DOPPIOZERO

Spazio latente, algoritmi e fake photography

Mauro Zanchi

28 Agosto 2024

A volte è necessario dare molta importanza all'etimologia, per cercare di comprendere più in profondità il senso di una questione complessa. L'etimo "postfotografia" ha origine negli anni Ottanta del Novecento ed è mutuato dalla nozione di postmodernismo, utilizzato per descrivere la profonda trasformazione della fotografia, innescata dalla rivoluzione digitale. Nel corso dei decenni, questa trasformazione ha portato a un radicale cambiamento nel modo in cui produciamo, distribuiamo e consumiamo immagini. Accanto alla fotografia ha preso corpo il "fotografico" e il "metafotografico", ovvero una serie di ricerche estensive, dove gli artisti hanno ibridato il medium con altre discipline e media e aperto questioni di matrice concettuale. La fotografia, un tempo legata a processi chimici e a supporti fisici, è diventata successivamente un flusso continuo di dati digitali, facilmente manipolabile e riproducibile. In *La furia de las imágenes* (2016; trd. it. *La furia delle immagini*, Einaudi 2018: qui la recensione di Elio Grazioli), Joan Fontcuberta connota la post-fotografia da un lato al riuso di fotografie già scattate, secondo un reimpiego in campo artistico, come per esempio alle opere degli artisti Erik Kessels, Joachim Schmid e Penelope Umbrico, e dall'altro a quelle immagini provenienti dalla seconda rivoluzione digitale, dall'era dei social network, del virtuale e di internet, dove giornalmente assistiamo allo spettacolo della pervasività, della moltiplicazione e dell'onnipresenza delle figurazioni, che fluttuano nell'iconosfera o nell'infosfera. Tutte queste immagini che viaggiano in rete sono state realizzate con macchine fotografiche digitali, poi ridotte in immagini JPEG, oppure scansionate da libri o riviste. Sono immagini che sottostanno ancora all'etimo "fotografia", ovvero riprese ottenute per mezzo della scrittura con la luce.



Alessandro Sambini, Human Image Recognition, Ansel - Archetipo 2 (settembre 2022), dipinto su NVIDIA Canvas, stampa colore, lettura HIR, fronte e retro. Courtesy Galleria Indice

Nell'era analogica la luce impressionava una lastra o una carta fotosensibile mentre nell'era digitale le presenze del mondo vengono memorizzate mediante dispositivi in grado di fissare immagini attraverso vari passaggi che portano al pixel. Sebbene secondo quest'ottica le immagini fotografiche che scorrono nei dispositivi tecnologici attuali abbiano cambiato sia il loro statuto sia la funzione, possiamo collocarle ancora nella cerchia di ciò che deve la loro esistenza all'azione della luce. L'era della postfotografia reale inizia quando un'immagine non viene più realizzata per mezzo della luce, e si va al di là di quella caratteristica che prevede la registrazione indicale di qualcosa che esiste nella realtà. Quindi solo da qui in avanti si può utilizzare il termine "postfotografia", ora che le AI text-to-image sono in grado di rendere visibili figurazioni molto realistiche (proprio come fossero fotografie credibili) attraverso parole, prompt, previsioni statistiche, vettori, oppure con disegni a macchie colorate, come per esempio con NVIDIA Canvas (si veda: Alessandro Sambini, *Human Image Recognition, Ansel – Archetipo 2*, Settembre 2022). Nel nostro tempo storico non è più necessario attingere qualcosa solo da dataset e datacenter, dove sono presenti miliardi di immagini fotografiche. L'era della postfotografia comincia solamente con l'avvento delle tecnologie come Midjourney, Stable Diffusion, Dall-E 2 o altri sistemi "text-to-image". Tutto quello che è stato realizzato prima è riuso concettuale di fotografie già scattate precedentemente da altri, è una serie di aperture estensive del medium, è ricerca metafotografica, dove il prefisso meta- riguarda le varie accezioni e declinazioni dei metalinguaggi, sono le immagini veicolate e rielaborate dal metamedium (primo nome dato al computer), sono le immagini contenenti metadati – che ora sono il nuovo petrolio, la risorsa più preziosa al mondo –, e tutto quello che circola in rete attraverso i social e le

applicazioni di tipo messaggistico di proprietà META (Whatsapp, Instagram e Facebook) e tramite Tik Tok. Appartengono agli approcci meta- anche gli algoritmi di correzione dell'immagine, il deep web, l'apertura al non visuale, la codificazione con stringhe di numeri, l'archivio reinterpretato, le corruzioni e gli sviluppi dell'inconscio tecnologico, l'utilizzo delle telecamere di sorveglianza e dello scanner invece di un obiettivo, l'immagine generata da un software, l'immagine soft, digitale, computazionale, operativa, non umana, in rete. Non appartengono a qualcosa che pare solido e finito, bensì a dimensioni mutevoli, legate a processi che si intersecano con le rimediazioni tecnologiche.



Trevor Paglen, It Began as a Military Experiment, 2017, dettaglio, Set of 10 pigment prints; Courtesy l'artista.

All'interno dei recenti software di intelligenza artificiale, immagini e testi vengono trasformati in complessi e velocissimi calcoli matematici attraverso un numero inimmaginabile di parametri, rappresentati da vettori in uno spazio multidimensionale. Questo “spazio latente” è invisibile e contiene un'infinità di possibili figurazioni. Ma in una immaginazione artificiale non ci sono immagini, solo probabilità vettorizzate. Generare un'immagine equivale a selezionare un

punto specifico all'interno di questa ulteriore dimensione, un processo guidato da probabilità e statistiche. Questa nuova modalità di creazione si discosta radicalmente dai metodi tradizionali, poiché l'immagine generata non è una semplice copia di un pezzo di mondo o un indice dell'esistente nella realtà, ma il risultato di complesse operazioni matematiche su dati astratti. L'intelligenza artificiale non "vede" le immagini nello stesso modo in cui lo facciamo noi umani, ma opera su rappresentazioni matematiche di esse, aprendo la porta a nuove forme di fruizione e a una concezione radicalmente diversa dell'immagine. Potremmo anche chiederci se esista un inconscio algoritmico e come agisca sulla nostra coscienza quando noi ci relazioniamo con le intelligenze artificiali. In questa relazione quali sono i riferimenti tra input, output e dataset?

I modelli di intelligenza artificiale operano all'interno di una sorta di scatola nera, uno spazio matematico multidimensionale dove le immagini e i testi vengono rappresentati da vettori. Le AI parlano una lingua completamente diversa dalla nostra, fatta di numeri e complessi algoritmi. Quando generiamo un'immagine partendo da comandi scritti con la tastiera del PC all'interno di Midjourney stiamo in realtà chiedendo alla macchina di "tradurre" una presenza matematica in una rappresentazione visiva. Questo processo è così complesso che risulta difficile comprendere o prevedere esattamente come l'intelligenza artificiale arrivi a creare un'immagine specifica. Di conseguenza, le immagini generate non possono essere considerate semplici copie o rielaborazioni di immagini esistenti o di fotografie JPEG caricate in rete e nei dataset, ma piuttosto il risultato di un processo completamente nuovo e diverso. Ogni immagine digitale, all'interno di un software, sia esso di intelligenza artificiale o meno, è in realtà una materia costituita da sequenze di numeri 0 e 1. Un computer non ha una percezione visiva e tratta le immagini come dati numerici. Il processo di generazione di un'immagine da parte di un'AI è quindi una sorta di traduzione da un linguaggio all'altro: dal linguaggio dei pixel a quello dei vettori numerici e di nuovo al linguaggio dei pixel. Le AI manipolano vettori numerici, sottoponendoli anche a complessi calcoli statistici. Il prompt, che fornisce all'AI delle indicazioni su quale tipo di immagine generare, può essere visto come una sorta di traduttore, che aiuta a guidare il processo di creazione. Nel caso dei testi, la rappresentazione numerica delle parole (*embedding*, che è un concetto determinante per il funzionamento dei modelli di *deep learning*) è ancora più complessa, in quanto implica una serie di assunzioni profonde sulla natura e struttura del linguaggio e del significato.



Ho Tzu Nyen, The Critical Dictionary of Southeast Asia (2017~present), Courtesy l'artista.

Quando dico o scrivo sulla tastiera "Sophia cerca di comprendere il funzionamento delle AI text-to-image", la parola "Sophia" si riferisce a una ragazza o donna chiamata Sophia in virtù di una catena causale complessa e di una tradizione del linguaggio verbale, che intendono il riferimento a "Sophia" come se fosse qualcuno che io conosco, o frequento, o a un personaggio di fantasia o a un riferimento alla Sapienza, secondo il pensiero greco antico. Non sappiamo a cosa si riferisca il termine "Sophia" quando è prodotto da un LLM, nonostante i Large Language Model rappresentino una forma evoluta dell'intelligenza artificiale, specializzati nella "comprensione profonda" (ammesso che siano in grado di comprendere secondo l'accezione umana) e nella generazione di testi. Si può argomentare che la teoria canonica del riferimento a qualcosa o a qualcuno crei a prima vista un problema per l'attribuzione di una semantica agli LLM. Un problema analogo si crea se si assume qualche collegamento causale tra il contenuto o la rappresentazione di un'immagine e l'oggetto rappresentato da quell'immagine (tra un computer e la foto di quello stesso computer, o meglio, il fatto che quella foto specifica rappresenti il computer). A differenza dei tradizionali algoritmi di apprendimento automatico, gli LLM sembrano in grado di cogliere le sfumature e le complessità del linguaggio umano, grazie a un addestramento su vasti corpora di testi. Questa capacità li rende strumenti preziosi per una miriade di applicazioni, dalla traduzione automatica alla generazione di contenuti creativi. L'avvento di modelli come GPT-3.5, ChatGPT e BARD ha ulteriormente potenziato le capacità delle macchine di dialogare in modo naturale e di fornire risposte informative a un'ampia gamma di richieste. Il Natural Language Processing (NLP) ha fatto passi da gigante grazie

allo sviluppo dei Large Language Model. Questi modelli, alimentati da enormi quantità di dati testuali, hanno rivoluzionato la capacità delle macchine di comprendere e generare linguaggio naturale. Grazie agli LLM, le interazioni uomo-macchina sono diventate sempre più fluide e naturali. Le applicazioni sono molteplici e in continua espansione, dalla generazione di testi creativi alla risposta a domande complesse. L'evoluzione degli LLM promette di portare a un futuro in cui le macchine e gli androidi saranno in grado di comprendere e interagire con gli esseri umani in modo ancora più profondo e sofisticato. Quando si afferma che un'immagine sintetica assomiglia a un'immagine reale si presuppone una comparazione oggettiva tra pixel, ignorando il complesso intreccio tra l'opera d'arte, la cultura di riferimento e lo spettatore. In realtà, la percezione di somiglianza è un'esperienza soggettiva, influenzata da un bagaglio culturale e da una serie di associazioni mentali. Un'immagine sintetica non è semplicemente una copia di un'immagine reale, ma il risultato di un processo complesso che coinvolge algoritmi, dati e interpretazione umana. È un errore trattare le immagini generate da AI come se fossero oggetti statici, trascurando il loro carattere relazionale. Per valutare un'immagine generata, è necessario andare oltre la semplice comparazione visiva e considerare il contesto culturale, le intenzioni dell'artista e le implicazioni etiche. Ovviamente questo valeva anche per le fotografie e per tutte le immagini in generale. Ora però sono messe in azione altre connotazioni e diversi effetti collaterali, di cui non siamo ancora a conoscenza e che forse comprenderemo più in là nel tempo.



Alper Yesiltas, Jimi Hendrix, 2022, dalla serie As if Nothing Happened, Courtesy l'artista.

Anziché riprodurre fedelmente realtà esistenti, le AI generano immagini originali esplorando un vasto “spazio latente”, un universo di possibilità visive nascoste. Si è passati dalla copia all'esplorazione. Tradizionalmente, immagini fotografiche, pittoriche, scultoree, video, erano copie della realtà. Con l'AI questo paradigma cambia radicalmente. Come abbiamo già detto sopra, le immagini non sono più semplici riproduzioni, ma il risultato di complessi calcoli matematici applicati a un enorme insieme di dati. Lo spazio latente è un concetto fondamentale nell'intelligenza artificiale generativa, un universo di innumerevoli possibilità**,** una dimensione dove le regole della realtà fisica non valgono più in toto. Esso offre una rappresentazione compatta e flessibile dei dati, permettendo di manipolare e combinare informazioni in modi nuovi e creativi. Le sue implicazioni sono vaste, dalla generazione di contenuti artistici all'ottimizzazione di processi industriali. Immaginiamo di comprimere un'enorme enciclopedia in un singolo punto, che pur essendo minuscolo contiene tutte le informazioni dell'intera enciclopedia. Nello

stesso modo, lo spazio latente è una rappresentazione compatta e astratta di un vasto insieme di dati, come immagini, testi, suoni, e altro. È uno spazio matematico multidimensionale, dove ogni punto corrisponde a una combinazione specifica di caratteristiche. Le reti neurali, gli strumenti matematici che alimentano l'intelligenza artificiale, sono in grado di imparare a mappare i dati da uno spazio visibile (per esempio un'immagine) a uno spazio latente. Questo processo è chiamato *encoding*. Una volta nello spazio latente, i dati possono essere manipolati e combinati in modi che sarebbero difficili o impossibili nello spazio iniziale. Si possono sommare due vettori per creare una nuova rappresentazione, che, decodificata, corrisponderà a un nuovo dato. Manipolando i vettori, è possibile generare nuovi dati che non esistono nel dataset di partenza. Lo spazio latente permette di esplorare combinazioni di caratteristiche che potrebbero non essere state considerate in precedenza, aprendo la strada a nuove idee e inediti percorsi combinatori. Questa dimensione è destinata a diventare un concetto chiave per comprendere il futuro della visione e della cultura. Ci permetterà di esplorare nuove forme di espressione artistica e di interagire con la tecnologia in modi sempre più innovativi. Un'immagine generata dall'AI è complessa e distribuita, è il risultato di un processo collettivo, che coinvolge sia gli sviluppatori delle intelligenze artificiali sia i dati utilizzati per l'addestramento. Le promptografie generate con AI text-to-image non sono false immagini, ma figurazioni “altre”, appartenenti a una nuova forma di realismo, che potremmo chiamare “realismo macchinico”, una presenza che non cerca di imitare la realtà, ma di esplorarne le numerose possibilità in potenza. È un realismo che ci invita a ripensare il nostro rapporto con l'immagine e con il suo utilizzo. La sensazione di aver già visto un'immagine generata dall'AI è comune. Questo fenomeno, noto come "déjà-vu digitale", è dovuto alla saturazione mediatica e alla ripetizione di motivi visivi. Tuttavia, le immagini generate dall'AI vanno oltre la semplice ripetizione. Esse offrono una nuova prospettiva sulla realtà, una prospettiva che ci costringe a riconsiderare le nostre nozioni di approcci creativi. Molte sintografie sono immagini estremamente realistiche, al punto che l'occhio umano potrebbe non riuscire a individuare le differenze rispetto a fotografie documentarie.



Joan Fontcuberta, OYM #2 (Alex DeLarge, Camden Market + Angiospermes gracilis, Panamá), 2024; courtesy l'artista. Una delle due immagini è una fotografia, l'altra è stata ottenuta con Midjourney

Joan Fontcuberta ha sempre avuto l’abilità di prevedere, prima di altri artisti e critici, questioni che sarebbero accadute in anni successivi, nel fluire cronologico della storia. Da buon profeta visionario aveva probabilmente già considerato il termine “postfotografia” legato alle immagini algoritmiche, in un periodo in cui la società non era ancora pronta per comprendere il salto o sopportare questo cambio repentino. Ha preparato l’arrivo delle immagini algoritmiche lavorando prima con opere di avvicinamento, dagli anni Novanta fino al 2019, attraverso tutte le sue serie di fake photography. Ora che gli algoritmi e l'intelligenza artificiale stanno andando via via sempre più a sostituire la macchina fotografica e l'occhio, Fontcuberta ripensa con le sue opere – realizzate con la tecnologia delle reti neurali generative – il ruolo delle immagini che finora hanno contribuito a costituire o a forgiare la nostra sensibilità. Nel 2021, con la serie *Déjà-vu*, innesca un processo che mette in discussione le nozioni di "artist*", "conservator*", “storic* delle immagini”, "curator*", paradigma desacralizzante della creatività e dell'arte. Pone una domanda cruciale: il significato di un'immagine non sta nella sua origine ma nella sua destinazione? Inoltre prende in considerazione e analizza, sia con lo sguardo dell’artista sia con quello del critico, gli errori del sistema stesso, i test falliti, i passaggi intermedi e l'inconscio tecnologico che emerge dagli incidenti. In questi cortocircuiti ritroviamo ancora qualcosa che apparteneva ad altre tecnologie e al passato degli esseri umani. Così, in una certa misura, tutto appare "déjà-vu", già visto e preimmaginato. Le

tecnologie deepfake sempre più pervasive occupano grandi spazi nell'iconosfera e generano continuamente volti di persone che non sono mai nate e mai esistite nella vita quotidiana degli esseri umani, milioni di ritratti iperrealistici che paiono della stessa specie delle persone che respirano nel mondo, ipotesi di storie non ancora accadute e che provengono da un probabile futuro. Entrambi i mondi sono ormai parte della realtà attuale. In questo momento non si tratta più di porsi le domande: “A quale verità dobbiamo credere?” o “Come dobbiamo porci in rapporto con tutto quello che sappiamo non essere totalmente vero e che appartiene a una costruzione di matrice tecnologica?”. Ora che abbiamo vissuto il passaggio da un realismo basato sull'ottica a un realismo basato sull'informatica, come orientiamo la nostra percezione sulle cose che accadono? Come agirà la nostra percezione, che è sballottata entro altri sistemi di formattazione dello sguardo e altri regimi di visione compattati nell'automazione computerizzata della rappresentazione?

Leggi anche
Sara Benaglia | Che cosa è il postfotografico?
Elio Grazioli | Fontcuberta. La furia delle immagini
Silvia Mazzucchelli | Descubrir Fontcuberta: un’intervista
Gigliola Foschi | Oltrefotografia
Aurelio Andrighetto, Le insidie delle immagini

In copertina, Joan Fontcuberta, Fungus viridis, 2023, della serie De Rerum Natura.

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.
Torna presto a leggerci e SOSTIENI DOPPIOZERO

