La fiducia nei confronti della Intelligenza Artificiale è meritata?

Data 19 ottobre 2025
Categoria Medicinadigitale

L'intelligenza artificiale (IA) è ormai una presenza quotidiana. Ciò pone una questione cruciale: possiamo fidarci di una macchina come ci fidiamo di una persona? E, soprattutto, che tipo di fiducia dovremmo accordarle?

Come ricorda J. Beger in un recente paper (1), la fiducia in medicina è sempre stata una faccenda "umana": si costruisce giorno per giorno nella vita professionale con ascolto, empatia, esperienza condivisa. Il medico infatti non è solo un tecnico ma anche un custode della fragilità del paziente. Fidarsi significa, in fondo, delegare: "mi affido a te, convinto che agirai nel mio interesse".
Con l'IA il discorso si complica: niente vera empatia, niente strette di mano rassicuranti, niente sguardi comprensivi. L'algoritmo non prova sentimenti e non sa cosa significhi "capire". Nonostante questi limiti è un dato di fatto che si possa nutrire un sentimento di " fiducia" verso la IA, ovvero un " **Atteggiamento, verso altri o verso sé stessi, che risulta da una valutazione positiva di fatti, circostanze, relazioni, per cui si confida nelle altrui o proprie possibilità, e che generalmente produce un sentimento di sicurezza e tranquillità** " (Vocabolario Treccani). E' giustificato questo positivo sentimento di sicurezza e tranquillità? Il dibattito è vivace e le conclusioni non sempre sono concordi.
Uno studio recente ha rivelato che le risposte fornite da un modello di IA sono state valutate come di qualità superiore rispetto a quelle dei medici. Queste risposte sono state percepite anche come significativamente più empatiche. Tuttavia, simulare empatia non equivale a provarla. L'IA, almeno nella sua forma attuale, non possiede capacità empatiche perché non vive esperienze emotive: non prova gioia né dolore. Di conseguenza, anche se può produrre risposte che sembrano partecipate emotivamente, in realtà non sono autentiche, poiché l'AI non condivide alcuna esperienza e non è in grado di manifestare preoccupazione e cura genuine. Le espressioni empatiche umane indicano inoltre disponibilità a sostenere un peso emotivo e ad impiegare risorse cognitive-emotive nell'interazione. Essendo peraltro tali risorse limitate, esse segnalano l'importanza e la vicinanza del destinatario. Al contrario, l'AI risponde a chiunque con lo stesso entusiasmo, senza costi, e perciò non comunica una attenzione autentica e non fa sentire il destinatario unico o importante. Sebbene l'IA conversazionale possa simulare abilmente interazioni empatiche, a volte persino dando l'impressione di superare l'empatia umana, il desiderio di empatia trascende la semplice ricezione di una risposta empatica ideale. Per questo, invece di cercare di umanizzare l'IA, occorre puntare a progettarla in modo eticamente solido e funzionale, al servizio del giudizio umano, non come suo sostituto. **L'IA non deve imitare il medico, ma rafforzarne le capacità, supportarne il giudizio e contribuire a un'assistenza più equa e condivisa, che includa il bisogno di una cura genuina e di un coinvolgimento emotivo autentico da parte di chi offre supporto** (2).
Nel rapporto con la IA la fiducia non nasce da un rapporto, ma dalla percezione che la tecnologia sia competente, trasparente, sicura ed eticamente progettata. Più che fiducia "umana", è, come afferma Beger, una fiducia calcolata, conseguenza di una valida ingegneria computazionale.
Non mancano rischi: alcuni medici seguono l'IA a occhi chiusi (**fenomeno noto come automation bias**), o ne sopravvalutano le capacità(**overconfidence**)altri ne diventano dipendenti (**overdependence**), altri ancora vedono le proprie abilità "arrugginirsi" (**deskilling**). Un recente studio clinico randomizzato che ha coinvolto 180 radiologi (3), con o senza il supporto della IA, ha valutato l'accuratezza dell'interpretazione delle radiografie del torace. Sebbene i sistemi di IA abbiano superato i radiologi nell'analisi complessiva, è stata riscontrata una marcata eterogeneità, con alcuni specialisti in radiologia che mostravano "negligenza nell'automazione": altamente fiduciosi nella propria lettura ignoravano le interpretazioni dell'algoritmo. In breve: c'è chi si fida troppo, chi troppo poco, e chi deve ancora trovare la misuragiusta.

La questione della responsabilità
Una delle colonne portanti della fiducia in medicina è la responsabilità personale: alla fine, l'attribuzione delle scelte personali, pur condivise con il paziente, è prerogativa del medico. Ma se un'IA propone una diagnosi sbagliata e il medico la segue, di chi è la colpa? Non dell'algoritmo, perché non è una persona. Si entra allora in una zona grigia di responsabilità diffusa e confusa, che rischia di minare la fiducia nel sistema. Ecco perché si parla tanto di governance: non basta un software brillante, serve un impianto di regole chiare e trasparenti.
Queste devono anche tenere di conto che l'algoritmo apprende e si aggiorna in silenzio, cambiando senza avvisare nessuno: un po' come se il medico cambiasse terapia ogni settimana senza dirlo al paziente. Occorre perciò monitorare costantemente le modifiche, fornire feedback e costruire una fiducia dinamica, capace di aggiornarsi insieme al sistema.

Trasparenza e incertezza: meglio dire "non lo so" che inventare
Il paziente non ha bisogno di conoscere i dettagli tecnici, ma di risposte comprensibili e sincere. A volte basta ammettere che non tutto è prevedibile. Niels Bohr lo diceva con ironia: "Fare previsioni è difficile, soprattutto sul futuro". L'errore più comune è quello di essere troppo ottimisti e di sbagliare la prognosi per eccesso. I clinici adottano spesso la formula della risposta sincera, per esempio alla fatidica domanda: "Dottore… cosa succederà?", dicono che purtroppo è molto difficile fare previsioni poiché possono rivelarsi sbagliate (4). E'essenziale una condivisione positiva dell'incertezza, variabile inevitabile, che non va negletta o combattuta ma accettata e affrontata, come occasione di confronto con il paziente, reso partecipe dei molti "non so", per arrivare a una condivisione delle scelte.
Qui entra in gioco anche il tema dell'explainability: un'IA che funziona come una scatola nera può produrre ottimi risultati, ma se non sappiamo come ci è arrivata, la fiducia vacilla, facendo venir meno il "diritto alla spiegazione" (5)

valido per i pazienti nei confronti dei medici ma paradossalmente non valido per le IA verso i medici e tutti gli altri possibili utenti . L'opacità dei sistemi di IA solleva l'interrogativo di come garantire il rispetto per i diritti delle scelte delle persone coinvolte, soprattutto quando una decisione suggerita dall'algoritmo abbia un impatto significativo sulla loro vita.

Conclusioni

Parlare di "fiducia" verso la Intelligenza Artificiale significa da un lato attribuire alla stessa potenzialità conformi ai propri desideri, ma significa anche associare un sentimento positivo ad una entità che non conosciamo e che non siamo in grado di controllare se non parzialmente. Attenzione quindi, perché la fiducia riduce il nostro senso critico, il che potrebbe impedirci di riconoscere errori o carenze nelle risposte che ci forniscono questi mezzi: ci troveremmo quindi nel paradosso di ricercare e correggere bias negli "umani" e di ignorarli nelle " macchine"…

Come gestire allora la nostra " relazione" con la IA? Considerandola un "particolare essere" verso il quale la nostra umanità ci induce a provare emozioni e talora perfino sentimenti, ma che da oggi in poi valuteremo con severità per trovarne tutti i difetti, criticarla e correggerla: tanto Lei non se la prende di certo! Usiamo la nostra umanità e la nostra intelligenza per controllare la IA, per controllarci e per valutare la relazione che con essa stiamo instaurando: lo facciamo quotidianamente con i nostri pazienti possiamo ben farlo con lei…

L'IA non serve a sostituire l'intuito e l'empatia umana, ma a rafforzarli. Il futuro sarà fatto di macchine intelligenti e medici ancora più "intelligenti": non perché in grado di calcolare più velocemente, ma perché sapranno realizzare ciò che nessun algoritmo potrà mai fare: astrarre, intuire, adattarsi, e soprattutto… prendersi davvero cura di un altro essere umano(6).

Giampaolo Collecchia e Riccardo De Gobbi

Bibliografia

1) Beger J. Not someone, but something: Rethinking trust in the age of medical AI www.linkedin.com/in/janbeger/

2) Matan Rubin; Hadar Arnon; Jonathan D Huppert; Anat Perry. Considering the Role of Human Empathy in AI-Driven Therapy. JMIR Ment Health 2024; 11:e56529) doi: 10.2196/56529

3) Agarwal N et al. Combining Human Expertise with Artificial Intelligence Experimental Evidence from Radiology 2023 https://www.nber.org/papers/w31422)

4) Rossi LR. Zona d'ombra. Roma: Il Pensiero Scientifico Editore, 2022

5) Mannelli C. Etica e Intelligenza artificiale. Il caso sanitario. Roma: Donzelli editore, 2022

6) Collecchia G, De Gobbi R. Intelligenza artificiale e medicina digitale. Una guida critica. Roma: Il Pensiero Scientifico Editore,2020