

Il volto latente tra ludico e politico

sensò e importanza degli stereotipi di DALL-E

Gianmarco Thierry Giuliana

Questo saggio è debitore dalla presentazione del Prof. Bruno Surace avvenuta durante il convegno “FACE IT! The New Challenge of Cognitive Visual Semiotics” tenutosi a fine gennaio 2023. Durante il suo intervento egli mostrava una immagine stilizzata di un simil-sé stesso generata da una intelligenza artificiale a partire da descrizioni verbali inserite dall’utente. Davanti a questa immagine stilizzata ma abbastanza somigliante di sé stesso, Surace descriveva il duplice effetto di somiglianza e alterità accompagnato dal duplice sentimento di soddisfazione e perturbamento davanti a questa rappresentazione.

Per comprendere l’efficacia semiotica e l’esito interpretativo di tali immagini bisogna, tuttavia, guardare innanzitutto alla loro generazione ancor prima che al loro contenuto. Esse sono il risultato di diverse tecnologie, tra cui l’impiego di *machine learning* basato sullo spazio latente che è l’oggetto al centro di questo libro. In termini non-specialistici, lo spazio latente può essere concepito come *uno spazio virtuale e matematico* in cui un certo oggetto dato come input (es: una immagine) viene *disposto* da una rete neurale ad una certa distanza da un altro sulla base di proprietà/qualità/caratteristiche comuni ricorrenti nel dataset (insieme del materiale usato per allenare la rete neurale). Queste proprietà vengono definite *dimensioni* dell’oggetto di partenza (input) e non sono riconosciute dalla macchina come dei significati o degli oggetti culturali ma sotto forma numerica. Questo riconoscimento permetterà poi non solo di identificare e categorizzare le varie occorrenze future ma anche di *produrre* una occorrenza *ex novo* sulla base di queste dimensioni e della loro topologia. L’esempio classico (LeCun 2019) di questo spazio è il seguente: forniamo come input a una macchina un migliaio di immagini di numeri scritti a mano (che presenteranno dunque variazioni grafiche come 5, 3, 3, 3, etc.), ed essa, senza sapere affatto cosa sia un numero, li disporrà ordinata-

mente in uno spazio secondo delle “coordinate” tali per cui numeri visivamente simili si ritroveranno vicini secondo le loro caratteristiche grafiche (troveremo dunque molti dei 3 molto vicini tra loro ma anche gli 8 molto vicini ai 3¹). In questo modo la macchina saprà non solo riconoscere futuri numeri ma anche produrne. Nel caso di tecnologie di *text-guided AI image generation*, come quella usata dal professor Surace, abbiamo uno spazio latente più complesso in cui le molteplici dimensioni di oggetti verbali (connessioni fra parole per cui pianoforte sarà, per esempio, caratterizzato dalla musica e più vicino alla parola “violino” che non a tostapane) condividono lo stesso spazio latente di oggetti visivi (l’esempio delle immagini di numeri precedenti ma anche foto di cani e gatti, etc.) creando così un metro di misurazione di *somiglianze semantiche* (di significati) tra immagini e parole. Se allo stato iniziale uno spazio latente è dunque uno spazio matematico di possibilità, nel suo stato “finale” esso diventa uno spazio di classificazione di informazioni e rappresentazioni. Come scrive Cristina Voto:

“All’interno delle Reti Generative Avverse, lo spazio di latenza è il risultato della compressione nascosta dei dati di ingresso e del passaggio che precede l’uscita della Rete Neurale, un passaggio che di solito rimane invisibile all’occhio umano, mantenendo la promessa di un effetto trasparente della realtà da parte dell’Intelligenza Artificiale. È: “lo spazio di possibilità della creazione, che definisce le coordinate di tutti i possibili output. Lo spazio latente puro è privo di vincoli e di significato: è una possibilità illimitata. Tuttavia, quando un modello generativo viene addestrato, restringe lo spazio di possibilità” (de Vries 2020: 2113).” (Voto 2022 p.55)

1

<https://pytti-tools.github.io/pytti-book/intro.html>

Tornando all'intervento del prof. Surace, noi vorremmo qui riflettere sul rapporto tra 1) la costituzione semiotica di queste immagini tramite una topologia latente e 2) l'effetto di senso interpretativo di questi volti creati tramite un uso su larga scala fondamentalmente ludico delle potenzialità dell'intelligenza artificiale. In altre parole, ci vogliamo qui chiedere: cosa vi è di socialmente significativo nel *giocare* con la somiglianza e la differenza percepita guardando l'immagine del volto di un simil-noi stessi creato tramite I.A.? Una domanda alla quale intendiamo rispondere tenendo insieme la doppia accezione di "latente": da un lato spazio *nascosto* e dall'altro lato come spazio *potenziale* del reale.

Per rispondere a questa domanda abbiamo replicato l'esperimento generando anche noi delle immagini a partire da descrizioni verbali tramite il sito di DALL-E², ma questa volta generando immagini fotorealistiche. La nostra prima ricerca è stata³: "un professore italiano di 35 anni che gioca ai videogiochi nella sua cameretta". Questa prima descrizione chiama in causa almeno quattro diversi volti latenti socialmente significativi: il volto di un professore, il volto di un italiano, il volto di un trentacinquenne e il volto di un videogiocatore. I risultati prodotti da questa prima descrizione, ovvero quattro immagini, sono già abbastanza interessanti:

- 3 su 4 hanno gli occhiali
- 4 su 4 hanno la barba
- 3 su 4 hanno dei capelli corti e uno su questi tre che ha un chiaro principio di calvizie
- 4 su 4 sono uomini (ricordiamo che il testo originale era in inglese e dunque senza indicazione di genere) di carnagione chiara

2 <https://openai.com/product/dall-e-2>

3 Traduciamo qui tutti i testi in italiano in quanto gli input devono essere attualmente in inglese

- 3 su 4 non hanno un viso magro pur senza essere in chiaro sovrappeso
- 3 su 4 hanno in naso relativamente grande
- 4 su 4 sono castani
- 4 su 4 non hanno tatuaggi o piercing di alcun tipo

Ricorrenza di volti a cui si associano poi anche altre ricorrenze non facciali come il fatto che tre su quattro giocassero seduti sul loro letto, che due camere da letto su quattro fossero molto disordinate, e che nessuno dei quattro fosse vestito un modo elegante o stravagante.

Nonostante molti tratti in comune, tuttavia, nessuna di queste quattro immagini rappresentava in modo verosimile il volto dell'autore che qui scrive. Un risultato mancato deludente ma al tempo stesso *rassicurante*. Abbiamo allora reiterato la produzione di queste immagini modificando volta per volta le descrizioni cercando di ottenere un risultato soddisfacente⁴. Abbiamo aggiunto parole connotate come “nerd”, caratteristiche fisiche come la bassezza, situazioni-locazioni diverse come una lezione universitaria, rimosso termini come “professore”, specificato che dovesse avere la barba, sostituito nerd” con “geek”, trasformato l'azione del giocare in una caratteristica inserendo il termine “gamer”, rimosso “nerd e geek”, aggiunto una descrizione emotiva come “felice”, cambiato la nazionalità da italiano a francese, specificato che dovesse avere gli occhiali, aggiunto che fosse celibe, detto che fosse in sovrappeso e poi magro, abbandonato del tutto l'area del gioco chiedendo che fosse rappresentato mentre mangiava la pizza, cambiato la sua emozione da felice a triste, alterato il vestiario in modo che indossasse una maglietta divertente e,

4 Avremmo potuto in alternativa generare nuove immagini senza cambiare le parole o chiedendo delle variazioni a partire da una di queste prime quattro immagini, ma per i fini di questo articolo era per noi più interessante rilevare il rapporto tra singolo termine e la sua connotazione visiva.

finalmente, inserito che dovesse avere uno specifico tipo di naso. Dopo sedici reiterazioni, producendo un totale di 64 immagini, abbiamo così finalmente raggiunto una immagine fotorealistica abbastanza somigliante da essere non solo significativa per chi l'ha generata ("sono io!") ma anche per altri "lettori" che, una volta caricata come foto profilo sui social, l'hanno o confusa come una foto effettiva oppure (quelli più intimi e generazionalmente abituati alle alterazioni digitali) percepita come una foto sulla quale fosse stato inserito un qualche *filtro*. Infine, per verifica abbiamo poi subito dopo inserito questa stessa frase tale e quale chiedendo questa volta una rappresentazione in forma di disegno e ottenuto nuovamente una immagine incredibilmente somigliante e anzi, in virtù della maggiore astrazione, ancora meno riconoscibile come "falsa". La frase specifica dalla quale si è generato questo volto latente è: "a 35 years old single and happy italian professor with beard, glasses and a bulbous nose playing videogames".

Questo piccolo esperimento è semioticamente interessante, e risponde dunque in parte alla nostra domanda iniziale, per almeno due diversi motivi.

Il primo è che queste ricorrenze fanno sembrare *vicine* delle unità culturali apparentemente non necessariamente correlate: esse danno una risposta abbastanza netta al quesito "cosa hanno oggi in comune un gamer, un docente, una persona di nazionalità italiana e un o una 35 enne?". Per un essere umano il quesito è quantomeno bizzarro poiché la molteplicità delle sue esperienze di queste tre tipologie di volti, e cioè della sua enciclopedia, è così variegata da non ammettere una ovvia risposta. Senza alcun dubbio un umano per rispondere non farebbe il tentativo di una media matematica della sua esperienza diretta (es: nucleo di amici coetanei) e indiretta (volti di attori) ma cercherebbe di pensare a *prototipi iconici* di ognuno di questi termini. Darebbe cioè un diver-

so *peso* a certi dati ritenendoli particolarmente rappresentativi. Farebbe questo cercando la quadra tra 1) i volti del suo cerchio sociale più vicino e che incontra più di frequente nel suo contesto sociale quotidiano e 2) quelli più al centro della sua semiosfera di riferimento e cioè con maggiore visibilità mediatica. Semplificando il quesito al solo genere maschile e alla contemporaneità, il risultato umano sarebbe un misto di conoscenze personali e contestuali con *personaggi* italiani iconici come Sergio Mattarella, Roberto Bolle, Khaby Lame, Ciccio Gamer, Giorgio Armani, Yotobi, Fedez, Amadeus, Favij, Pierfrancesco Favino, Carlo Cracco, Umberto Galimberti, Mario Balotelli, Massimo Recalcati e così via. Come è facile vedere da questo piccolo ma significativo campione, è piuttosto difficile per noi trovarvi delle regolarità così *forti*. Questa in qualche modo con-naturale perplessità epistemologica dinanzi alla diversità dell'esperienza umana e delle rappresentazioni a cui siamo esposti, dovuta non solo alla varietà delle manifestazioni facciali osservate ma anche alla coscienza di tutti i possibili volti divergenti e alternativi al nostro modello personale, è tuttavia assente dalla programmazione della macchina che 1) non ha che "esperienze" indirette e 2) non ha come mondo di riferimento che un insieme testuale chiuso senza eccezioni o esclusioni (il dataset).

Da un lato è certo possibile, e anzi necessario, riflettere criticamente sul fatto che le rappresentazioni e regolarità così create non possono essere ritenute del tutto epistemologicamente valide: esse non sono rappresentative del reale e nemmeno delle sue rappresentazioni ma solo di una parte delle sue rappresentazioni usate come input. Questo vale sia per il contenuto figurativo di questi materiali (risultati di selezioni estetiche e biopolitiche di natura storica) che per la natura visiva stessa del materiale in cui un fattore intrinsecamente situato e performativo come l'identità viene disincarnato e narcotizzato (Voto 2022). Dietro all'apparente oggettività di un processo macchinico di classificazione *plastica* si nasconde così

una retorica rappresentativa che occulta come i risultati ottenuti a partire da datasets non siano mai culturalmente neutri in quanto sono necessariamente situati storicamente e culturalmente, favorendo dunque alcune rappresentazioni al posto di altre sulla base dell'input opaco (Voto 2022). Input che per altro viene "letto" secondo criteri interpretativi non umani incapaci di distinguere l'oggetto dal contesto così che, per esempio, ciò che stava per "gamer" erano di fatto delle cuffie (che non sono affatto necessarie per giocare!) mentre "nerd" portava nell'immagine un vestiario dandy retrò e occhiali tondi che rimandano al nerd come caricatura del perdente e non alla rappresentazione degli appassionati "reali" di giochi e storie di fantasia. Si tratti di fenomeni ben noti e accomunabili alle I.A. che non sanno distinguere un peso dal braccio che lo solleva o che discriminano tra cane e lupo sulla base dello sfondo (Ribeiro *et al.*, 2016).

Dall' altro lato, tuttavia, non solo i risultati contemporanei del *machine learning* sono possibili solo grazie a grandissime quantità di dati (e dunque quella porzione di rappresentazione non è poi così piccola e culturalmente insignificante) ma è anche impossibile negare che rappresentazioni stereotipate come "professore=barba" abbiano un qualche senso socialmente fondato, un *sentore* di vero legato al nostro immaginario che li ammette facilmente sulla base di una esperienza del reale necessariamente parziale ma non per questo falsa nel senso ideologico del termine "discorso" (ovviamente senza per questo postulare la necessità ontologica di queste correlazioni). Lo spazio latente inteso come *modo di produzione* rappresenta cioè *potenzialmente* una verità rappresentazionale del mondo di tipo statistico che, sebbene sembri funzionare come uno stereotipo,⁵ in principio non ne è uno, dato che rileva in modo *accurato*

5 È di fatto per una deformazione etica dovuta alla nostra sensibilità contemporanea che vediamo in questo *processo* a priori uno stereotipo, ma per definizione lo stereotipo può essere tale solo se *fallibile*.

e *puntuale* delle occorrenze sulla base di un dataset situato e costituito storicamente dalla catena dei segni della cultura proprio come lo è necessariamente il Soggetto (Eco 1984 in Paolucci 2010).

In altre parole, le ricorrenze del volto latente non sono rappresentative del reale, e cioè del mondo vero in tutta la sua complessità, ma sono invece rappresentative della realtà e cioè del mondo in cui gli esseri umani interpretano il reale cercando di dargli un senso attraverso certi discorsi sociali e testi che vengono rivelati dalle I.A. come regolarità produttive.

Il secondo motivo per cui questo esperimento è interessante riguarda invece il fatto che tutte queste caratteristiche sono effettivamente esatte rispetto al volto del qui presente autore: la più banale e stereotipata rappresentazione di quella descrizione è, casualmente ma effettivamente, una realtà. Tuttavia, al tempo stesso, nonostante l'innegabile fattualità di queste caratteristiche nessuno di quei primi quattro volti rappresentava il qui presente autore in maniera anche solo approssimativamente accurata. Per riuscire a ottenere un volto vagamente somigliante sono state necessarie decine di reiterazioni e specificazioni: da cui il senso di *assicurazione* dinanzi alla propria banalità simbolica ma non iconica. Un fenomeno, questo dell'apparente irriducibilità linguistica del volto, che caratterizza la lettura e scrittura di opere letterarie (Magli 2016). Vediamo dunque qui una delle caratteristiche più semioticamente rilevanti del volto umano: il suo possedere una complessità multidimensionale tale da stare per analogia per l'unicità identitaria ideologicamente irriducibile dell'individuo. Inoltre, ancora più importante, la somiglianza del volto finale davanti a cui si è manifestato il pensiero "sono io!" non era affatto di tipo matematica ma aveva invece a che vedere col modo in cui una certa ibridazione di tratti simboli ed iconici aveva la capacità di *evocare attraverso un volto diverso una stessa identità* latente. In questo senso, vediamo come la nozione di latenza sia metaforicamente applicabile anche al

volto umano dal punto di vista della sua generazione (il volto di ogni neonato è una particolare *attualizzazione variabile* di possibilità virtuali costrette da fattori genetici e cioè di una ricorrenza iscritta e a noi nascosta), della sua trasformazione (ogni volto adulto umano è al tempo stesso una immagine già trasformata da cause bio-culturali e una immagine sempre ulteriormente trasformabile), della sua interpretazione (ogni volto *altro* davanti a noi richiama in noi il virtuale-ricorrente attualizzato da cui è stato generato e al tempo stesso non smette mai di mantenere qualcosa di nascosto oltre l'evidenza della sua simbolicità e iconicità) e persino della sua identificazione (ogni volto proprio *assume il senso/sentore* di proprio non solo per continuità/coerenza fisiognomico-morfologica col volto passato e "purezza" ma riconoscendo o sconoscendo⁶ in quei tratti una *storia semiotica di significati sia consolidati e socialmente condivisi che mai realizzati e individuali*). Ecco perché, in virtù della sua multidimensionalità irriducibile alla sua espressione, un volto senza nulla di nascosto (e da nascondere) e senza un potenziale ancora inespresso (completamente dato) è sostanzialmente il volto di una non-persona.

Da tutti e due i punti di vista è facile vedere come *l'enjeu* di una riflessione umanista sulla generazione ludica di volti-sé tramite uno spazio latente sia innanzitutto etico: lo spazio latente è una *topologia invisibile che identifica e categorizza* secondo logiche macchiniche dei testi-documenti storici prodotti da degli umani e che a partire da queste *procedure produce rappresentazioni del mondo potenzialmente interpretabili da umani come verosimili* e persino *rappresentative* del reale. Rappresentazioni, classificazioni e identificazioni che, nel caso del volto, diventano determinazioni artificiali di aspetti identitari che hanno effetti di veridizione in quanto enunciativamente prodotte da un discorso che non solo

6 Come accade nel romanzo "The Face of Another" (Abe 1964).

non esplicita la sua situazione storico-enciclopedico (in altre parole che il suo punto di partenza è già compromesso e contaminato) ma che adotta molti tratti del discorso tipo scientifico: impersonalità dell’agente, osservazione reiterata, verità statistica e controllabile del risultato.

Da un punto di vista etico, così descritte, è facile ritenere le produzioni di immagini tramite algoritmi come quello di DALL-E siano non solo mendaci ma anche in qualche modo preoccupanti: lo spazio latente *nasconde* dietro la ricorrenza di un volto tutte le sue possibili ed anzi probabili divergenze. Un tema specialmente rilevante nel contesto politico attuale in cui da molti discorsi emerge una chiara insofferenza per le generalizzazioni e categorizzazioni opponendo spesso a queste la rivendicazione dell’autodeterminazione come unica forma di determinazione eticamente accettabile ed epistemologicamente valida.

In un tale contesto, non sorprende infatti che diversi articoli giornalistici abbiano esplicitamente accusato i sistemi di rappresentazione latente come DALL-E di diffondere stereotipi e di essere socialmente pericolosi⁷: troppe persone, secondo queste critiche, sono escluse dalla stessa possibilità di dire “sono io!” (proprio come è accaduto a noi) o addirittura “potrei essere io!”.

Andando verso le conclusioni del nostro breve saggio, vorremmo tuttavia qui intervenire a titolo critico contro questo tipo di atteggiamento censorio e ragionamento unilateralmente polarizzato. Vorremmo cioè riconoscere a tutti gli effetti gli evidenti stereotipi che formando lo spazio latente (tramite il dataset) fondano le immagini facciali (bias ammessi dagli stessi creatori di DALL-E) ma, attraverso una riflessione semiotica sul senso identitario dei volti così generati e sull’uso ludico che ne facciamo, opporre alla presunta pericolosità di queste rappresentazioni il loro grande valore conoscitivo e immaginativo.

7 <https://www.vox.com/future-perfect/23023538/ai-dalle-2-openai-bias-gpt-3-incentives>

Intendiamo opporci a queste critiche innanzitutto *a priori* ricordando la necessità cognitiva umana di usare categorizzazioni ed euristiche. Dal punto di vista delle scienze della mente, è infatti indiscutibile che esse siano necessarie per agire efficacemente nel mondo: la mente umana è perennemente coinvolta in processi di classificazione del “simile” mirati ad *automatizzarne* il processo di riconoscimento e a velocizzare (Kahneman 2011) la presa di decisioni sulla base di giudizi intuitivi (Gilovich et al., 2002). Processi decisionali che quando applicati al volto determinano il potere delle prime impressioni sulla nostra immagine della persona di fronte a noi (Todorov 2017). Ma l’uso di “tipi” è anche altrettanto riconosciuto dalle scienze della cultura, dei linguaggi e delle interpretazioni come fondamentale per la comprensione interpersonale. Possiamo trovare infatti questi fenomeni descritti sia all’interno dell’epistemologia linguistico-strutturalista della semiotica fondata da Greimas e Courtés (1979) che all’interno dell’epistemologia logico-culturale della semiotica Echiana fondata sul lavoro di Peirce (1990). Nel primo caso il concetto chiave è quello di *figuratività*, nel secondo è quello di *selezione enciclopedica*. In entrambi i casi, la possibilità di dare senso al mondo e alle sue rappresentazioni/narrazioni dipende da una capacità e predisposizione alla generalizzazione inscritta poi nelle culture come unità culturali di una enciclopedia proprio sulla base della loro ricorrenza. Da questo punto di vista, vi è indubbiamente qualcosa di molto (troppo?) umano nel modo in cui una data rete neurale viene allenata ed è paradossalmente una critica non alla macchina ma all’uomo quella che rivolgiamo a DALL-E accusandolo di non saper immaginare un gamer senza cuffie o il volto, oggi assai comune, di un italiano con tratti fortemente asiatici. Che lo spazio latente categorizzi i volti per differenza di *tipi* secondo criteri storicamente costituiti e che potrebbero non piacere a chi si ritrova incluso o escluso dalla categoria, insomma, non può essere un problema *a*

priori perché altrimenti ogni rappresentazione diventa *a priori* problematica. Laddove DALL-E e l'umano divergono, tuttavia e per fortuna, è nell'autocoscienza critica della natura pragmatica di queste produzioni che implicano per definizione, dunque, a loro volta sempre la possibilità, se non la necessità, di concepire altrettanto pragmaticamente la loro *potenziale falsità*. Ancora una volta qui le critiche a DALL-E non sono critiche all'intelligenza artificiale: il rischio sociale di rappresentazioni stereotipate non sta nelle rappresentazioni e nel loro modo di generarsi ma nei suoi lettori a-critici e ineducati che, di certo, non sono il prodotto delle I.A. degli ultimi dieci anni. Infine, va anche sottolineato come quando tali I.A. applicano questi stessi processi ad oggetti non-facciali, e cioè non così fortemente identitari e politici, le comunità umane sembrano molto meno inclini a vedere un problema in questo modo stereotipato di ri-enunciare la testualità del mondo presa a campione. Ma concentrandoci in particolare sul volto e sulla sua relazione con le identità personali, vi è una questione filosofica ancora più profonda che fa sì che accusando i volti di DALL-E si sta di fatto accusando l'intelligenza collettiva delle comunità umane contemporanee. Questa questione riguarda il fatto che se è vero che le categorizzazioni e interpretazioni generalizzanti possono essere lette come un atto di potere e di oggettificazione dei singoli del mondo, altrettanto vero è però che non può esservi vita sociale e comprensione del mondo senza di essi. Come postulato dal Sartre del *l'Essere e il Nulla* (1958) e in continuità con le prospettive fenomenologiche (Rosato 2014), l'oggettificazione dell'altro comincia con lo sguardo fra due esseri umani che si sminuiscono a vicenda per potersi cogliere (Pertseva 2017) e si fonda precisamente sull'opacità e conflittualità del rapporto tra *Io* e *alterità*. Così come noi non possiamo conoscere e agire senza tipizzare, allo stesso modo noi non possiamo interagire con l'altro se non oggettificandolo in un primo momento: ignorandone la latenza virtuale per

identificarlo col suo esito attuale. È proprio per questo che il compito educativo fondamentale di ogni società è precisamente di dare ai loro membri gli strumenti di saper concepire l'eccezione oltre il tipo e il Soggetto oltre l'oggetto: in altre parole di saper mettere in crisi ogni discorso ed *in primis* il proprio. In questo senso, gli intellettuali che si occupano oggi del pensiero umano hanno lo scomodissimo e ingrato compito di difendere la legittimità cognitiva della semplificazione generalizzante e la legittimità culturale del discorso (unico modo tramite cui possiamo dare senso al mondo in modo condiviso) per evitare che, ingenuamente e senza accorgersene, non si faccia altro che sostituire una presunta rappresentazione ideologica con un'altra. Proprio chi studia i linguaggi e la comunicazione, in particolare, ha a disposizione tutte le argomentazioni scientifiche per confutare l'idea di un rapporto 1:1 fra l'esposizione alla necessaria semplificazione e prospettività di ogni sistema di rappresentazione (come DALL-E) e i gravissimi fatti politico-sociali di ingiustizia, discriminazione e violenza verso la diversità compiuti dai Soggetti del reale. Questo abuso di potere è infatti compiuto da soggetti incapaci di pensare la complessità e la possibilità per cause, anzi colpe, che è inaccettabile attribuire a Shutterstock anziché alle famiglie e ai governi degli Stati. Allo stesso modo, l'incapacità di alcuni soggetti di riconoscere le immagini in quanto tali, nonostante le tendenze visivo-estetiche contemporanee mirino a questo effetto di ambiguità (Pinotti 2021), non può essere colpa delle immagini stesse. Un conto è dunque fare battaglia a certe rappresentazioni errate del reale o a delle tendenze stereotipanti di certi canoni, un altro è voler, nei fatti, occultare ogni singola rappresentazione che non sia in linea con alcuni valori contemporanei tacciandola come violenza o discriminazione attraverso la retorica del "pericolo". Questo da un lato perché i nostri valori contemporanei sono, in quanto tali, storicamente situati e costruiti e dunque *in principio* non meno cri-

ticabili dei valori passati a cui essi si oppongono. Dall'altro perché solo attraverso un presupposto ideologico si può negare il mondo come spazio costituito, sia fattualmente che cognitivamente, da occorrenze maggioritarie e minoritarie o si può affermare che ogni rappresentazione "maggioritaria" sia a priori discriminatoria e pericolosa. Posto anche che il mondo sia fatto solo da entità così complesse e uniche che ogni loro riduzione e riconduzione ad una "parte" rappresenta una violenza, noi non abbiamo alcun accesso non mediato alla verità e all'intimità di questa complessità e unicità dell'altro. Motivo per cui l'unica alternativa alla rappresentazione è un mondo fatto di monadi che si guardano in un eterno silenzio. Una immagine, questa del silenzio, che sembra talvolta essere il triste destino di chi, per i più nobili motivi, insegue ingenuamente ideali di "rappresentazioni assolute". Essi, infatti, sembrano proprio finire spesso con l'imporre nuove categorie (es: cis gender), nuove ideologie (il ruolo delle figure femminili nei testi di pop-culture contemporanei è prevedibile esattamente come lo era quello del protagonista maschile negli anni '90) e di coniare termini *irrappresentabili*: dalla "schwa" al nome del movimento "LGBTQIA+" che, nel doppio tentativo di rimuovere le marcature storiche da un lato e di includere tutte le occorrenze dall'altro, finiscono entrambe per essere *non pronunciate*. Prendiamo qui l'esempio del femminismo non per criticarne obiettivi e principi, che anzi condividiamo, ma come esempio per eccellenza di una mancata prospettiva pragmatica sulla natura locale-cooperativa della significazione e di conseguenza sulla latenza attuata come fenomeno rivelatore anziché mendace. È infatti proprio in quanto manifestazione dei discorsi non culturalmente neutri che le rappresentazioni facciali di DALL-E sono epistemologicamente valide in termini di rappresentazione umana del mondo (realtà): è solo attraverso i volti presenti-maggioritari che possiamo dare significato e riflettere sui volti *assenti* che sono a tutti gli effetti par-

te dell'output. Facciamo un esempio concreto: nell'articolo prima menzionato sui pericoli di DALL-E, il primo esempio di stereotipo rappresentato riguardava la ricorrenza di volti maschili e bianchi scrivendo il termine "avvocato", ebbene perché indignarsi di questa rappresentazione e di che diritto definirla come "mendace" quando ancora nel 2020 una ricerca demografica sulla composizione delle corti giudiziarie faceva emergere proprio il dato di una giustizia "maschile e bianca" 8? In questo senso i volti prodotti tramite sistemi matematici di latenza sono per noi invece preziosi in quanto rivelatori, e non essere rappresentati da questi sistemi può essere un torto tanto quanto esservi presenti! Anziché una distopica mossa di *resa* nei confronti dei Soggetti che diamo per scontato come incapaci di un pensiero critico, serve allora oggi arricchire la semiosfera contemporanea con la produzione di nuovi volti latenti a partire da diversi dataset, vedendo non nella censura delle rappresentazioni ma nella loro molteplicità il nostro più sicuro scudo contro un mondo ideologico (Giuliana 2023).

Vorremo, infine, difendere a *posteriori* le immagini prodotte da DALL-E ritornando sulla dimensione ludica delle pratiche di generazioni di immagini tramite I.A. e in particolare di referenti del mondo reale. In che senso noi *giociamo* con DALL-E? Ebbene lo facciamo secondo una delle accezioni più consolidate e attuali di gioco come messa alla prova del reale, ovvero come *paidia* (Caillois 2000) e *risemantizzazione* (Thibault 2020). Una reinvenzione del mondo che è molto più di un carnevalesco breve momento di sovvertimento e che invece diventa strumento di conoscenza e rivelazione destinato a perdurare dopo la fine del gioco. Da questo punto di vista ognuno dei 64 volti generati nel nostro esperimento ha rivelato qualcosa di preziosissimo: un altro-se' e una alterità intima

8 <https://www.americanprogress.org/article/examining-demographic-compositions-u-s-circuit-district-courts/>

palesatesi sulla intrinseca latenza del volto (esposta qui sopra) e fondate sulla profonda natura intersoggettiva dell'identità. Ogni gioco è infatti una esperienza cosciente di latenza, ovvero un giocare alterando e combinando le possibilità di un dato sistema (non solo di regole ma anche, come in questo caso, storico) entro un quadro di ricorrenze al fine di vederne gli esiti. Una messa alla prova che, in quanto compiuta nel gioco direttamente dal lettore empirico, ha il potere di rivelare il molteplice nel singolo e la natura *viva* della latenza della nostra identità. Ecco come guardando alla pratica ludica attorno a DALL-E possiamo assumere una posizione critica rispetto a una prospettiva identitaria esistenzialista secondo cui il fondamento dell'identità ha come condizione la sua irriducibilità ad aspetti biologicamente, socialmente o culturalmente codificati e cioè intersoggettivi. Difendiamo al contrario l'idea che la parte più vera e più simile di noi rappresentata in quelle rappresentazioni di volti di DALL-E sia precisamente quella *differenza* col nostro volto fattuale. Una differenza la cui possibilità di manifestazione dipende dalla produzione segnica di unità di uno spazio latente costituito come luogo di enunciati passati che sono costitutivi dello spazio attuale del reale. Ogni volto *non-nostro* prima integrato nello spazio latente e poi ri-prodotto da DALL-E ci ricorda non solo delle possibilità (passate, presenti e future) che pulsano dentro di noi e ci definiscono ma anche delle molteplici persone che costituiscono intersoggettivamente la nostra persona (Paolucci 2020). In questo senso, lo sconforto per la banalità del volto del sottoscritto riprodotto da DALL-E è stato accompagnato dalla gioia di riconoscere in quel volto tondo, occhialuto, bianco, castano e barbuto una somiglianza con una famosa vignetta di Umberto Eco. In termini forse meno astratti, proprio DALL-E come gioco fa emergere quanto legare la propria identità (e di conseguenza la sua rappresentazione) all'apparenza somigliante del volto (che nelle critiche è sempre la parte incriminata) con sé stessi anziché con altri sia assurdo in un duplice

senso: proprio per causa della latenza del volto noi siamo noi sia in quanto simili ad altri che in quanto inevitabilmente dissimili dal nostro stesso volto accomunabile a milioni di non-io. Una conclusione questa che, lungi dall'essere una pura farneticazione filosofica lontana dalla realtà sociale, sembra esistere come coscienza sociale proprio in quei testi ludici come la serie di meme "literally me"⁹ e ancora più in particolare nella produzione di questo meme che usano il caso della sirenetta di colore¹⁰ proprio per deridere un tale idea di rappresentatività degli individui del reale laddove l'esperienza ludica-immaginativa insegna a *essere come molteplicità*. Per questo le critiche alle rappresentazioni di DALL-E non possono essere ritenute delle valide critiche alla sua effettiva rappresentatività (e cioè alla dimensione interpretativa) a meno di non assumere, paradossalmente, posizioni discriminatorie (es: il colore della pelle o il sesso come parte fondante dell'individuo e prevalente rispetto a tutto il resto della *persona*).

Concludendo, sia come *processo* che come *esito* ci sembra di poter affermare che i volti latenti, tanto come prodotti dalle più recenti tecnologie quanto come prodotti dalle pratiche ludiche nate attorno ad essi, siano oggetti estremamente preziosi per le scienze sociali e filosofiche che, dallo studio della loro generazione a quello delle loro interpretazioni, attraverso le intelligenze artificiali possono osservare lo specifico semiotico del modo *tutto umano* di costruire e dare senso alla nostra realtà sociale.

9 <https://knowyourmeme.com/memes/literally-me-syndrome-wow-this-is-literally-me>

10 https://www.reddit.com/r/memes/comments/xy526g/thats_literally_me/

Bibliografia

Caillois R. (2000). *I giochi e gli uomini. La maschera e la vertigine*. Bompiani, Milano.

Courtés J. e Greimas A.J. (1979) *Sémiotique. Dictionnaire Raisonné De la Théorie Du Langage*. Hachette, Paris.

Eco U. (1990) *I limiti dell'interpretazione*. Bompiani, Milano.

Gilovich T., Griffin D., e Kahneman D. (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge.

Giuliana G.T. (2023, in stampa) “Retorica dello sforzo e ideologia nei giochi digitali”. *Lexia* 43-44.

Kahneman D. (2011) *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New-York.

Le Cun Y. (2019) *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob, Paris.

Magli P. (2016) *Il volto raccontato*. Raffaello Cortina, Milano.

Paolucci C. (2010) *Strutturalismo e interpretazione*. Bompiani, Milano.

Paolucci C. (2020) *Persona. Soggettività nel linguaggio e semiotica dell'enunciazione*. Bompiani, Milano.

Pinotti A. (2021). *Alla soglia dell'immagine*. Einaudi, Torino.

Pertseva A. (2017) « Le regard de l'autre chez Sartre : L'entre-deux de l'imagination et de la perception », *Bulletin d'Analyse Phénoménologique* [En ligne], 13 (2), URL : <https://popups.uliege.be/1782-2041/index.php?id=948>. DOI: 10.25518/1782-2041.948

Ribeiro M., Singh S., e Guestrin C. (2016). ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier.” In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, San Diego, Association for Computational Linguistics, pp. 97–101.

Rosato J. (2014). “Sartre and Levinas as Phenomenologists”. *Philosophy Today* 58 (3):467-475.

Sartre J.P. (1958) *L'Essere e il Nulla*. Mondadori, Milano.

Thibault M. (2020) *Ludosemiotica. Il gioco tra segni, tesi, pratiche e discorsi*. Aracne, Roma.

Todorov A. (2017) *Face Value: The Irresistible Influence of First Impressions*. Princeton University Press, Princeton NJ.

Voto C. (2022) “From archive to dataset. Visualizing the latency of facial big data*”, *Punctum* 8(1):47-62.