

Spazio latente

Sull'utilità e il danno delle IA generative per l'immaginazione.

di [Pietro Montani](#) – 30 Marzo 2026



A poco più di 5 anni dall'irruzione delle intelligenze artificiali generative nella nostra quotidianità sembra possibile fare un primo bilancio delle cose che crediamo di aver capito sul loro funzionamento e di quelle che ci piacerebbe capire meglio a proposito dei loro effetti e potenzialità.

Proverei a muovere da un profilo antropologico affidabile. Inserirei, cioè, i chatbot di carattere linguistico (quelli che generano immagini richiedono un approccio un po' diverso di cui dirò più avanti) nel numero delle grandi *esternalizzazioni* di cui l'essere umano si è servito nel corso della sua evoluzione specifica. Per riferirci all'esternalizzazione più direttamente implicata con i *Large Language Models* (LLM) in uso nei chatbot pensiamo alla scrittura alfabetica, forse la più ricca di conseguenze profonde e durature per le nostre forme di vita. **Ebbene, il primo punto da mettere in chiaro, e da tenere ben saldo, è che i LLM non esternalizzano "il linguaggio" (posto che esista qualcosa come "il" linguaggio) ma esternalizzano – cioè ne automatizzano la gestione delegandola a un dispositivo computazionale di carattere statistico e predittivo – le procedure dell'argomentazione coerente.** Nessuno stupore, dunque, che ChatGPT o Claude argomentino meglio di noi. Ci meravigliamo, forse, che un qualunque

calcolatore ci surclassi nella pratica del far di conto?

Per tranquillizzarci un po' aggiungiamo subito un dato di fatto che dovrebbe farci riflettere. Le principali esternalizzazioni a cui noi umani abbiamo fatto ricorso sono state, di regola, occasione di formidabili rielaborazioni creative. La scrittura alfabetica, per dire, è di certo collegata con le più antiche forme di mnemotecnica (pensiamo ai "token" e alle "bulle" di argilla incise usate per tenere il conto di beni d'uso e di scambio), ma la sua nascita coincise a tutti gli effetti con un'*emergenza creativa originale* in questo senso preciso: che non era contenuta analiticamente nel paradigma funzionale della mnemotecnica (come *ex post* siamo inclini a ritenere) ma fu il risultato di una brillante *integrazione innovativa* ("sintetica" potremmo dire con una terminologia kantiana) immaginata da qualcuno. **Così, dell'immaginazione tecnica umana possiamo dire in modo non equivoco che è creativa proprio per questa sua capacità di escogitare nuove regole per integrare elementi eterogenei** (nella fattispecie le iscrizioni mnemotecniche, da un lato, e l'articolazione linguistica, dall'altro) **ottenendone dispositivi che, da quel momento in poi, ci si presentano come "individui" autonomi, capaci di co-evolvere insieme a noi.** È una delle cose che ci ha insegnato Gilbert Simondon.

Ora chiediamoci: sono capaci di "integrare" i chatbot? Certo che sì. E lo fanno mirabilmente. Ma sono anche capaci di integrare *in modo innovativo*? Cioè di ipotizzare e sperimentare *nuove* regole di integrazione? **E qui la risposta dev'essere negativa: le regole di integrazione i chatbot le estraggono pari pari dagli immani repertori testuali sui quali si addestrano e in nessun caso (almeno per il momento) sono in grado di escogitarne di nuove.** Così, se conveniamo che la proprietà saliente dell'immaginazione umana è quella di sperimentare nuove regole per integrare elementi eterogenei dovremo ammettere che i LLM ne difettano.

È facile, a questo punto, trarre un paio di conseguenze significative da questa conclusione. La prima è che il conservatorismo strutturale dei chatbot ne plasma da cima a fondo l'interazione con gli umani, della quale non dovrà stupirci che essa susciti empatia e altri sentimenti. In che senso la plasma? Nel senso che anche la tonalità affettiva di base dei nostri colloqui con i chatbot si restringe nei confini di una convenzionalità altrettanto marcata di quella semantica e concettuale: un desolante appiattimento sul "buon senso comune" nell'accezione più triviale. **La seconda conseguenza è che l'effetto "disciplinare" (involontario?) prodotto dall'azione congiunta del conformismo cognitivo e di quello affettivo erogato a piene mani dai chatbot possa costituirsi come un modello per le nostre relazioni intersoggettive cosicché, come ha denunciato Vittorio Gallese nel suo recente *Il Sé digitale*, il vero**

rischio delle IA generative non è che ci sostituiscano ma che ci modellino, deprimendo e uniformando i nostri processi di individuazione. Vien voglia di chiedersi, a questo punto, se l'altra versione delle IA generative, quella degli algoritmi che riconoscono, analizzano e producono immagini, non possa in qualche misura riequilibrare questa malinconica omologazione.

Se muoviamo anche in questo caso dal fenomeno dell'esternalizzazione, tuttavia, dovremo registrare una discrepanza rilevante, e cioè che questi modelli di IA generativa non si possono sussumere sotto la stessa categoria funzionale dei precedenti per il fatto, macroscopico e tuttavia inavvertito, che essi *non automatizzano prestazioni semiotiche umane allo stesso titolo dei chatbot linguistici*. Voglio dire che mentre l'argomentare coerente *fa parte* della nostra competenza linguistica standard e che i LLM la acquisiscono processando un numero strabiliante di testi già esistenti, intere immense biblioteche convertite in database, le IA che generano immagini si addestrano su repertori approntati *ad hoc* costituiti da miliardi di "Text-Image Pairs", cioè da figure associate a un'etichetta verbale appositamente predisposta. Che la differenza tra i due sistemi generativi sia rilevante già sul piano dei materiali destinati al loro training è confermata dal fatto che essendo l'associazione tra un'immagine e una definizione verbale il risultato di una scelta largamente arbitraria, queste IA generative sono afflitte da *biases* culturali particolarmente vistosi, talvolta gravi, alla cui decostruzione si sono dedicate con grande zelo intere correnti di attivismo digitale e di arte militante.

Ma è il processo generativo come tale a far registrare la differenza più significativa in quanto l'immaginazione artificiale si fonda, com'è noto, sul lavoro di ottimizzazione progressiva dell'output affidato alle cosiddette GAN (*Generative Adversarial Networks*), la cui prestazione competitiva (tra un *Generator Network* che abbozza una certa selezione di tratti pertinenti iconici e un *Discriminator Network* che provvede via via a emendarla) è inadatta a processare i testi linguistici che, operando con unità discrete dotate di senso quali sono le parole, non potrebbero sottoporsi con vantaggio ad una analoga procedura di modellaggio e raffinamento dell'output conclusivo. Ciò da un lato conferma l'irriducibilità dell'immaginazione artificiale a un processo di esternalizzazione: non è un caso, del resto, che la generazione si articoli all'interno di un "*latent space*", com'è stato definito, una scatola nera di cui con certezza si può solo dire che non condivide nulla con le procedure semiotiche amministrare dall'immaginazione umana. **Dall'altro, e proprio in forza di questa indubbia originalità produttiva, è occasione di un'intera serie di interrogativi che qui mi limiterò a discutere sotto un solo profilo, parziale, certo, ma anche saliente: l'attitudine dell'immaginazione artificiale a innescare *autentiche* relazioni interattive.** Ne evidenzierò due grandi tendenze facendomi guidare da questa domanda: nei processi interattivi implementati

dall'immaginazione artificiale si fa valere, e in che modo, quella capacità di sperimentare *nuove regole di integrazione* che ho presentato come uno dei tratti distintivi dell'immaginazione creativa di noi umani e di cui ho denunciato l'assenza nei LLM?

La prima tendenza, fin qui la più ricca, consiste nel dialogare con gli algoritmi dell'immaginazione artificiale per saperne di più a proposito del "latent space", delle sue procedure occulte e delle "allucinazioni" nelle quali incappa. Ora, se nel far questo l'interesse si focalizza sull'eventualità che queste procedure possano condurre all'emergenza di regole integrative nuove la risposta che si è subito presentata è che ciò può accadere solo a condizione che la GAN si lasci modellare da *interventi mirati* eseguiti dall'interlocutore umano con cui interagisce. Ciò può accadere secondo diverse modalità procedurali, la più significativa delle quali consiste nello specializzare il campo d'azione dell'immaginazione artificiale ricorrendo al cosiddetto "*fine-tuning*", cioè alla scelta accurata di un corpus ristretto di dati su cui addestrare l'algoritmo e in base a cui orientare le prestazioni generative richieste.

Alcuni artisti visivi, ad esempio, insegnano alle GAN a lavorare in sinergia con la loro prassi esecutiva – di cui il sistema impara rapidamente a imitare lo stile, a proporre varianti e a riconoscerne le anomalie come occasioni di "serendipity" – costruendo in tal modo una *memoria operativa condivisa* che permette di trattare lo sviluppo di un lavoro in comune *anche* come un laboratorio per la sperimentazione di eventuali emergenze integrative di carattere innovativo. Quale sia il limite di questo tipo di "*fine-tuning*" è evidente: il lavoro creativo compiuto nel corso dell'interazione tra artista e GAN è in via di principio circolare e autoriferito, finendo per somigliare a quelle opere manieristiche che sono originali, sì, ma non "originanti". Macchine celibi senza progenie.

Chiuderò con l'uso del "fine-tuning" che giudico più promettente e che vorrei esemplificare con i protocolli complessi messi a punto in diversi anni di esperienza dal gruppo di *Forensic Architecture* nel corso delle sue inchieste per documentare crimini di guerra, violazioni dei diritti umani, misfatti di carattere ecologico. Il gruppo com'è noto ricorre a tecniche di modellazione in 3D e di analisi spaziale che vengono implementate sullo sfondo di un presupposto teorico generale: quello secondo cui gli archivi audiovisivi (oggi unificati nel formato dei *big data*) si possono esplorare attivamente tramite selezioni, comparazioni e rimontaggi di carattere multi e intermediale (non solo visivi, dunque) di volta in volta afferenti a un compito investigativo determinato – ad esempio: l'analisi satellitare delle modifiche del territorio della striscia di Gaza negli ultimi due anni – includendo in questo lavoro il "*fine-tuning*" di diverse GAN cui si chiede di ricavare dal materiale indagato componenti che risultino *capaci di*

integrare mappature multimodali del target in grado di far emergere dai dati un *nuovo* senso, una vera e propria ri-figurazione.

Il presupposto, al tempo stesso semplice e rivoluzionario, di questa pratica è che dal mondo "datificato" si possano *estrarre* gli elementi di *verità* che esso *già* contiene nella forma di tracce multimodali di volta in volta rfigurabili in modo appropriato ed eloquente. Si tratta di una pratica "estrattiva" virtuosa, dunque, che inaspettatamente attualizza uno dei nomi più antichi della verità: il greco *a-letheia* che designa, alla lettera, il risultato di un "trarre fuori" o di un "dis-occultare".