

L'eclissi del pensiero critico? Intelligenza Artificiale, debito cognitivo e responsabilità

09 Gennaio 2026

© Riproduzione riservata

Quotidiano Sanità <https://share.google/d6FO82ovASFD5wEsj>



LUCIO ROMANO

Gentile Direttore, ci è mai capitato di non ricordare il numero di telefono di un nostro familiare perché salvato in rubrica? O di sentirci smarriti senza Google Maps anche in una zona che conosciamo? Ecco, questi rappresentano degli esempi semplici e ricorrenti che dimostrano gli effetti svantaggiosi del *cognitive off-loading* o “scarico cognitivo”. Inteso come affidamento a supporti esterni su cui scaricare funzioni mentali come memoria, calcolo, analisi dei dati, decisione.

L'*off-loading* cognitivo offre anche vantaggi immediati come riduzione del carico mentale, maggiore efficienza, ottimizzazione dei tempi e risparmio di energie per attività complesse. Tuttavia, se utilizzato in modo acritico e sistematico come vera e propria dipendenza, con limitazioni dell'impegno mentale, può produrre come conseguenza un debito cognitivo. Ovvero un beneficio a breve termine per comodità e velocità che si paga nel tempo con l'indebolimento della capacità critica, della profondità di analisi e dell'autonomia di pensiero. Fino ad una sorta di atrofia del pensiero critico che crea l'illusione della competenza: leggiamo un output generata dall'IA e crediamo di aver capito l'argomento ma in realtà abbiamo solo letto il risultato finale senza aver fatto il percorso logico necessario per arrivarci.

Un focus è stato dedicato al tema dal NEJM nell'articolo dal titolo Educational Strategies for Clinical Supervision of Artificial Intelligence Use e, in particolare, in merito al ricorso a sistemi artificiali intelligenti (IA) da studenti e medici in formazione.

Per quanto i concetti di “scaricamento cognitivo” e “debito cognitivo” siano stati analizzati in un settore ben preciso, quale appunto quello della formazione biomedica, sono certamente estendibili ai più vari ambiti. Sempre più ampi viste le molteplici applicazioni dei sistemi di IA generativa. Ebbene il debito cognitivo, conseguenza dell'uso dell'IA come sostituto del ragionamento clinico

piuttosto che come supporto (*informing*), comporta diversi rischi per lo sviluppo delle competenze: perdita progressiva di abilità (*deskilling*), mancato sviluppo di abilità (*never-skilling*) e sviluppo distorto di abilità (*mis-skilling*).

Sulla base della letteratura scientifica, misurando l'uso di strumenti di IA, il *cognitive off-loading* e le capacità di pensiero critico, si è riscontrata una significativa correlazione negativa tra l'uso frequente di sistemi di IA e le abilità di pensiero critico. Nello specifico si è rilevato un rapporto diretto tra maggiore dipendenza dall'IA e ridotto coinvolgimento nella risoluzione autonoma dei problemi e nel ragionamento analitico. Risultati che indicano uno spostamento verso output generati dall'IA piuttosto che verso uno sforzo cognitivo personale. Vale a dire che questo risultato è riconducibile a una interazione in cui gli utenti aggirano il proprio giudizio e si limitano ad assecondare l'output del sistema. Possiamo dire che fino a ieri delegavamo la memoria di numeri, date, mappe. Oggi, con l'IA stiamo iniziando a delegare funzioni che richiedono capacità predittive in contesti di incertezza.

Un ambito particolarmente critico è quello del *mis-skilling* (sviluppo distorto di abilità) che può verificarsi quando l'apprendimento è affidato totalmente e senza analisi critica a modelli inaccurati di IA. Ad esempio, in una ricerca pubblicata su JAMA, i clinici ai quali venivano mostrate previsioni diagnostiche generate dall'IA con bias sistematici erano più propensi ad adottare tali previsioni errate. Ciò significava, appunto, uno sviluppo distorto di abilità nel riconoscere e correggere i *bias* dell'IA fino a rinforzare ulteriormente lo stesso *mis-skilling*.

I rischi che abbiamo rappresentato significano forse una visione catastrofista o un atteggiamento neoluddista nei confronti dell'IA applicata in medicina? Certamente no. Ignorare il supporto dell'IA e le relative raccomandazioni, anche quando sono corrette, sarebbe una perdita di opportunità basate sul ricorso all'IA. Al contrario, la combinazione del ragionamento umano e dei risultati dati dall'IA migliora le prestazioni complessive, suggerendo che un'elevata e imprescindibile competenza di base favorisce un uso efficace e sicuro dell'IA.

Ma allora quale modello di collaborazione tra umano e IA è più appropriato? Come riportato nel citato articolo del NEJM, si distinguono due modalità simboliche: il "modello centauro" e il "modello cyborg". Nel modello centauro vi è una netta separazione dei compiti: l'IA fornisce supporto operativo (es: analisi dei dati, stesura di bozze, documenti) ma il giudizio finale e le decisioni restano di competenza umana. Possiamo dire che in questo modello, pur partendo da una separazione delle competenze, si rileva una cooperazione tra sistemi di IA e umano. Il modello cyborg, invece, prevede un'integrazione molto più stretta, con una co-produzione continua tra umano e IA, particolarmente efficiente per attività ripetitive e a basso rischio ma esposta al rischio di *deskilling* e dipendenza. Si evidenzia che la competenza chiave del clinico non è l'adesione rigida a uno dei due modelli, ma la capacità di passare in modo flessibile dall'uno all'altro in base al rischio del compito, al grado di validazione dello strumento e al proprio livello di esperienza.

Rimane un interrogativo: come promuovere il pensiero critico, fondamentale baluardo a fronte di un'eccessiva dipendenza dall'IA? Come arginare i fenomeni di *deskilling*, *never-skilling* e *mis-skilling*? Come promuovere il pensiero critico come struttura di supporto per accelerare lo sviluppo di competenze adattative e, allo stesso tempo, migliorare l'alfabetizzazione all'IA?

Per dare una risposta a questi interrogativi, è stato proposto il modello DEFT-AI che può offrire un approccio strutturato e di buon senso per promuovere il pensiero critico nelle interazioni con i sistemi di IA.

Nello specifico, nella fase di "Diagnosis/Discussion/Discourse" (D) il docente chiede allo studente una descrizione circa l'uso specifico che ha fatto dell'IA, in particolare di esplicitare il proprio ragionamento, quali diagnosi aveva in mente, che informazioni aveva considerato, come aveva utilizzato l'IA e quali prompt aveva inserito. Nella fase "Evidence" (E) si verificano gli output generati dall'IA e si procede a una valutazione delle evidenze cliniche a favore di una certa ipotesi diagnostica così le evidenze a supporto dell'uso di uno specifico sistema di IA. Nello step del "Feedback" (F) si chiede di riflettere sulle opportunità di crescita nell'uso dell'IA, come valutare l'utilizzo nel caso specifico e come migliorare l'impiego. Con il "Teaching" (T) si forniscono insegnamenti mirati sulla base dei risultati emersi e si raccomanda se, quando e come utilizzare l'IA in modo sicuro in futuro. In particolare, si rafforzano le abilità e le competenze, individuando limiti e bias. Infine, con "Recommendation for AI engagement" (AI) si forniscono raccomandazioni personalizzate per un uso sicuro dell'IA. Vale a dire esercitarsi nell'utilizzare l'IA per **informare** il ragionamento, piuttosto che **sostituirlo**. Prima verificare e poi fidarsi. In sintesi, saper riconoscere quando poter fare affidamento (modalità *cyborg*) e quando invece è necessario confermare criticamente gli *output* (modalità *centauro*).

Queste procedure significano sostituire la logica del "*trust and go*" (fidarsi e andare avanti) con la metodologia del "*verify and trust*" (verificare e fidarsi) in cui l'output dell'IA viene prima valutato secondo il principio di esplicabilità e poi utilizzato. In questa ottica si acquisiscono nuove competenze o si potenziano quelle esistenti per adattarsi a tecnologie e pratiche emergenti (*up-skilling*). Si favorisce lo sviluppo di capacità critiche e decisionali, migliorando l'uso di sistemi innovativi di IA. In definitiva con l'*up-skilling* non solo si incrementa efficienza e sicurezza ma ci si protegge anche dai rischi di *deskilling*, garantendo che l'apprendimento e la pratica professionale evolvano insieme alla tecnologia.

Lucio Romano Commissione Scientifica Centro Interuniversitario di Ricerca Bioetica (CIRB)