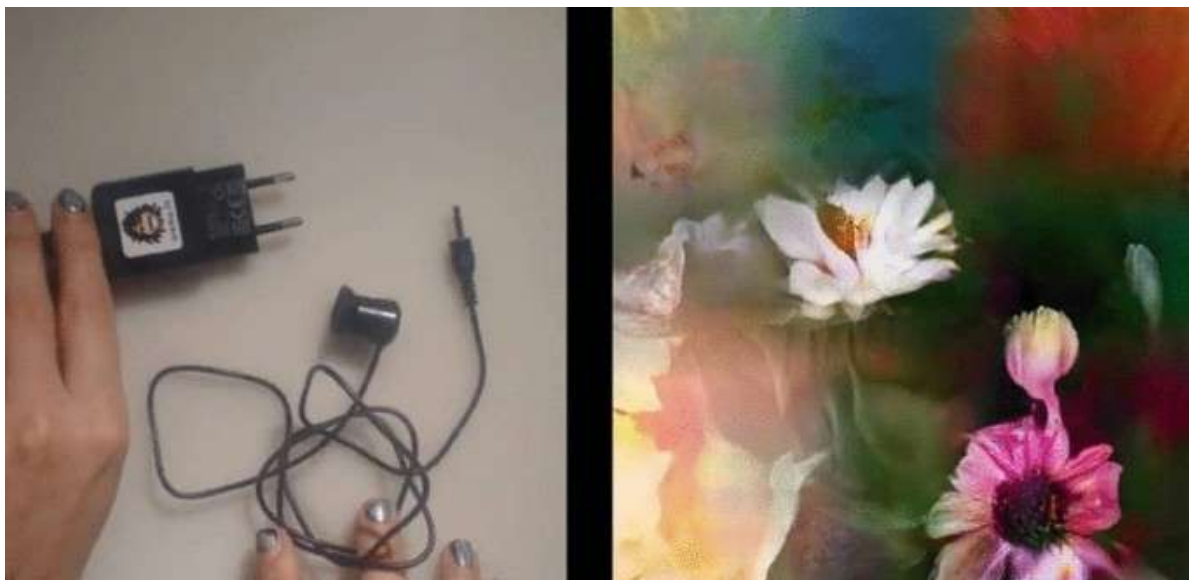


Fig. 33: Memo Akten, *Learning to see*, 2017.

É possibile, per una macchina, “imparare a vedere”? Come fanno le macchine a riprodurre un oggetto e, soprattutto, possono darne un'interpretazione? Questo è ciò che si chiede Memo Akten in *Learning to see*.

Siamo nel maggio del 2019, a Londra, presso il Barbican Centre. Durante una mostra l'artista e ricercatore nel campo dell'intelligenza artificiale si chiede proprio questo e cerca di dare una risposta, esponendo delle opere che ci aprono nuovi orizzonti sul modo che hanno le macchine di vedere gli oggetti. Anche in questo caso, gli spettatori sono portati ad interagire con l'opera, ma stavolta sottoponendo non la propria immagine e il proprio volto

¹⁸¹ Per altri lavori di Tamiko Thiel: <https://tamikothiel.com>.

alla macchina, bensì diversi oggetti di uso comune appositamente scelti per essere sottoposti al lavoro di una GAN: pezzi di stoffa, chiavi, telecomandi, caricabatterie, occhiali...

Durante il processo di rielaborazione succede qualcosa di inaspettato: la GAN presenta un modo completamente nuovo e diverso rispetto all'osservatore umano di vedere le cose: è così che un comune caricabatterie sotto lo sguardo della macchina si trasforma nella rappresentazione di fiori e petali colorati, quasi fosse un dipinto ad olio (Fig. 32). Un pezzo di stoffa qualunque, invece, può diventare nella prospettiva della GAN un panorama con mare e orizzonte, cielo e mare in tempesta. (Fig. 33), Muovendo il panno nella realtà vediamo il mare nella prospettiva della GAN muoversi, le onde infrangersi come fosse in tempesta (sono stati fatti in questo caso alcuni video dall'artista, oltre alle riproduzioni fotografiche¹⁸²).

Dunque: la GAN è in grado di interpretare? Quelli che sono oggetti di uso comune possono essere interpretati dalla macchina come paesaggi o oggetti naturali? Osservando *Learning to see* di Memo Akten pare di sì. Si tratta di un "mancato riconoscimento"¹⁸³ oppure di una trasformazione e reinterpretazione in maniera creativa dell'oggetto? Secondo l'autore si può rispondere così: "Una rete neurale artificiale osserva il mondo esterno, tentando di dare un senso a ciò che vede, in funzione di ciò che ha visto precedentemente. Proprio come noi, può vedere soltanto ciò che conosce già"¹⁸⁴.

Da queste parole si può capire il perché la macchina veda, ad esempio, una nuvola o un fiore al posto di un caricatore: questo è dovuto principalmente al *training set* e al modo in cui essa è stata allenata. L'intelligenza artificiale riconosce i dati che ha a disposizione, il set tramite cui viene allenata rappresenta una sorta di "memoria": non può "immaginare" indiscriminatamente, bensì può immaginare e creare sulla base dei dati che sono stati forniti da chi la guida e la istruisce. Dunque, se la GAN è stata allenata con immagini (solitamente prese da Google Art, ImageNet e altri siti *opensource*) di elementi naturali, di conseguenza riconoscerà questi dati, avrà a disposizione questo *background* di riferimento e genererà immagini di questo tipo. In questo possiamo vedere una sorta di similitudine, forse, con

¹⁸² È possibile vedere alcuni video al sito <http://www.memo.tv/works/learning-to-see/>.

¹⁸³ Cfr. A. Barale, *A hideaway in the wastelands: nuove sfide filosofiche dell'AI art*, in A. Barale, *Be my Gan*, Jacobo, Milano, 2020, p. 226.

¹⁸⁴ Ivi, cit. p. 226.

l'uomo, che immagina e crea in base anche alla propria esperienza, a ciò che ha vissuto e "immagazzinato" durante le proprie esperienze, al proprio bagaglio culturale e di vita. La sua percezione e riproduzione della realtà cambierà anche in base a questo, così come nel caso della macchina cambierà il modo di riconoscere i dati e di generare in base al training set attraverso cui è stata allenata.

È interessante, dunque, la spiegazione dell'artista e il parallelismo con l'uomo: "Proprio come noi, può vedere soltanto ciò che conosce già". Il mondo esterno è per noi, infatti, qualcosa che varia in base alle percezioni, non è possibile darne una riproduzione identica e ugualmente valida per ogni persona. Ognuno ha un modo di "processare" e rielaborare le informazioni che arrivano dagli stimoli esterni diverso, che varia in base alla sensibilità, alle esperienze pregresse, alla capacità di comprendere, all'istruzione ecc. Ci sono tanti dettagli che possono fare la differenza nel caso dell'uomo. Certamente della macchina non possiamo dire lo stesso, non vi è la stessa possibilità di interpretare e dare una comprensione del mondo esterno (per farlo dovrebbe avere capacità cognitive, di linguaggio, sensibilità e molto altro), ma sicuramente nel caso del background esperienziale dovuto in questo caso all'addestramento e al training set di partenza, può esserci una similitudine: "L'immagine che vediamo nella nostra mente cosciente non è un'immagine speculare del mondo esterno, bensì una ricostruzione basata sulle nostre aspettative e sulle nostre convinzioni preesistenti"¹⁸⁵. Dunque, possiamo dire che questa ricostruzione del mondo esterno viene fatta dalla GAN sulla base dell'archivio di memoria precostituito, che viene scelto e guidato da chi la progetta e la allena.

¹⁸⁵ *Ibidem*.



Fig. 34: Memo Akten, *Gloomy Sunday – Learning to see*, 2017. Installazione interattiva al Barbican Centre, 2019.

Il progetto *Learning to see* è frutto di un percorso portato avanti dall'artista dal 2017, risultato di un continuo lavoro di ricerca e sperimentazione che è ancora in corso. L'installazione al Barbican Centre di Londra, invece, risale come abbiamo detto al 2019 e permette al pubblico un'interazione viva e presente con la macchina. Si tratta in questo caso di una telecamera puntata su un tavolo in cui le reti neurali analizzano e interpretano in tempo reale gli oggetti presenti, che possono essere toccati, spostati e manipolati dagli spettatori, vedendo sullo schermo il modificarsi in tempo reale delle immagini e delle interpretazioni dell'intelligenza artificiale. L'intento dell'artista è proprio indurre lo spettatore in maniera diretta, immediata, coinvolgendolo in prima persona, a ragionare su quella che è la visione del mondo di ognuno, su quanto l'esperienza della realtà sia frutto di esperienze e interpretazioni personali. Anche in questo caso, nella visione dell'artista, la rete neurale è liberamente ispirata alla corteccia umana e, attraverso conoscenze preesistenti, cerca di dare un senso alla realtà che osserva¹⁸⁶.

Un altro progetto interessante proposto dall'artista nello stesso anno è *Learning to Dream: Supergan*. In questo caso il progetto è incentrato interamente sull'arte: la macchina

¹⁸⁶ Cfr. IArte, *Memo Akten: vedere e sognare con l'Intelligenza Artificiale*, pubblicato su *IArte: Nuove frontiere tra arte e Intelligenza Artificiale*, il 19 giugno 2019, http://iarte.altervista.org/memo-akten-intelligenza-artificiale/?doing_wp_cron=1631961150.9074130058288574218750 URL visitato in data 18 settembre 2021.

è stata addestrata attraverso immagini di storia dell'arte, migliaia di opere di artisti scaricate dal sito *Google Art Project* di differenti generi e con diversi soggetti, che spaziano dal cubismo, al realismo all'astrattismo, da immagini con soggetti religiosi ad immagini di paesaggi naturalistici. Allenando in questo modo la macchina, in maniera differente dal caso precedente, è come se metaforicamente parlando le facessimo “studiare la storia dell'arte” per poi, infine, sulla base delle conoscenze ricevute dal training set di allenamento, far riprodurre alla rete neurale in maniera creativa delle opere d'arte.

Da questo punto di vista il progetto può essere simili ad altri visti nell'esposizione di alcuni *gan artists*, ma la peculiarità di Memo Akten¹⁸⁷ è proprio questo parallelismo tra la GAN e la mente umana, la capacità della GAN di “vedere” qualcosa, di dare un senso a ciò che le si sottopone sulla base del modo in cui viene addestrata.

3.3. Creatività e *machine learning*: un dibattito ancora aperto

Facendoci guidare dalla ricerca e dallo studio di alcuni artisti nel campo della *Gan art* e attraversando brevemente il loro percorso artistico a contatto con queste nuove tecnologie, siamo giunti a nuove consapevolezze. Grazie al loro sguardo creativo abbiamo visto, infatti, come sia possibile, in maniera tangibile, una collaborazione tra macchina e uomo: a partire dal discorso teorico sulle *Creative Adversarial Networks* e dalla riflessione sull'ipotetica creatività delle macchine, abbiamo visto sul piano concreto come queste sperimentazioni abbiano dato luogo a nuove scoperte, alla generazione di nuove ed originali opere. Dunque, ci troviamo davanti ad una risposta concreta alle domande poste inizialmente.

È necessario, tuttavia, ricordare sempre che ci stiamo addentrando in un campo che è ancora in via di sviluppo e sottoposto ad una costante ricerca. Di conseguenza, nonostante le risposte di questi innovativi artisti, dobbiamo tenere a mente l'impianto teorico di partenza, la risposta sul piano pratico e portare a posteriori una nuova riflessione.

¹⁸⁷ Per altri lavori di Memo Akten: <http://www.memo.tv/works/#selected-works>.

Proprio in quanto innovativo, questo campo è sottoposto ancora ad un forte dibattito e ad opinioni scettiche riguardo la creatività delle macchine e del *machine learning*: ci troviamo davvero davanti ad un nuovo sviluppo della creatività e a nuove possibilità per l'arte o si tratta di una svalutazione del lato artistico umano, come affermano i più scettici? Dobbiamo farci guidare dall'opinione di alcuni esperti.

“Le opere che implicano un qualsivoglia impiego dell'intelligenza artificiale suscitano dibattiti sulla validità della creatività delle macchine, sull'identità del vero artista e sulla qualità estetica dei risultati. In questo momento filosofi, esperti di scienze del computer ed esponenti del mondo dell'arte devono affrontare interrogativi di grande complessità”¹⁸⁸.

Come mette in luce giustamente Mirian Mazzone, le domande sul *machine learning* che ci stiamo ponendo e che caratterizzano questo dibattito, toccano più implicitamente un'altra domanda a cui bisogna rispondere (attraversando quindi diversi campi disciplinari): cos'è l'arte? Cosa può essere definito effettivamente creativo? Cosa intendiamo con questi termini e a cosa ci riferiamo, dunque, noi esseri umani che creiamo opere artistiche e fruiamo di esse? Se non riusciamo a dare una risposta innanzitutto a queste domande e a valutare il valore artistico ed estetico delle opere, sarà difficile dare una risposta relativa al *machine learning* e ad affermare, dunque, se l'essere creativo possa appartenere al solo essere umano o anche alle macchine.

Riprendendo la definizione di creatività di Margaret Boden¹⁸⁹ (sia negli esseri umani che nelle macchine), l'autrice porta la riflessione su una distinzione: partendo dal principio secondo cui la creatività è “la capacità di creare idee nuove e dotate di valore”¹⁹⁰, vengono distinti due diversi livelli di creatività: esistono infatti “idee psicologiche” considerate innovative, ma solo per il singolo che le genera e non riconosciute universalmente anche dagli altri individui, e “idee storiche”, che hanno la caratteristica di essere nuove per il singolo individuo che le genera, ma anche collettivamente per gli altri individui. L'autrice sottolinea

¹⁸⁸ Cfr. M. Mazzone, *Le GAN e la questione della creatività nell'arte e nell'intelligenza artificiale*, in A. Barale *Be my Gan*, Jacabook, Milano, 2020, cit. p.51.

¹⁸⁹ Cfr. M. Boden, *The Creative Mind: Myths and Mechanisms (second edition)*, Routledge, 2003.

¹⁹⁰ Cfr. M. Mazzone, *Le GAN e la questione della creatività nell'arte e nell'intelligenza artificiale*, in A. Barale *Be my Gan*, Jacabook, Milano, 2020, cit. p.53.

come la maggior parte delle idee appartengano alla prima categoria, mentre è molto più raro e difficile farle rientrare nella seconda categoria.

“Oggi siamo riusciti ad addestrare macchine e a elaborare processi generativi che, per usare la terminologia della Boden, possiedono “la capacità di generare idee nuove e dotate di valore”¹⁹¹. Seppure quindi vi siano diversi dubbi a riguardo, M. Mazzoni afferma che attraverso l’intelligenza artificiale vi sia la possibilità di avere un soggetto creativo in grado di creare opere d’arte: nonostante il livello di creatività sia paragonabile ancora ad un piano di idee “psicologiche”, vi è la possibilità di passare al piano di “idee storiche”. Si tratta ancora di un’aspirazione per chi lavora in questo campo, ma la possibilità concreta esiste. Bisogna precisare in ogni caso che secondo l’autrice i parametri di valutazione vanno sempre tenuti a mente in questo dibattito.

Bisogna tenere, inoltre, presente come all’interno dell’elaborazione dei dati da parte dell’intelligenza artificiale vi sia sempre uno spazio per delle “anomalie”, uno spazio che può consentire il conseguimento di innovazioni e dunque il raggiungimento della creatività. È proprio quello spazio per gli errori ciò che permette l’esplorazione di quella zona adibita al raggiungimento della creatività. Soprattutto nel caso delle GAN (che rappresentano ad oggi la tipologia di algoritmi ad oggi in grado di produrre risultati più sorprendenti) le immagini generate risultano molto simili a quelle del *training set*, ma d’altra parte non sono completamente identiche. È proprio grazie all’attività antagonista delle due reti neurali che si crea quello “spazio” per gli errori e per la creatività. La costante competizione tra le due reti neurali in cui una tenta di “ingannare” l’altra, migliorandosi di volta in volta e incrementando i risultati prodotti, dà la possibilità di esplorare zone che i normali algoritmi non riuscirebbero a raggiungere: si tratta di un meccanismo molto più sofisticato rispetto a quelli comuni.

Possiamo dire che nel caso delle GAN i risultati si riferiscono indubbiamente agli input iniziali, che, come abbiamo visto, fanno riferimento al database di allenamento, ma d’altra parte vengono manipolati e rielaborati, in modo tale che i risultati del processo di

¹⁹¹ Ivi, cit. p. 53.

elaborazione possano discostarsi anche notevolmente da questi input iniziali: gli esiti possono essere così imprevedibilmente creativi ed inaspettati.

Sebbene l'uso delle GAN possa essere ampliato a differenti ambiti, bisogna precisare che la funzione imitativa, volta alla creazione di opere sulla base di uno stile artistico preesistente o all'animazione di fotografie non viene considerata propriamente appartenente all'ambito artistico in senso stretto (quello più specificatamente creativo). Il livello di creatività dipende dalla scelta del database e da quanto una GAN sia sofisticata: queste azioni implicano senza dubbio una grande partecipazione dell'artista e una stretta "collaborazione" con la macchina. Di conseguenza, gli artisti che utilizzano questo tipo di tecnologie si possono considerare artisti a tutti gli effetti.

Durante l'esposizione del funzionamento delle *Creative Adversarial Networks*, è stato posto in evidenza il processo creativo alla base della creazione delle opere da parte della macchina: nel caso della GAN vi è una tensione tra il tentativo di seguire gli esempi forniti nel database di partenza e la penalizzazione quando il risultato si avvicina eccessivamente all'originale: "L'obiettivo è creare immagini simili – ma non esattamente identiche – a quelle usate per l'addestramento, evitando i risultati eccessivamente nuovi e quindi sgradevoli o irriconoscibili"¹⁹². Possiamo dire quindi che, in maniera comunque controllata e entro spazi limitati, vi sia margine per l'innovazione e la sperimentazione: si tratta di capire se queste possibilità siano paragonabili a quelle della mente umana (la difficoltà rimane tuttavia anche nel capire il funzionamento di quest'ultima).

Riguardo quest'ultimo punto nello specifico, ovvero il parallelismo tra mente umana e "mente" computazionale nella generazione di immagini, M. Mazzone fa notare una differenza fondamentale¹⁹³: lo stimolo di partenza. È evidente che per quanto una macchina possa ricevere stimoli simili e produrre anche opere simili a quelle prodotte dall'uomo (ad esempio, una macchina potrebbe certamente generare il dipinto di un paesaggio o il ritratto di un volto), questa non potrà mai ricevere quello stimolo con l'esperienza diretta, calandosi nella realtà e vivendo, per poi rielaborarla, una determinata situazione. La macchina non può "essere fisicamente" in una situazione, non avendo un corpo. L'uomo può vivere le

¹⁹² Ivi, cit. p. 57.

¹⁹³ Ivi, cit. p. 65.

esperienze nella realtà, trarre stimoli dal mondo esterno, calarsi nella natura e riprodurre in maniera più o meno verosimile la propria esperienza, la macchina non può farlo. Il suo unico punto di riferimento è la mente dell'uomo: egli solo può insegnare e trasmettere il proprio modo di elaborare la realtà esterna.

Abbiamo già messo in luce più volte nel corso dell'esposizione l'importanza della collaborazione tra uomo e macchina: è proprio la diversità dell'approccio e il livello di relazione a dare diverse sfumature al processo e al risultato delle opere. Nell'esposizione degli artisti, infatti, abbiamo visto quanto possa cambiare l'immagine generata in base ad alcuni elementi, tra cui:

- Grado di indipendenza e autonomia della macchina nel processo di generazione delle immagini;
- Partecipazione più o meno attiva dell'artista che la guida, volontà di controllo sulle azioni della macchina (l'artista può infatti dipingere in prima persona i dipinti da inserire nel database di partenza per avere una partecipazione maggiore nel processo creativo, come nel caso degli artisti Anna Ridler e Chris Peters);
- Interazione partecipazione dello spettatore con l'opera generata o nel processo attivo di costruzione dell'opera (seguendo gli esempi di Anna Ridler, Mauro Martino, Tamiko Thiel, Memo Akten).

In base alla presenza di questi fattori e all'intersecarsi di essi cambierà notevolmente lo stile di un artista rispetto ad un altro, pur usando il medesimo strumento. Ed è proprio ciò che va sottolineato: il fatto che le GAN siano anche uno strumento, non sistemi autonomamente creativi e generativi indipendentemente dalla mano e dalla mente dell'uomo, bensì strumenti in grado di potenziare e di dare uno slancio diverso alla creatività e alla personalità dei diversi singoli artisti che ne fanno uso. In questo senso, si aprono molte strade all'uso creativo dello strumento per una proficua collaborazione. Il ruolo dell'essere umano in relazione al processo di elaborazione e generazione di opere da parte delle macchine, appare a questo punto cruciale per il dibattito.

Per quanto i progressi siano in continua evoluzione, per quanto queste tecnologie offrano una possibilità di indagine ancora ampia e, per quanto sia vero che un margine di creatività sia possibile nel processo di generazione da parte delle GAN, questa riflessione

rende esplicita la fondamentale importanza della partecipazione dell'uomo nel processo creativo.

L'evoluzione tecnologica ci sta facendo scoprire possibilità inimmaginabili, gli sviluppi in questo campo nel giro di pochi anni hanno raggiunto traguardi imprevedibili. L'intelligenza artificiale offre la possibilità all'uomo di potenziare le proprie capacità e resta un ambito che attira l'attenzione di diverse discipline, che possono collaborare tra loro. È così che viene riunita l'attenzione di scienziati, informatici, filosofi, giuristi, storici dell'arte. Fino a che punto arriveranno queste possibilità? Vi sono effettivamente dei parallelismi con il funzionamento della mente umana e se sì, lo sviluppo e l'accrescimento delle capacità da parte delle macchine dove può portare? La creatività della macchina e il modo di analizzare, vedere e rielaborare la realtà è effettivamente paragonabile alla mente umana?

Resta ancora difficile dare una risposta definitiva a tutti questi interrogativi, ma indubbiamente, per quanto riguarda l'uso in campo artistico, la partecipazione dell'uomo appare ancora essenziale: egli è ancora in una posizione di potere, potendo guidare, addestrare e collaborare con la macchina.

L'analisi degli artisti e l'esposizione delle loro opere, arricchita dalla riflessione sulla creatività, ci ha fatto comprendere come un'interazione sia possibile e come, spesso, questa possa dare vita anche a nuove ed inesplorate vie creative per l'artista. La collaborazione non toglie nulla alla "mano umana" in questo caso: ogni artista sceglie quanto e come farne uso.

Il futuro resta da scoprire, ma ad oggi lo spirito artistico dell'essere umano, la sua sensibilità e la sua capacità di espressione attraverso l'arte, non sono ancora sostituibili. Una macchina non può dare un senso a ciò che crea, non può, ad oggi, formulare propriamente un concetto e fornire di senso una esperienza vissuta. Queste sono tutte capacità umane, che restano ancora tali. Rimane, dunque, all'uomo la scelta: fare uso o meno di questo strumento: questa libertà è ancora in suo potere. Come tutte le grandi rivoluzioni, il margine di riflessione e di dibattito a riguardo resta ancora ampio. Ma, con esso, anche la possibilità di sorprenderci.

Conclusioni

Riprendendo gli intenti iniziali, si spera di aver dato una nuova luce e delle nuove prospettive su un argomento molto attuale.

Inizialmente sono stati presentati i fondamenti teorici del discorso, facendoci guidare dal pensiero di alcuni importanti filosofi come Husserl e Baudrillard. La prospettiva fenomenologica di Husserl è stata approfondita per quanto riguarda l'aspetto teorico dell'iperrealismo, ma adattandola anche al discorso più critico sulla realtà contemporanea e sull'epoca della simulazione dovuta ai nuovi media immersivi e alla diffusione di *fake news*.

È stata messa in luce una possibile “via negativa” per quanto riguarda l'uso dell'intelligenza artificiale: spiegando in maniera più tecnica cosa si intenda con i termini deepfake e GAN ne sono stati mostrati gli aspetti, se vogliamo, più pericolosi e potenzialmente diffamatori. In questo caso non ci si è limitati a definire e spiegare un fenomeno, ma si è cercato di dare anche uno spazio per la riflessione sulle eventuali problematiche, sulle difficoltà create da un uso sbagliato di questi mezzi e su eventuali possibili soluzioni, lasciando dunque aperto un dibattito e una possibile ricerca su un impiego diverso, più positivo e su possibili metodi per “difendersi” da questi nuovi mezzi tanto sofisticati quanto potenzialmente pericolosi.

Infine, è stata mostrata una nuova, possibile strada da intraprendere: un uso dell'intelligenza artificiale e delle GANs, già sperimentato con successo da alcuni artisti e ricercatori, applicato al mondo dell'arte e alla creazione di opere d'arte che uniscano capacità umane a capacità computazionali. Ritengo che sotto questo punto di vista ci sia ancora ampio margine per la ricerca: sia per quanto riguarda gli sviluppi, nonché i possibili usi e applicazioni in diversi campi delle GANs (o CANs) sia per quanto riguarda il dibattito sull'intelligenza artificiale: può essere proficua un'interazione uomo/macchina? Le macchine possono essere creative? La macchina è in grado di potenziare le capacità umane e il lavoro dell'uomo in diversi campi e se sì, come avviene e che conseguenze ha tutto ciò in campo artistico?

Si è cercato di dare una risposta ad alcuni di questi interrogativi, di mostrare il lavoro di ricerca e di sperimentazione artistica compiuto fino ad ora in questo campo e di delineare a grandi linee il dibattito contemporaneo, ma resta ancora ampio margine per la ricerca e l'approfondimento in un campo che è in continua evoluzione.

La speranza è quella di aver definito un quadro generale sull'argomento e di aver fornito allo stesso tempo alcuni spunti per future riflessioni.

Appendici: Intervista a Mauro Martino

1) **C:** Da quanto tempo lavora con l'intelligenza artificiale applicata all'arte e, in particolare, con le Generative Adversarial Networks?

M: Ho iniziato a lavorare con l'intelligenza artificiale, con l'obiettivo di comprenderla e visualizzarne i modelli, a partire dal 2014. Ovviamente il mio lavoro è stato sempre un ibrido tra attività di ricerca o di progetti attinenti all'università (soprattutto in collaborazione con Carlo Ratti, tante visualizzazioni fatte per giornali come *Nature* e *Science*) e il desiderio di avere uno spazio personale come artista, o meglio come persona che si esprime attraverso gli strumenti che ha. Nel mio caso sono bravo a scrivere codici e a partecipare all'attività creativa, visuale, filmica attraverso questo stile di *computational design*: un design che si esprime soprattutto attraverso codici. Il codice prima significava soprattutto "linguaggi parametrici": modi in cui le forme cambiano, evolvono e sono generative ma all'interno di equazioni super controllate. Dunque, una complessità notevole, ma nulla a confronto di ciò che si raggiunge attraverso le reti neurali. Lo scarto importante è stato quando dal 2013-14 e soprattutto 2015 ho iniziato ad occuparmi di reti neurali, il che è stato molto faticoso. Ho iniziato a collaborare con colleghi nuovi, sono entrato a far parte di una nuova unità che si occupava tanto di intelligenza artificiale, circondato di esperti di A.I., in un nuovo ufficio a Manhattan. Il grande boom è scoppiato nel 2012, quando un *paper* ha dimostrato che le reti neurali erano pronte a risolvere meglio di altri sistemi matematici alcuni problemi di comprensione di *astrature data*. Ovviamente l'analisi dei dati

unstructure (video e testi fatti anche a mano libera, tutto quello che non è “dentro una tabella”) è sempre stato visto in maniera molto importante, tanto da investire molto denaro per svilupparlo anche manualmente. L’archiviazione di queste informazioni è cruciale per capire anche la storia e le tendenze, i *trend*, e per prevedere in qualche modo il futuro. Da quando c’è il *deep-neural-net* e ci sono questi computer così potenti, tutto questo linguaggio in CUDA per programmare questi modelli all’interno delle architetture Nvidia, tutta questa potenza che inizialmente emergeva grazie alla *cryptocurrency* (Nvidia aveva un boom di vendite, i prezzi delle schede grafiche erano aumentati notevolmente), tutti questi soldi che arrivavano dalla *cryptocurrency* erano utilizzati da Nvidia per evolversi. Dunque, Nvidia al tempo aveva abbastanza soldi per iniziare a pensare a come indicare queste reti non soltanto per i videogiochi, ma anche per la *cryptocurrency* e per il *trend* che stava per emergere dell’*Artificial Intelligence*. Personalmente, mi sono ritrovato tra tutti questi esperti, tra questi scienziati e ho capito cosa sono e come si comportano le *deep-neural-network*. Inizialmente la possibilità di visualizzazione non è andata subito “in porto”: anche se ne ho visualizzate tante, nessuna di queste visualizzazioni dava una chiara idea di quello che stava avvenendo all’interno. Sono, per definizione, *black box*, perché è troppo complessa l’attività che queste reti fanno per essere visibilmente intelligibili, ci sono altri strumenti per semplificarne il messaggio e i comportamenti. Al tempo mi occupavo di avere il mio spazio in campo artistico: quindi presente in diversi musei con il mio lavoro, mi sono trovato a fare la stessa cosa con le nuove *expertise* e le nuove conoscenze, quindi anche le A.I. diventavano un nuovo strumento per esprimermi. La maggior parte dei miei lavori che si conoscono sono soprattutto quelli artistici perché quelli scientifici spesso sono dei *white paper* interni, dedicati agli scienziati, che non si rivolgono al grande pubblico. Sul mio sito web, così come nelle interviste con i giornalisti, ci si dedica quasi sempre alla sfera artistica. C’è stata questa esplosione delle A.I. all’interno dell’attività artistica più legittima e ufficiale, più famosa di me, e quindi mi sono ritrovato come “uno tra i tanti”, ma tuttavia credo di essere stato tra i primi a fare questi esperimenti, anche esperimenti come *news explore* in realtà hanno tantissimo di A.I., è una visualizzazione costante di un flusso di informazioni che viene costantemente selezionato da *engine* di *artificial intelligence*. Quello che è A.I. e quello che non lo è, sarebbe un’altra ampia domanda.

2) **C:** Giustamente ha sottolineato l'aspetto scientifico dei suoi lavori, e infatti volevo farle una domanda un po' più specifica riguardo le GAN, al funzionamento di queste reti neurali attraverso cui ha realizzato appunto anche alcune delle sue opere. Dal punto di vista scientifico non mi era molto chiaro il funzionamento delle due reti neurali: il *discriminator* e il *generator*. Mi è chiaro il lavoro del *discriminator* (che funziona attraverso un *training set*, basandosi su questo *database* di immagini, per dare una risposta che sia "vero" o "falso", per dire se l'immagine creata dal *generator* sia artificiale o meno). Invece non mi è molto chiaro l'*input* iniziale del *generator*: quando inizia a creare, prendiamo ad esempio, un volto, in molti schemi l'*input* iniziale viene identificato con uno *scratch*, un rumore casuale di pixel attraverso cui inizia a creare, ma non riesco a capire se non avendo un *database* su cui basarsi inizi a creare "a scatola chiusa", e come possa iniziare a creare un'immagine così complessa come, ad esempio, un volto: come possa quindi creare senza l'iniziale risposta del *discriminator* o senza basarsi su un *database* di immagini. Dunque, inizia davvero a creare attraverso un "rumore casuale", uno *scratch* di pixel, oppure c'è qualcosa su cui si basa per iniziare a generare queste opere, che possono essere volti o paesaggi o anche altre opere?

M: La domanda è giusta perché c'è un fraintendimento su come funziona il *discriminator*. Se il *discriminator* fosse in grado di riprodurre artificialmente le immagini dei volti e dei paesaggi così come sono dall'inizio, avendoli già visti, non ci sarebbe nulla da fare per il *generator*, perché non avrebbe quasi nessun dialogo. Nessun *output* sarebbe vicino al *target* e sarebbe costantemente "bocciato". Ma come fa a riprodurre quelle forme? Seguendo questo ragionamento si potrebbe aspettare all'infinito. Qui si sbaglia nella visione del *discriminator*: il *discriminator* in realtà parte "senza sapere nulla" a sua volta, ci troviamo così davanti a due componenti che, entrambi "non sanno nulla". Quindi all'inizio è come se andasse bene qualunque cosa: il *generator* veramente viene attivato con un *random vector*, inizia a creare da zero, da un *random number*, ma anche il *discriminator*. È lo stato iniziale, bisogna partire da qualcosa. Il discriminatore sa cosa fare: deve imparare a rappresentare queste immagini che sono nel *training data*, anche il generatore sa cosa deve fare: deve

iniziare a produrre immagini vicine al *discriminator*. Il generatore e il discriminatore per fortuna sono identici, quindi man mano che evolve il *discriminator*, man mano che i pixel si collocano in maniera più vicina ad un volto, che da “una nube” prendono sfumature sempre più definite, per il generatore non sarà difficile avvicinarsi a sua volta. Il processo è simultaneo: se il *discriminator* impara troppo in fretta sparisce il dialogo costruttivo con il *generator* e, di conseguenza, il *generator* non evolve in fretta. Se invece manteniamo “lenta” e soprattutto simultanea l’evoluzione di entrambi, facendo così il *generator* riesce ad evolvere velocemente e a seguire gli stadi di evoluzione del *discriminator*. Pian piano si portano insieme verso la costruzione di immagini sempre più realistiche. Quindi, in realtà, si tratta di un “crescere insieme” che permette al *generator* di passare dalla “nube” di *random vector* al *target*: perché la stessa cosa fa il *discriminator*.

C: Pur basandosi quest’ultimo sul *training set*, se non sbaglio.

M: Certo, lui guarda il *training set*. L’unico che infatti può evolvere in questo percorso insieme è il *generator*, tuttavia però l’altro lo segue, perché la sua missione è seguirlo. Lo segue, lo segue e lo porta al risultato. Perché si fa questa cosa? Per permettere a noi di avere immagini che sono simili a quelle del *training data*, ma che non appartengono al *training data*. È molto utile nel caso in cui vogliamo fabbricare *data* e fare il *training* di altri modelli. È costosissimo, ad esempio, e anche legalmente molto difficile, scambiarsi radiografie di pazienti e sono rare le radiografie con un certo tipo di *disease*, dunque cosa si fa? Le si fabbrica artificialmente in modo tale da avere tantissimi dati che altrimenti sarebbero molto difficili da avere e questi dati artificiali possono essere usati per tantissimi altri modelli che hanno lo scopo di trovare quel tipo di *disease* (diventano settoriali). Quindi è utilizzato molto per lo sviluppo di *fake data*, per fare il *training* di altri modelli.

C: Capisco, avevo proprio questo dubbio sul *generator* e su questo *input* iniziale, *from scratch*.

M: Perché in realtà si tratta di un *misunderstanding* del *discriminator*.

- 3) **C:** Come ha detto giustamente, attraverso il lavoro di queste reti neurali, c’è lo spazio per creare immagini che sono simili al *training set*, ma d’altra parte diverse. Quindi non semplicemente delle copie del *training set*, ma delle immagini rielaborate dalle

reti neurali e dall’A.I. Questo se non sbaglio lascia spazio, possiamo dire in un certo senso, anche alla creatività della macchina. A questo proposito le volevo chiedere se secondo lei è possibile dire che le macchine sono dotate, in questo senso, di creatività. Perché c’è ancora un forte dibattito attorno a questo argomento della creatività delle macchine, dunque volevo chiederle cosa ne pensa a riguardo: se le macchine possono essere intese, in qualche modo, come “dotate di creatività” o se questa è una caratteristica che appartiene solamente all’uomo.

M: Beh, bisogna dire intanto che ci sono diversi tipi di creatività. Ci sono anche poeti che mettono delle parole in un sacchetto e poi le estraggono per fare una poesia. In quel senso anche una macchina potrebbe farlo. La creatività intesa come capacità di guardare il mondo, di capirlo, di interpretarlo secondo le proprie sofferenze e le proprie passioni e quindi “soggettivizzare” la visione del creato, il bello della natura e renderlo artefatto che gli altri possano leggere, comprendendo, in qualche modo, anche l’autore, questa è un’utopia. L’unico modo sarebbe costruire in maniera artificiale una personalità di una macchina: si creano così dei finti personaggi robotici che sembrano avere una loro storia, diventando l’autore di quel robot che lo programma, e quindi guardando i suoi artefatti si cerchi di trovare questo aspetto. Però sarebbe tutto inutile perché il bello dell’arte rimane la comprensione di noi stessi. Gli artefatti prodotti dalle reti neurali sono sempre da riferire ad un artista che usa la rete neurale, non si scappa da qui. Il discorso può essere anche molto machiavellico, molto cinico, può essere anche molto sofista, ma il succo è che la nostra interpretazione di creatività nell’arte è quasi sempre legata al piacere di scoprire un’altra anima. Quindi qualunque sia il senso dell’anima dato dalle diverse culture, essa si esprime anche attraverso l’uso di reti neurali, che però rimangono uno “scalpello moderno”. È sicuramente giusto cercare di capire l’estetica delle A.I., però questa è legata alla percezione dell’estetica di chi in quel momento sta utilizzando quel determinato strumento. Non ci illudiamo che le reti neurali siano le A.I.: queste non le abbiamo ancora in mano, l’intelligenza artificiale ancora non è tra noi e non sappiamo quando arriverà. Abbiamo ad oggi strumenti un po’ più sofisticati: con le *deep-neural-network* abbiamo fatto dei grandi passi avanti, si tratta di una grandissima tecnologia. Sono però dei modelli matematici che rimangono nel confine dei modelli matematici. Chi li sa usare con coscienza li definisce strumenti, chi li usa senza avere la più pallida idea di

come funzionano può trovare “il divino” le A.I., o quello che vuole. Quindi che sia creativa la macchina dipende dalla profondità e dalla capacità di lettura di chi guarda l’opera.

C: Certamente. E giustamente, come ha detto, si tratta sempre di uno strumento che è in mano all’artista. Quindi è l’artista che in base alla propria sensibilità, al proprio modo di vedere il mondo, al proprio sguardo e la propria personalità che utilizza questo strumento. Quindi ogni opera creata da ogni artista, pur usando lo stesso strumento (che possono essere le GAN, le reti neurali) avrà qualcosa dell’artista. Ci sarà, diciamo, uno scambio, un arricchirsi vicendevolmente tra la macchina e l’artista, ma sempre con la presenza umana dell’artista nell’opera. Questo va in direzione contraria rispetto a molte visioni che possono essere anche negative sull’uso delle macchine, che possono anche spaventare: credo ci sia ancora, da parte della maggioranza delle persone, una sorta di scetticismo sull’uso di questi strumenti, anche a tratti fantascientifico. Anche nel caso delle prime opere create attraverso l’uso di questi strumenti si è creato inizialmente un forte dibattito. Come lei giustamente ha detto, però, si porta sempre qualcosa dell’artista nell’opera: la sua sensibilità, la sua personalità.

M: Purtroppo spesso con la facilità di replicare questi codici assistiamo anche a “truffe”, ad artisti che replicano in maniera poco originale il lavoro di altri senza portare molto della propria soggettività. Diciamo che il 99% è *scan*: si tratta di gente che ha il proprio gusto estetico e che prende gli algoritmi fatti da altri (da scienziati ad esempio) usandoli a proprio modo, vendendoli come propri. Sarebbe come se nel passato un grande pittore avesse definito a proprio modo una tecnica e proteggendola, prima era più semplice perché magari l’artista aveva la propria bottega e poteva proteggerla e trasmetterla ai propri discepoli. In questo caso nel mondo scientifico non si dà molto peso all’arte, essendoci anche molti altri obiettivi, quindi si lascia quasi tutto *free*. Questo, però, lascia spazio al “prendere” senza, a volte, dare credito. Spesso si tratta anche di marketing e di vendite a prezzi elevati. Personalmente penso che un pennello e dei colori diano campo espressivo a tutta l’umanità, mentre maneggiare modelli di reti neurali e renderli propri, esprimersi attraverso di essi richiede livelli di conoscenza che non tutti hanno. C’è una parte di A.I. che è fatta dagli scienziati per gli artisti: questo già succede, ma sicuramente arriverà quel momento in cui uno scultore modellerà quella creta e cliccherà un pulsante per far diventare quell’oggetto di

creta di marmo, grazie ad un braccio che si autoprogramma e prende spazio e volumi (anche qui si tratta di reti neurali, che sono ovunque non solo nelle GAN). Così non dovrebbe più usare lo scalpello per produrre un'opera di marmo, prima invece il braccio meccanico non era in grado di fare perché il controllo numerico non era basato su algoritmi così sofisticati come adesso. Un pittore magari in futuro potrà dire alle A.I. che opera vorrebbe, prima insegnandogli come dipinge, insegnando il proprio gusto estetico, per poi descriversi. Potrà farsi vedere da questo *tool*, far capire chi è, la propria "spiritualità", il proprio modo di vedere, per poi descrivere come vuole che venga fatta un'opera. Il sistema potrà così dipingerla, chiedendo con quale tecnica. In un mondo in cui abbiamo un controllo sempre più sofisticato di queste tecnologie, potremmo avere artefatti che partono dal nostro linguaggio o dai nostri pensieri. Arrivati a quel livello qualunque artefatto sembrerà un artefatto tradizionale: una scultura fatta secondo lo stile di Canova, un dipinto secondo lo stile di Da Vinci. Sicuramente sono però tecnologie che ci spingono verso un'arte concettuale.

C: Capisco. Sarà quindi molto complicato vedere anche l'evoluzione di queste tecnologie, siamo ancora in un campo molto di sperimentazione se non sbaglio.

M: Di grande confusione del mondo artistico, del mondo umanistico, sicuramente. Però saranno tecnologie considerate sempre più necessarie, piacevoli, capaci di fare tutto. Ora, che l'arte sia la capacità di dipingere, la capacità tecnica è stata per molto tempo uno dei fondamenti dell'essere artista. Però abbiamo perso questa connessione con la tecnica già da tanto tempo. Quindi che l'arte sia diventato concetto lo vediamo anche nel mondo dell'architettura: lì per primi ci si è mossi dal talento dell'architetto al talento dello studio (inteso come un gruppo di persone che lavorano per te). Le A.I. in qualche modo racchiudono questo concetto: non sei solo tu, tu dai l'idea e lo studio in qualche modo la fa. Nello studio dell'architetto contemporaneo spesso l'architetto mette solo la firma. Quindi si è persa la connessione tra l'autore e le sue conoscenze tecniche. Le A.I. consentiranno di farlo a molte più persone, non solo quelle di successo che possono avere degli impiegati. Le A.I. considerate come "agente che si sostituisce all'uomo" esistono già, in qualche modo.

C: Molto interessante come pensiero. Quindi non una svalutazione dell'arte, ma se vogliamo, una democratizzazione in qualche modo e un contribuire alle capacità dell'artista.

Quindi non “macchina che sminuisce l’uomo”, che lo desensibilizza, che ne toglie la capacità di interpretare la personalità e lo sguardo.

M: Abbiamo tante delusioni riguardo gli artefatti prodotti dall’umanità, internet ci ha convinto che offrendo conoscenze a tutti *free*, la più alta conoscenza possibile, allora avremmo avuto tanti geni nel mondo. Magari i nostri figli, la nuova generazione sarebbe stata molto più preparata, avendo tutto. Invece si è creato un fenomeno opposto: l’idea che avere tutto disponibile renda la cultura poco importante. C’è la percezione forte che, ad esempio, perché posso fare una ricerca in pochi secondi, allora tanto vale non farla. E anche in questo caso: l’idea che un dipinto fatto con tutte le tecniche giuste, siccome a queste tecnologie si possono insegnare diverse tecniche, diversi *modus operandi* per fare un dipinto ad olio, ad acquerello, scultura su marmo, qualunque sia la tecnica delle A.I. resa disponibile a tutti, veramente democratizza o rende l’oggetto, l’artefatto inutile agli occhi della popolazione? In questo senso non sappiamo dove si va. Se l’artefatto, per quanto bello, per quanto sofisticato, può essere prodotto da chiunque, l’artefatto avrà un valore? La popolazione darà ancora valore a qualcosa che tutti possono fare? È un grosso dilemma, non sappiamo in che direzione si va. Sicuramente non perché qualcosa è disponibile per tutti, rende quel qualcosa di valore. Può anzi perdere completamente il suo valore.

C: E però, come abbiamo già detto, non sappiamo quello che succederà.

- 4) **C:** Per quanto riguarda gli *A.I. Portraits*: in quest’opera, nello specifico, c’è un modo di coinvolgere in maniera diretta lo spettatore, il fruitore. Chi si trova ad usare questa piattaforma poi è direttamente coinvolto nell’uso, non solo: vedrà la propria immagine riprodotta attraverso l’elaborazione della macchina in diversi stili pittorici. Secondo lei questo è stato un modo per avvicinare anche il pubblico con meno conoscenze possibili riguardo l’argomento dell’intelligenza artificiale, riguardo le reti neurali, per coinvolgere quindi una maggioranza di persone che ancora non sono ben a conoscenza di questi argomenti, oppure siamo ancora distanti da questo coinvolgimento?

M: In una conferenza¹⁹⁴ per un evento sul futuro del design alla *Harvard University* spiego ufficialmente quello che volevo ottenere da *AI Portraits*. Ad oggi posso dire qualcosa in più sulla capacità di democratizzare, di diffondere: forse la risposta è no. L'idea che io possa rendere il mio volto dipinto da Raffaello o altri artisti già ad oggi non interessa più, è visto quasi come un filtro di due anni fa. L'idea che ci possa essere qualcosa che prima non c'era e oggi c'è, quindi lo usi, non ha funzionato, così come non ha funzionato l'idea che siccome la conoscenza veniva maggiormente diffusa e sarebbe quindi stata usata. Quello che sta accadendo è che vi sia un'idea di *trendy* e *non trendy*: adesso aspetto che l'A.I. faccia qualcosa di nuovo che non ho mai visto prima. Se esce un'A.I. in grado di animare un volto statico, per un paio di settimane diventa *trendy*. Ma dopo aver capito che l'A.I. lo sa fare, si perde l'interesse. Questi avanzamenti, quindi, non stanno procurando un avvicinamento all'arte, sicuramente stanno procurando un senso di vago riempimento, di filtro di moda o non di moda. Una volta che esiste uno strumento in grado di dipingerti come un grande artista del passato, non c'è poi una versione 2 o 3. Quindi l'idea che qualcosa di così sofisticato passi di moda può fare male, anche se nel mio caso il progetto è stato volutamente chiuso "all'inizio del picco". Però nonostante questo, per quanto sia durato poco, 10 milioni di persone si sono fatte fare il ritratto. Spero l'abbiano conservato, perché non ce ne sarà un altro fatto da *AI Portraits*. Ma la difficoltà che ve ne sia un altro dopo è che quell'effetto visivo è già stato consumato, che rimane in qualche modo un effetto visivo (siamo anche vittime della cultura di *Photoshop* che avanza ogni anno, ogni anno infatti vengono creati nuovi filtri). E così nel mondo dell'estetica l'immagine viene trasfigurata attraverso modelli di reti neurali, dove all'inizio c'era lo *style-transfert*: compagnie come Prisma che creavano *app* (prendo la tua foto dando i colori di un'altra foto, non ne modifico le forme ma i *pattern*, la foto rimane la stessa). Poteva essere "filtrata" come un quadro di Van Gogh: con gli stessi colori, ma la foto era la stessa. Le foto rimanevano identiche e questo non era verosimile, perché Van Gogh non avrebbe mai dipinto in quel modo, non mi avrebbe fatto uguale identico alle foto: l'artista aveva il suo modo di dipingere, tutto tornava in una sua estetica.

194

<https://vimeo.com/395227911?fbclid=IwAR2myqFBB0KW2uvLt1txPpsvBKz3Sxjqp1xWtbqMmWG9J88BtSQeO9JUw6I>. URL visitato in data 25 settembre 2021.

Era poco credibile che sostituire i colori fosse bastevole per ridare il senso di essere dipinto da un altro artista. Quindi quello sforzo andava in quella direzione: di sentirsi ridisegnato e anche non somigliante: quello sforzo di calibrare la somiglianza, perché nessun artista ti disegna per quello che sei, ma per come ti vede. Ed è vero: l'artista ti vede a modo suo, ha i suoi "vezzi", le sue ricorrenze estetiche che riportano il suo modo di interpretare. La cosa più bella da regalare a chi si fa dipingere è l'idea di farsi sentire un po' inadeguato ogni volta che vede il proprio dipinto. Chi è seduto di fronte all'artista quando guarda il risultato non apprezza ciò che vede molto spesso, dice "io non sono così". Lo so perché mio padre dipinge da prima che nascessi e ha fatto tanti *portraits*, ma spesso anche se sapeva disegnare le persone non erano soddisfatte, dicevano "c'è qualcosa di me": questo era il commento di chi veniva ritratto. Chi si vede ritratto non si auto-riconosce facilmente. (Tornando a *A.I. Portraits*) era piacevole vedere tutto quel successo, mi sarebbe stato utile anche per vedere il prosieguo di questa tecnologia, di vederla applicata da altri attori. In realtà quando si cerca di capire lo stile di un artista, di darlo nel fare il *portrait*, è diverso da (ad esempio) trovare i canoni estetici di un cartone animato 3D, perché in quel caso ci sono delle ricorrenze (5 o 600 personaggi in totale, bisogna vedere quello che somiglia di più al tuo volto, però si va verso lo stereotipo, la cartonificazione è più un processo di stereotipizzazione). Il mio lavoro però non andava in questa direzione, andava più nella direzione di comprendere come un artista era abituato a modificare le forme, l'estetica e non solo il colore.

C: E giustamente, come tutte le cose innovative, è stato un processo che ha coinvolto molte persone ma come ha detto è stato anche un processo veloce.

M: Volutamente veloce: prevedendo i tempi in base ad un obiettivo che mi ero prefissato ho voluto chiuderlo in tempo.

C: Resta, comunque, un buon modo per avvicinare le persone a questo argomento.

M: L'argomento delle A.I. è molto affascinante, non perché è affascinante l'evoluzione delle A.I. che non accade mai. È bellissimo vedere come l'argomento sia sempre attuale nel tempo: lo era nell'Ottocento e lo è adesso, sembra sempre che ci sia. C'è sempre qualcuno nella storia dell'umanità, da quando abbiamo la tecnologia e da quando c'è il computer, c'è sempre qualcuno che crede ci possa essere un computer come noi, che parla come noi, è una ricorrenza. Lo si vede in tantissimi testi, non solo di fantascienza: l'idea che

sta per arrivare, è quasi lì, che il test di Turing è stato soddisfatto. Non smettiamo mai di crederci e questo è molto affascinante.

C: Certo, purtroppo c'è anche molta influenza da parte della fantascienza relativamente a questo discorso, per la maggioranza rimane un discorso relegato a questo ambito, mentre invece è un argomento molto attuale e applicabile a molti campi concreti, in qualche modo.

M: Sì, tantissime cose. In realtà non c'è ancora un'intelligenza artificiale, non c'è e non sappiamo quando ci sarà (intesa proprio come intelligenza umana riprodotta artificialmente), però noi abbiamo la percezione che ci sia perché la maggior parte delle persone attorno a noi hanno accettato lavori manuali (o sono stati costretti). Alcuni di questi lavori manuali ci sono sempre stati, eppure ora per la maggior parte non vengono più svolti dalle persone ma dalle macchine, con una velocità impressionante. È un grande problema sociale quello che viviamo (lavori "meccanici" spesso fatti da esseri umani). L'arrivo di un'intelligenza artificiale sarebbe la singolarità: un'entità capace di essere come noi, creativa, investigativa, in grado di porsi degli obiettivi, tenace, in grado di andare avanti nei nostri percorsi di ricerca e soprattutto in quello suo: "come mi rendo io migliore". Noi non riusciamo a renderci facilmente più intelligenti, spesso si tratta di fattori casuali, ma non c'è un'evoluzione naturale che deve andare verso l'intelligenza, anzi, alcuni dicono che abbiamo raggiunto "il massimo" in alcuni campi (consumo energetico ecc.). Però la macchina, se fosse come noi, potrebbe evolvere e anche velocemente. Se evolvesse velocemente ci sarebbe sempre più intelligenza ("uomo 2, uomo 3 ecc.") e verrebbero scoperte cose che non si sarebbero mai scoperte altrimenti. Questo porterebbe alla singolarità: una sorta di "miracolo" in cui scoperta quella, tutto il resto si scoprirebbe nell'arco di pochissimo tempo. Siamo lontanissimi da questo, ma tutto quello che accade ora (macchina che si guida da sola, macchina che lavora al posto dell'uomo, ad esempio, in una fabbrica) scoprono un nervo sociale. Tante persone sono felici a svolgere questi lavori (più manuali che pensanti/che implicano l'uso di intelligenza). Il fatto che questo problema sia stato quasi risolto pone però altri problemi sociali giganteschi (ad alcune persone piace fare quei lavori, e cosa farebbero una volta sostituiti dalla macchina?). C'è una forma di "robotizzazione" dell'umano. Sicuramente il meccanismo non sarebbe un'immediata sostituzione di alcuni lavori con

l’A.I., non funziona così. Tutti questi non sono aspetti riguardanti solamente l’A.I., si tratta di modelli matematici sempre più sofisticati, sempre più veloci, in grado di sostituire alcune attività umane non pensanti. Noi consideriamo, ad oggi, intelligente l’umano perché era in grado, a differenza della macchina, di dialogare con il mondo (“io vedo, io sento, la macchina no”). Ecco, i modelli neuronali hanno scardinato questa differenza perché la macchina ora vede e sente molto meglio di noi, vede anche cose invisibili all’occhio umano. Questo evolversi di algoritmi matematici sempre più complessi pone problemi seri al contesto sociale e si svilupperanno o meno anche in funzione del contesto sociale, sicuramente.

C: Grazie, la sua riflessione va molto oltre le mie domande ed è molto profonda.

Bibliografia

Monografie (libri):

Barale, A., *Be my gan*, Jaca Book, Milano, 2020.

Baudrillard, J., *Il delitto perfetto: la televisione ha ucciso la realtà?*, Trad. di Gabriele Piana, Milano, R. Cortina, 1996.

Baudrillard, J., *Simulacri e impostura: bestie, Beaubourg, apparenze e altri oggetti*, a cura di Matteo G. Brega, Milano, Pgreco, 2008.

Boden, M., *The Creative Mind: Myths and Mechanisms* (second edition), Routledge, 2003.

Conte, P., *In carne e cera. Estetica e Fenomenologia dell'iperrealismo*, Quodlibet, Macerata, 2015.

Husserl, E., *Fantasia e coscienza di immagine*, a cura di C. Rozzoni, Rubbettino Editore, Soveria Mannelli, 2017.

Husserl, E., *Ricerche logiche* (1900-1901), tr. it. e cura di G. Piana, 2 voll., il Saggiatore, Milano 2001, vol. 2, pp. 229-230.

Riva, G., *Fake news. Vivere e sopravvivere in un mondo post-verità*, Il Mulino, Bologna, 2018.

Schlosser, J., *Storia del ritratto in cera* (1911), a cura di P. Conte, intr. di G. Didi-Huberman, postfaz. di P. Conte, Quodlibet, Macerata, 2011.

Wiesing, L., *Artificial Presence. Philosophical Studies in Image Theory* (2005), Stanford University Press, Stanford, CA, USA, 2010.

Wiesing, L., *Artifizielle Präsenz Studien zur Philosophie des Bildes*, Suhrkamp, 2005.

Articoli e saggi:

Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, Marian Mazzone, *CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms*, arXiv:1706.07068v1 [cs.AI].

Barale, A., *A hideaway in the wastelands: nuove sfide filosofiche dell’AI art*, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020.

Berlyne, D.E., *Aesthetics and psychobiology*, volume 336. JSTOR, 1971.

Berlyne, D.E., *CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms*, p. 3.

Calì, C., *Husserl e l’immagine*, in *Aesthetica Preprint Supplementa*, Centro Internazionale Studi di Estetica, 10 aprile 2002, p.35.

Chandell E. Gosse, Jacquelyn Burkell, *Politics and porn: how news media characterizes problems presented by deepfakes*, in *Critical Studies in Media Communication*, FIMS Publications, 2020.

Colton, S., *Creativity versus the perception of creativity in computational systems*, in *AAAI spring symposium: creative intelligent systems*, volume 8, 2008.

Conte, P., *Mockumentality: From Hyperfaces to Deepfakes* in WORLD LITERATURE STUDIES, vol. 11, pp. 11-25 (ISSN 1337-9690),

Conte, P., *Unheimlich. Dalle figure di cera alla Uncanny Valley*, PSICOART n. 2, 2011-12, pp.13-14.

Ferraris, M., *Postverità e altri enigmi*, Il Mulino, Bologna, 2017.

Freud, F., *Il perturbante* (1919), tr. It. di S. Daniele, in Id., *Opere*, a cura di C.L. Musatti, 12 voll., Bollati Boringhieri, Torino 1977.

Harwell, D., *Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*, in *The Washington Post*, 31 dicembre 2018.

Husserl, E., *Zur Lehre vom Bildbewusstsein und Fiktumbewusstsein* (1912), in *Husserliana*, vol. 23, cit., pp. 486-494.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, *Generative Adversarial Networks*, su *Communication of the Acm*, vol. 63, n.11, Novembre 2020.

Klingemann, M., *Memories of Passersby I*, 2018, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020.

Kris Belden-Adams, *ThisTheoryDoesNotExist: Historicizing and Understanding Artificial-Intelligence-Generated, Hyper-Real, Photographic Data Visualizations*, University of Mississippi, USA, Publishing House EIDOS, 2019.

Mazzone, M., *Le GAN e la questione della creatività nell'arte e nell'intelligenza artificiale*, in A. Barale *Be my Gan*, Jacabook, Milano, 2020

Mori, M., *The Uncanny valley* (1970), tr. ingl. di Karl F. MacDorman and Norri Kageki, IEEE Spectrum, 2017.

Moruzzi, C., *Alla ricerca della creatività: le GAN come paradigma dell'autonomia nel software per la composizione musicale*, in A. Barale, *Be my gan*, JacaBook, Milano, 2020.

Obvious, *La Famille de Belamy e i Sogni elettrici di Ukiyo: reinterpretazioni e accelerazioni*, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020.

Purgar, K., *The meaning of hyperrealism today: Reality, Iconic Difference, and Perception of Hyper-Transparent Images*, testo disponibile all'indirizzo:

https://webcache.googleusercontent.com/search?q=cache:2zt8PS5QVh0J:https://www.bib.irb.hr/961019/download/961019.Znaenje_hiperrealizma_danas.pdf+&cd=11&hl=it&ct=clnk&gl=ch

Ridler, A., *Set di dati e decadenza: Fall of the House of Usher*, in A. Barale, *Be my Gan*, Jacabook, Milano, 2020.

Shu Hu, Yuezun Li, and Siwei Lyu. (2020). *Exposing GAN-generated Faces Using Inconsistent Corneal Specular Highlights*. arXiv:2009.11924 [cs.CV].

Simmel, G., *La cornice del quadro. Un saggio estetico*, in *I percorsi delle forme, i testi e le teorie*, a cura di Maddalena Mazzocut-Mis, Bruno Mondadori, Milano, 1997, p. 210.

S. Zhang, H. Dong, W. Hu, Y. Guo, C. Wu, D. Xie, and F. Wu. *Text-to-image synthesis via visual-memory creative adversarial network*, in *PCM*, 2018.

Tero Karras, Samuli Laine, Timo Aila, *A Style-Based Generator Architecture for Generative Adversarial Networks*, arXiv:1812.04948v3 [cs.NE] 29 Mar 2019.

Thibaud, C., *Intervista con Jean Baudrillard*, in *Cybersphere*, n. 9, tr. it. di A. Venieri e I. Spelletti *Cybersphere*, n° 9, Arci S:M:S: di Rifredi-Special!, 1998.

Westerlund, M., *The emergence of deep fake technology: a review*, in *Technology Innovation Management Review*, 9(11) : 40-53.

Yuezun Li, Cong Zhang, Pu Sun, Honggang Qi, Siwei Lyu, *Deepfake-o-meter: an open platform for Deepfake Detection*, Ocean University of China, 2021, arXiv preprint arXiv:2103.02018.

Sitografia:

Ali Breland, *The Bizarre and Terrifying Case of the “Deepfake” Video that Helped Bring an African Nation to the Brink*, pubblicato su *Mother Jones*, 15 marzo 2019, <https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>

Alison Flood, *'Post-truth' named word of the year by Oxford Dictionaries.*, su *theguardian*, 15 Novembre 2016.

Anna Ridler, *Alice and Bob*, 2017, <http://annaridler.com/quantum-computing-art>

Anna Ridler, *No Replacements Found*, 2015, <http://annaridler.com/no-replacements-found>

Anna Ridler, *Leaked Documents: Wikileaks*, 2015-2017 <http://annaridler.com/leaked-documents>, video on Vimeo <https://vimeo.com/165036571>

Anna Ridler <http://annaridler.com/works>

Arcangelo Rociola, *Il finto video in cui Zuckerberg dice di avere il controllo delle nostre vite*, pubblicato su *AGI-Agenzia Italia*, il 12 giugno 2019, https://www.agi.it/economia/zuckerberg_facebook_deep_fake-5647497/news/2019-06-12/

Bankead M., *How to spot deepfakes? Look at light reflection in the eyes*, University at Buffalo, 10 marzo 2021, <http://www.buffalo.edu/news/releases/2021/03/010.html>

Cook, S., *Maintaining Composure: An Interview with Tamiko Thiel*, pubblicato su *Unthinking Photography*, gennaio 2021. <https://unthinking.photography/articles/maintaining-composure-an-interview-with-tamiko-thiel>

Cole, S., *AI-Assisted Fake Porn Is Here and We're All Fucked*, su *Vice*, 11 Dicembre 2017. <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>

Dan Robitzski, *Here's What AI Thinks A Bunch of Famous People Look Like*, pubblicato su *Futurism* il 18 agosto 2018. <https://www.fastcompany.com/90376689/what-you-look-like-as-an-renaissance-painting-according-to-ai>

Deepfake: l'Università di Firenze in prima linea contro i video falsificati, pubblicato su *InToscana* il 28 ottobre 2020, <https://www.intoscana.it/it/articolo/deepfake-universita-firenze/>

IArte, *Memo Akten: vedere e sognare con l'Intelligenza Artificiale*, pubblicato su *IArte: Nuove frontiere tra arte e Intelligenza Artificiale*, il 19 giugno 2019, http://iarte.altervista.org/memo-akten-intelligenza-artificiale/?doing_wp_cron=1631961150.9074130058288574218750

IArte, *Memories of Passerby: l'algoritmo che sogna volti*, pubblicato su *IA arte, Nuove frontiere tra arte e Intelligenza Artificiale*, il 27 maggio 2019, <http://iarte.altervista.org/memories-of-passerby-lalgoritmo-volti/>

Intervista a Mauro Martino, pubblicata su *Brera Design District*, 2021.
<https://www.breradesigndistrict.it/people-and-stories/mauro-martino/>

Jason Bailey, *AI Artist Robbie Barrat And Painter Ronan Barrot Collaborate On “Infinite Skulls”*, pubblicato su *Artnome* il 6 febbraio 2019.
<https://www.artnome.com/news/2019/1/22/ai-artist-robbie-barrat-and-painter-ronan-barrot-collaborate-on-infinite-skulls>

Katharine Schwab, *What you look like as a Renaissance painting, according to AI*, pubblicato su *FastCompany* il 16 luglio 2019.
<https://www.fastcompany.com/90376689/what-you-look-like-as-an-renaissance-painting-according-to-ai>

La minaccia del deep fake: come affrontarla in Italia? Discorso pronunciato nella Sala Koch di Palazzo Madama il 9 dicembre 2019, http://www.senato.it/4171?atto_presidente=4701

Le Università di Trento e Firenze contro il deepfake. Rilevare in rete contenuti multimediali manipolati e diffamatori, redazione ANSA, 27 ottobre 2020,
https://www.ansa.it/toscana/notizie/2020/10/27/le-universita-di-trento-e-di-firenze-contro-il-deep-fake_e5f7aea9-a0ca-4472-9b16-9170ee5310a8.html

Lend me your Face! - Tamiko Thiel, on *The Photographers Gallery*, 2021.
<https://thephotographersgallery.org.uk/whats-on/lend-me-your-face-tamiko-thiel>

Maria Francesca Fortunato, *Fatevi fare un ritratto (vero) dall'AI. Si ispira ai grandi maestri, ma inventa come un pittore*, pubblicato su *Italian.Tech* il 16 luglio 2019.
https://www.italian.tech/2019/07/16/news/fatevi_fare_un_ritratto_vero_dall_intelligenza_artificiale-302647695/

Mario Klingemann su Artsy.net, <https://www.artsy.net/artwork/mario-klingemann-appropriate-response-1>

Mario Klingemann: <https://quasimondo.com>

Mauro Martino: <https://www.mamartino.com/projects.html>

Massimo Gramellini, *Profondo Falso*, in *Corriere della Sera*, 26 settembre 2019. https://www.corriere.it/caffe-gramellini/19_settembre_26/profondo-falso-e1d251c8-dfcf-11e9-aa5f-fbca0c81b7c9.shtml

Memo Akten: <http://www.memo.tv/works/learning-to-see/>

Memo Akten: <http://www.memo.tv/works/#selected-works>

Moro, M., *Speciale Nft. Le case d'asta*, pubblicato su *Il giornale dell'arte* il 21 giugno 2021, <https://www.ilgiornaledellarte.com/articoli/speciale-nft-le-case-d-asta/136368.html?fbclid=IwAR0oB7cU-ULGOi7aeCUbj10wkyq4sE2gyxWze8HC7y-ri2P72EucozZYrxQ>

Obvious, <https://obvious-art.com>.

Oscar Schwartz, [*You thought fake news was bad? Deep fakes are where truth goes to die*](#), in *The Guardian*, 12 November 2018.

Raffaele Angius, *Un'intelligenza artificiale ha imparato a creare dei perfetti volti di persone che non esistono*, pubblicato su *La Stampa* il 20 febbraio 2019. <https://www.lastampa.it/tecnologia/news/2019/02/20/news/un-intelligenza-artificiale-ha-imparato-a-creare-dei-perfetti-volti-di-persone-che-non-esistono-1.33682515>

Rina Diane Caballar, *What is the uncanny valley?* Su IEEE Spectrum, 06 Novembre 2019 (<https://spectrum.ieee.org/what-is-the-uncanny-valley>)

Robbie Barrat: <https://aiartists.org/robbie-barrat>, <https://robbiebarrat.github.io/>

Srishti Deoras, *How GAN Was The True Artist In 2019*, pubblicato su *Analytics India Magazine*, il 10 dicembre 2019, <https://analyticsindiamag.com/how-gan-was-the-true-artist-in-2019/>

Tamiko Thiel: <https://tamikothiel.com>

Tensordream: <https://tensordream.ai>

The Lost Robbies, pubblicato su *SuperRare Editorial*, il 5 settembre 2020.

<https://editorial.superrare.com/2020/09/05/the-lost-robbies/>

Tom Simonite, *How a Teenager's Code Spawned a \$432,500 Piece of Art*, pubblicato su *Wired*, il 20 novembre 2018, <https://www.wired.com/story/teenagers-code-spawned-dollar-432500-piece-of-art/>

Vincent, J., *A never-ending stream of AI art goes up for auction*, pubblicato su *The Verge*, il 5 marzo 2019, <https://www.theverge.com/2019/3/5/18251267/ai-art-gans-mario-klingemann-auction-sothebys-technology>

When the artist isn't there, pubblicato su *Higher Nationals*, il 23 luglio 2019.

<https://www.highernationals.com/blog/when-artist-isnt-there>

[https://www.treccani.it/vocabolario/deepfake_\(Neologismi\)/?fbclid=IwAR2hbOxum5hh4hmd3CkE1qL1flKug8Q931nfi-eb7VQ0dKT3q323ngo6Wsc](https://www.treccani.it/vocabolario/deepfake_(Neologismi)/?fbclid=IwAR2hbOxum5hh4hmd3CkE1qL1flKug8Q931nfi-eb7VQ0dKT3q323ngo6Wsc)

Elenco delle illustrazioni

- 1) Schema grafico della *Uncanny Valley*, M. Mori.
- 2) Schema grafico del movimento della *Uncanny Valley*, M. Mori.
- 3) David Hanson (*Hanson Robotics*), robot umanoide *Sofia*.
- 4) Hiroshi Ishiguro, *Geminoid Hi*.
- 5) Hiroshi Ishiguro, *Telenoid*.
- 6) Università degli studi di Washington, progetto deepfake *Synthesizing Obama*, 2017.
- 7) Frame del video deepfake Donald Trump.
- 8) Max Planck Institute for Informatics, *Deep video Portraits*, Germania, 2018.
- 9) Schema di funzionamento dell'elaborazione di immagini di una GAN.
- 10) Schema grafico di una *Generative Adversarial Network*.
- 11) Schema grafico che mostra il miglioramento di una GAN negli anni 2014-2020.
- 12) Schema grafico di GAN nell'elaborazione di immagini *Text-to-image*.
- 13) Esempi di immagini generate attraverso l'uso delle GAN: i ritratti del sito *thispersondoesnotexist*.
- 14) Schema grafico di funzionamento del sito *thispersondoesnotexist*.
- 15) University of China, schema grafico dell'architettura della piattaforma *Deepfake-ometer*, 2021.
- 16) Raphaël Fabre, *CNI, french ID card with 3D portrait*, 2017.
- 17) Schema tecnico dell'architettura delle *Creative Adversarial Networks*, 2017.
- 18) Esempi di immagini artistiche generate attraverso l'uso di *Creative Adversarial Networks*.
- 19) *Edmond de Belamy, Obvious*, 2018. Algoritmi GAN, stampa inkjet su tela, 70x70 cm.
- 20) *La Famille de Belamy*, Obvious, 2018.
- 21) Algoritmo usato da *Obvious*, posto come firma in basso a destra alle opere de *La Famille de Belamy*.

- 22) *Memories of Passersby I*, Mario Klingemann, 2018. GAN multiple, due schermi, console realizzata a mano in legno di castagno che ospita il cervello dell'intelligenza artificiale e ulteriore hardware. Console: 70 x 60 x 40 cm. Ogni schermo: 82,9 x 145 x 3,8 cm.
- 23) *Uncanny Mirror*, Mario Klingemann, 2018.
- 24) *Appropriate response*, Mario Klingemann, 2020.
- 25) *Fall of the House of Usher*, Anna Ridler, 2017.
- 26) *Alice and Bob*, Anna Ridler, 2017.
- 27) *Infinite Skulls*, Robbie Barrat/Ronald Barrot, 2019, olio su tela, stampa UV su plexiglass.
- 28) *AI Generated Nude Portrait #1*, Robbie Barrat, 2018, immagine digitale prodotta da GAN.
- 29) Alcuni esempi di immagini generate da *AI Portraits Ars*, Mauro Martino, 2019.
- 30) *AI Portraits*, Mauro Martino, 2018.
- 31) *AI Paintings*, Chris Peters, 2019.
- 32) *Lend Me Your Face, Go Fake Yourself!*, Tamiko Thiel, Metronom, 2021.
- 33) *Learning to see*, Memo Akten, 2017.
- 34) *Gloomy Sunday – Learning to see*, 2017, Memo Akten, Installazione interattiva al Barbican Centre, 2019.

Ringraziamenti

Arrivati a questo punto, vorrei esprimere alcune parole per coloro che hanno contribuito non solo alla creazione di questo elaborato, senza il cui contributo non sarebbe stato possibile, ma anche per chi è rimasto durante tutto questo percorso di crescita personale e professionale. Innanzitutto, un ringraziamento va al mio relatore, il professor Pietro Conte, per la pazienza avuta e la perseveranza che mi ha insegnato, per i suoi consigli, gli incentivi e gli stimoli che mi ha trasmesso prima e durante la stesura dell'elaborato.

Un ringraziamento a chi ha contribuito ad arricchire il progetto, con suggerimenti e chiarimenti specifici in materia: Mauro Martino (fondatore e direttore del Visual Artificial Intelligence Lab presso IBM Research e professore presso Northeastern University) per l'intervista che mi ha concesso, il tempo dedicato al dialogo e agli interessanti spunti di riflessione. Ad Alfio Ferrara (professore di informatica presso Università degli Studi di Milano) per il suo tempo e le preziose spiegazioni in materia di reti neurali, GAN e scienza computazionale. Ad Alessandro Drigo (ricercatore e assistente presso l'Università di Lucerna) per le sue gentili risposte e i chiarimenti in ambito di Intelligenza Artificiale e GAN. Grazie a chi è rimasto sempre in questi anni: ai miei genitori e alla mia famiglia, che mi hanno sempre sostenuto e appoggiato nelle mie decisioni, dalla scelta del percorso di studi fino alla fine di esso e al raggiungimento di questo importante traguardo e che mi hanno sempre incoraggiato, anche nei momenti più difficili.

Grazie ai miei amici e colleghi, una seconda famiglia che mi ha accompagnato e sostenuto in questo percorso, condividendo gioie e spensieratezza, ma anche sacrifici e momenti di sconforto, stimolandomi con la loro tenacia e ambizione, aiutandomi con le loro parole, ma anche con la capacità di ascoltare in silenzio. Grazie per esserci stati fisicamente e moralmente.

Grazie alla “mia GAN”, che spesso negli scontri e nel dialogo mi ha fatto capire punti di vista molto diversi dai miei, facendomi crescere. Che mi ha insegnato cosa significhi avere coraggio nella vita, non aver paura davanti alle sfide, non aver paura di sognare troppo in grande. Che mi ha dato la giusta leggerezza nei momenti di pesantezza, ma mi ha anche

mostrato l'importanza di essere determinati e di avere fiducia in sé stessi e nelle proprie forze, prima di tutto.

Grazie ai momenti più belli, ma soprattutto ai più difficili, che mi hanno permesso di crescere e insegnato cosa significhi perseverare, nonostante tutto.