L’illusione di sapere: l’eclissi del dubbio nell’era della IA

Data 15 novembre 2026
Categoria Medicinadigitale

L’intelligenza artificiale generativa sta trasformando il nostro rapporto con la conoscenza, producendo risposte e ragionamenti sempre più simili a quelli umani. Ma parlare come un essere umano non significa pensare come lui.

Quattro recenti studi esplorano la questione, interrogandosi non solo sulla natura dell’intelligenza artificiale, ma anche su come essa stia cambiando la percezione della nostra stessa intelligenza. Quattrociocchi, Capraro e Perc individuano profonde differenze tra i processi attraverso cui gli esseri umani e i modelli linguistici producono i propri giudizi. L'intelligenza umana si sviluppa attraverso le esperienze corporee, il ricordo, le emozioni, le motivazioni, le interazioni sociali e il contesto, la capacità di valutare criticamente le proprie convinzioni. I grandi modelli linguistici, invece, elaborano rappresentazioni statistiche apprese da enormi quantità di dati e generano risposte attraverso processi computazionali. La differenza non riguarda necessariamente la qualità apparente del risultato. Una macchina può fornire una risposta corretta, convincente e articolata senza averla prodotta attraverso gli stessi percorsi cognitivi di un essere umano. La somiglianza delle risposte non dimostra peraltro la somiglianza dei processi che le generano. Gli autori introducono a questo proposito il concetto di **Epistemia**: una condizione nella quale la plausibilità linguistica sostituisce la valutazione epistemica. In altre parole, riceviamo una risposta che sembra conoscenza senza aver necessariamente compiuto il lavoro di verifica, comprensione e giudizio necessario per farla nostra. Il rischio non consiste soltanto nell'accettare informazioni false ma soprattutto nell'abituarsi a considerare la disponibilità immediata di una risposta come equivalente alla comprensione di un problema.
Questa preoccupazione teorica trova un importante riscontro sperimentale in un altro studio. Attraverso cinque esperimenti che hanno coinvolto complessivamente 3.132 partecipanti, i ricercatori hanno analizzato come cambia il giudizio umano quando si ha a disposizione il suggerimento di un'intelligenza artificiale. Le domande erano difficili, i partecipanti potevano scegliere di non rispondere e i suggerimenti dell'IA erano deliberatamente sbagliati. I risultati sono particolarmente significativi: la semplice disponibilità dell'assistenza artificiale ha ridotto drasticamente la propensione a dichiarare la propria incertezza. Le persone hanno risposto a più domande, ma la loro accuratezza è scesa a circa un terzo di quella osservata senza l'IA, mentre la fiducia nelle proprie risposte è quasi raddoppiata. Anche gli incentivi economici a rispondere correttamente hanno attenuato, ma non eliminato, questo fenomeno. Si delinea così un paradosso: **più ci affidiamo a uno strumento che sembra sapere tutto, meno siamo disposti a riconoscere ciò che non sappiamo** . Eppure il dubbio è una componente essenziale del pensiero critico. Dire «non lo so» non rappresenta un fallimento intellettuale, ma il riconoscimento di un limite da cui può iniziare una ricerca autentica. Se questa capacità si indebolisce, rischiamo di sostituire progressivamente la costruzione personale della conoscenza con l'accettazione di risposte già confezionate. Gli esperimenti dimostrano questo effetto nelle condizioni studiate, resta da verificare in che misura esso si manifesti stabilmente nella vita quotidiana, nell'istruzione e nelle decisioni professionali.

L'LLMorfismo: quando iniziamo a interpretarci come macchine
Un'altra conseguenza, più sottile, è descritta da Valerio Capraro. Siamo ormai abituati a discutere dell'antropomorfismo dell'intelligenza artificiale, cioè della tendenza ad attribuire alle macchine qualità umane. Capraro propone di considerare anche il fenomeno inverso: l'LLMorfismo, ossia la **tendenza a interpretare il funzionamento della mente umana attraverso le caratteristiche dei modelli linguistici** . Se una macchina riesce a parlare come noi, potremmo essere portati a concludere che anche il nostro pensiero funzioni essenzialmente come una macchina che elabora e predice sequenze linguistiche. Questo slittamento culturale può avere conseguenze importanti. L'apprendimento rischia di essere ridotto all'elaborazione di informazioni, la creatività alla ricombinazione di elementi esistenti, il giudizio morale a una procedura di calcolo. L' LLMorfismo è, per ora, una proposta teorica e un'ipotesi culturale da studiare empiricamente, non un fenomeno di cui sia già dimostrata la diffusione. Il suo valore consiste nel segnalare un pericolo: non soltanto attribuire troppo alle macchine, ma arrivare ad attribuire troppo poco agli esseri umani.
Queste riflessioni convergono nel più recente lavoro di Capraro, Gary Marcus e Quattrociocchi. Gli autori contestano l'idea che la capacità di imitare un essere umano debba rappresentare il principale criterio per valutare i progressi dell'intelligenza artificiale. Il tradizionale test di Turing, basato sulla difficoltà di distinguere una macchina da un interlocutore umano, misura infatti una forma di somiglianza comportamentale, non necessariamente la qualità del ragionamento o l'affidabilità della conoscenza. Ma la critica è anche etica e sociale: perché dovremmo costruire macchine che imitano gli esseri umani, riproducendone eventualmente errori e pregiudizi, invece di sviluppare strumenti capaci di ampliare le possibilità della nostra intelligenza? La proposta degli autori è quella di spostare l'attenzione dall'imitazione al benessere umano. Il successo di un sistema artificiale dovrebbe essere misurato considerando quanto migliori effettivamente la capacità delle persone di comprendere, decidere, apprendere e agire autonomamente. Questo richiede ovviamente nuove responsabilità per chi sviluppa, finanzia e adotta tali tecnologie: non limitarsi a misurarne le prestazioni, ma valutarne gli effetti concreti sulle persone, sulla distribuzione dei benefici e sull'autonomia individuale.

Conclusioni
Considerati nel loro insieme, i quattro lavori delineano una questione che va oltre l'affidabilità tecnica dell'intelligenza artificiale. La trasformazione più profonda potrebbe riguardare il nostro rapporto con il sapere. Le macchine producono risposte sempre più sofisticate, ma la conoscenza umana non coincide con la semplice disponibilità di risposte. Conoscere significa anche comprendere le ragioni di un'affermazione, riconoscerne i limiti, confrontarsi con l'esperienza, assumersi responsabilità e saper cambiare idea. La sfida, quindi, non è contrapporre l'intelligenza umana a quella

artificiale, né rinunciare alle straordinarie opportunità offerte dai nuovi strumenti. È progettare una relazione nella quale la tecnologia rafforzi, anziché indebolire, le nostre capacità cognitive. Nella scuola, nel lavoro, nell'informazione e nella vita pubblica, ciò significa preservare lo spazio del dubbio, della verifica e del confronto. Significa insegnare non soltanto a utilizzare l'intelligenza artificiale, ma anche a interrogarla criticamente. Il vero progresso non si misurerà allora dalla capacità delle macchine di assomigliarci sempre di più, bensì dalla loro capacità di aiutarci a pensare meglio. Il rischio maggiore infatti non è che le macchine imparino a parlare come noi, ma che noi, affascinati dalle loro risposte, smettiamo gradualmente di esercitare la nostra capacità di giudicare.

Giampaolo Collecchia e Riccardo De Gobbi

Riferimentibibliografici

Capraro, V. (2026). LLMorphism: When humans come to see themselves as language models. arXiv:2605.05419.
Capraro, V., Marcus, G., & Quattrociocchi, W. (2026). Human likeness is a misplaced benchmark for artificial intelligence. osf.io/preprints/psyarxiv/zn5u9_v1
Marcoccia, C., Quattrociocchi, W., & Capraro, V. (2026). AI advice suppresses people's willingness to say “I don't know”, even when the advice is wrong and accuracy is incentivized. arXiv:2607.13562
Quattrociocchi, W., Capraro, V., & Perc, M. (2025). Epistemological fault lines between human and artificial intelligence. arXiv:2512.19466.