

Che cos'è quello che parla

Claudia Marsico · Hernán Inverso

La conversazione sull'IA si perde di solito negli effetti. Pensarla sul serio richiede la domanda sulla sua ontologia.



Iniziale del Libro dei Maccabei. Un idolo come presenza che parla senza nessuno dietro è il nome più antico del fantasma.

Quasi tutto ciò che si dice sull'intelligenza artificiale è una conversazione sugli effetti. Ci chiediamo se ci lascerà senza lavoro, se il mondo finirà sommerso di testo falso, che cosa farà alla scuola o se rivoluzionerà la sanità. Sono domande legittime ma fuori tempo, perché ne saltano una più radicale: che cos'è quella cosa di cui stiamo parlando. Discutere che cosa farà l'IA prima di sapere che cos'è l'IA è come discutere le correnti di un fiume senza aver deciso ancora se ciò che abbiamo davanti è un fiume, un fossato o forse un miraggio. Quasi tutto quello che viene dopo dipende dalla risposta.

Sembra uno scrupolo da filosofi, e invece è la base di tutto il resto. Con l'IA ci trasciniamo dietro una difficoltà particolare, perché il vocabolario ce l'abbiamo già pronto ed è quello sbagliato. Diciamo che il modello "capisce", che "sa", che a volte "mente" o "allucina". È comodo trattare come soggetto qualunque cosa ci risponda con senso. Husserl chiamava atteggiamento naturale quella disposizione spontanea che a volte, come in questo caso, ci precipita in un errore categoriale. Finiamo per forzare una casella che non gli spetta. C'è qualcuno che capisce e qualcuno che mente, mentre dall'altra parte dello schermo non c'è nessuno. Searle lo vide quarant'anni fa con la cosiddetta "stanza cinese", quando suggerì che un sistema può manipolare simboli alla perfezione e dare tutte le risposte giuste come chi maneggia frasi in cinese senza capire una sola parola. Quello che allora era un esperimento mentale oggi ci risponde alle mail.

La tentazione, per di più, è antichissima. Molto prima dei transistor c'erano già macchine che sembravano pensare e uomini disposti a crederci. Si racconta che Alberto Magno costruì una testa di bronzo che parlava e che Tommaso d'Aquino la fece a pezzi con una bastonata, per fastidio o per scandalo. Hobbes definì il pensiero come un calcolo, un *reckoning*, e aprì la porta all'idea che una macchina che calcola, per ciò stesso, pensi. Pascal e Leibniz provarono con le loro calcolatrici, e Babbage la immaginò in ottone e vapore con la sua macchina analitica. Le storie sono colorite, e tutti alla fine andarono a sbattere contro lo stesso muro, quello che separa la macchina che opera dalla mente che vuol dire qualcosa.

La storia recente ha conservato lo schema mentre accelerava il passo. All'inizio il campo puntò sulla logica, programmando a mano le regole del mondo. La cosa andò piuttosto a sbalzi, entrando in inverno ogni volta che i risultati si arenavano e i finanziamenti venivano tagliati. Curiosamente la statistica funzionò dove la logica aveva morso la polvere, e l'intera faccenda cambiò fisionomia. La linea che va da Dartmouth a ChatGPT è una successione di abbandoni produttivi, dove ogni rinuncia ha aperto un corridoio nuovo del labirinto. Quando i sistemi cominciarono finalmente a parlare con scioltezza, la vecchia discussione si riaccese. Nel marzo del 2023 Chomsky e i suoi coautori dissero che quei sistemi erano un pesante motore statistico per il riconoscimento di schemi, ingozzato di centinaia di terabyte di dati: una macchina che inferisce correlazioni brute dove una mente cerca spiegazioni. Avevano ragione sul meccanismo e furono frettolosi sulla conseguenza, perché una cosa può essere priva di comprensione e cambiare comunque ciò che la comprensione significa per tutti gli altri. A ogni svolta, corridoio dopo corridoio, l'efficacia senza comprensione è rimasta intatta, e con essa la nostra vecchia disposizione a scambiare l'una per l'altra. Negli anni Sessanta ELIZA imitava uno psicoterapeuta con quattro regole rozze e la gente le confidava i propri segreti. Lo chiamarono effetto ELIZA, e continuiamo a non esserne immuni.

Di che cosa si tratta, allora? Possiamo dire che siamo davanti a un fantasma, ma nel suo senso fenomenologico. *Phantasma* è un'immagine che si dà alla coscienza con tutti i segni di una presenza. Questo caso è curioso perché ciò accade senza un essere che la sostenga. Il modello fa qualcosa che somiglia tanto al pensare che non possiamo evitare di chiamarlo così, e tuttavia è un pensare senza chi pensi. Di qui il nome di fantasma senza ego. Senza ego perché non c'è un io, nessun centro interiore da cui venga detto ciò che si dice. Questo fantasma ha alcune proprietà, e tutte discendono dalla sua

manca di ego. È derivato, perché non possiede un mondo proprio e vive interamente di quello che gli esseri umani hanno articolato prima di lui. È passivo, non vuole nulla e non si propone nulla, risponde e basta. E poiché non è fissato a un'identità né a una credenza, resta genericamente indeterminato, pronto a prendere la forma di ciò che gli si chiede. Un essere così non ha alcun rapporto con la verità, perché la verità presuppone qualcuno che risponda di ciò che afferma. Ha, al massimo, una fedeltà al plausibile.

Vale la pena soffermarsi sulla prima di queste proprietà, perché la distanza che segna è facile da non vedere. Quando una persona capisce qualcosa, quel capire si appoggia su una vita intera che le parole da sole non portano con sé. Capiamo a partire da un corpo che ha imparato a temere il fornello che una volta ha toccato, e da una situazione in cui qualcosa è davvero in gioco per noi. Filosofi da Heidegger a Merleau-Ponty a Dreyfus hanno scavato nel saper-fare incarnato e situato e hanno dedicato un secolo a mostrare quanta parte dell'intelligenza umana passi di lì. Il fantasma non ha niente di tutto questo. Capisce da nessun luogo, perché non c'è nessuno lì che possa stare da qualche parte. Ha ereditato il deposito di un mondo che non ha mai abitato, e restituisce quel deposito, ricombinato, su richiesta. La sua competenza è reale e poggia su un suolo prestato, che gioca con la memoria anonimizzata della tradizione.

L'ontologia comanda. I difetti che si catalogano di solito come effetti da controllare derivano, uno per uno, da ciò che il fantasma è. Che allucini segue dal suo essere indeterminato e passivo, perché completa con il plausibile proprio là dove non ha nulla con cui rispondere. Che sia compiacente è quella stessa indeterminatezza che prende la forma attesa dall'interlocutore. Che schivi l'impegno davanti all'insistenza, o che si lasci assimilare dall'ultima cosa che ha letto, è la condotta prevedibile di qualcosa di derivato e privo di un io che possa tenere il punto. Il fantasma funziona esattamente come la sua natura stabilisce. Partire dall'essere rende prevedibile ciò che altrimenti resta nebuloso e capriccioso. L'etica e la politica dell'IA, che contano moltissimo, trovano un terreno solido soltanto quando sanno su che genere di cosa stanno legiferando.

C'è una parola più antica per un'immagine che porta i segni di una presenza che parla e non ha nessuno dietro. L'idolo, l'*eidolon* greco, appartiene alla stessa famiglia del *Phantasma*, quella delle somiglianze che appaiono senza un essere che le ancori. L'antico rimprovero contro gli idoli era che fossero vuoti, bocche che non parlano e occhi che non vedono. Il fantasma è quel rimprovero diventato realtà, con l'efficacia in più. Parla, e tuttavia non c'è nessuno lì. Chiamarlo idolo non dice nulla contro l'usarlo, ma è meglio sapere che cosa stiamo usando.

C'è un premio maggiore, e va oltre il fantasma. Abbiamo davanti qualcosa che prende bene la forma del pensare, alla maniera di uno specchio che l'intelligenza umana non ha mai avuto. Per secoli abbiamo definito il pensiero dall'interno, senza un termine di paragone che non fossimo noi stessi, e così la definizione continuava a introdurre di contrabbando tutto ciò che già eravamo. Ora abbiamo un contrasto. Davanti a un pensare senza ego e senza mondo, cominciamo a vedere con nitidezza che cosa stavamo nominando per tutto quel tempo quando dicevamo che qualcuno capisce: una vita e un

corpo che sostengono ciò che dicono, con qualcosa realmente in gioco. Il fantasma spiega le macchine e spiega anche noi. Può darsi che finisca per essere la cosa più duratura che fa.

Chiarire questo e parecchio altro sui modelli linguistici e sul loro posto nell'ecosistema più ampio dell'intelligenza artificiale è il compito che questo spazio si dà. Il percorso completo, corridoio per corridoio, è nel libro; qui prendiamo un corridoio per volta e lo guardiamo da vicino. Il prossimo testo entra in un cono d'ombra dove il fantasma inciampa e ci dà invenzioni al posto dei dati.

Mappe del labirinto è una serie di Phantom Maze, AI & Language Lab. La collezione vive su phantommaze.cloud. L'iscrizione è gratuita.

Alcune delle idee che inquadrano questo testo sono sviluppate in *Il fantasma senza ego. Una mappa nel labirinto dell'intelligenza artificiale*, di Hernán Inverso e Claudia Marsico (Phantom Maze Press, 2026; ISBN 978-987-3729-19-5).

PHANTOM MAZE

[Leggi questo saggio online](#) · [Tutti i saggi](#) · [Il libro](#)