

L'IA è affidabile per la gestione della salute?

Eugenio Santoro

Unità per la Ricerca in Sanità digitale, Dipartimento di Ricerca in Oncologia Clinica, Istituto di Ricerche Farmacologiche Mario Negri IRCCS

L'intelligenza artificiale (IA) è entrata con straordinaria rapidità nel mondo della salute. Dal lancio di ChatGPT nel 2022, medici, ricercatori e pazienti hanno iniziato a utilizzare strumenti di IA generativa per ottenere informazioni cliniche, redigere documenti, supportare decisioni e persino interagire con i pazienti. Tuttavia, la domanda fondamentale rimane aperta: quanto sono realmente affidabili queste tecnologie?

La letteratura scientifica degli ultimi anni suggerisce una risposta prudente. I grandi modelli linguistici (*Large Language Models, LLM*), come ChatGPT, Gemini e Claude, hanno dimostrato prestazioni sorprendenti quando chiamati a rispondere a test per superare esami universitari o per superare le prove che devono superare i laureati in medicina per diventare medici. Prestazioni altrettanto sorprendenti si sono ottenute nella genera-



Foto di Todd Images - stock.adobe.com

“ Molte applicazioni di IA già integrate nella pratica clinica, dalla radiologia alla cardiologia e all’oncologia, appartengono prevalentemente all’ambito predittivo piuttosto che a quello generativo.

zione di test clinici. Tuttavia, le loro capacità non coincidono necessariamente con l’affidabilità richiesta nella pratica clinica reale. Studi recenti evidenziano come le eccellenti performance nei test teorici non garantiscano una corretta gestione dei casi complessi, l’aderenza alle linee guida o l’integrazione nei flussi di lavoro sanitari.

Uno dei problemi più discussi è quello delle cosiddette “allucinazioni”: risposte formulate in modo plausibile ma contenenti errori fattuali, riferimenti bibliografici inesistenti o raccomandazioni cliniche inappropriate. In un contesto ad alto rischio come la medicina, un errore apparentemente marginale può tradursi in conseguenze significative per il paziente. La capacità degli LLM di esprimersi con sicurezza anche quando l’informazione è errata rappresenta una sfida senza precedenti per la valutazione critica delle loro risposte.

Un secondo elemento riguarda la trasparenza. Molti sistemi di IA generativa sono sviluppati come “scatole nere”: né gli utenti né, talvolta, i ricercatori conoscono pienamente i dati utilizzati per l’addestramento, i meccanismi decisionali o le modifiche introdotte con gli aggiornamenti dei modelli. Questo limita la riproducibilità scientifica e rende difficile validare i risultati secondo gli standard tradizionalmente richiesti in medicina. Trasparenza, spiegabilità e tracciabilità delle fonti rappresentano quindi prerequisiti indispensabili per un utilizzo clinico sicuro.

La questione della validazione scientifica è altrettanto cruciale. Una revisione sistematica pubblicata su *Nature Medicine* nel 2026 ha identificato oltre 4.600 studi sugli LLM in medicina, ma soltanto una minoranza utilizzava dati reali di pazienti e appena 19 studi erano *trial* randomizzati prospettici. La maggior parte delle evidenze deriva ancora da simulazioni, *benchmark* o domande d’esame,

contesti che non riflettono la complessità della pratica clinica quotidiana.

Questo scenario contrasta con quello dell’IA predittiva, che rappresenta la forma di intelligenza artificiale con la maggiore storia di applicazione in sanità. I modelli predittivi basati su dati clinici strutturati, immagini diagnostiche, segnali fisiologici o dati genomici sono sviluppati per stimare probabilità di eventi specifici: mortalità, rischio cardiovascolare, recidive tumorali o risposta ai trattamenti. Sebbene anch’essi presentino limiti e possibili *bias*, il loro sviluppo è generalmente accompagnato da procedure consolidate di validazione interna ed esterna, metriche di accuratezza ben definite e studi prospettici di implementazione. Per questo motivo, molte applicazioni di IA già integrate nella pratica clinica, dalla radiologia alla cardiologia e all’oncologia, appartengono prevalentemente all’ambito predittivo piuttosto che a quello generativo.

Non sorprende, quindi, che le applicazioni più mature dell’IA in sanità siano quelle orientate al supporto decisionale quantitativo, mentre l’IA generativa trova oggi maggiore utilità nelle attività amministrative, nella sintesi documentale, nella redazione di referti e nel supporto informativo. Alcuni studi mostrano che l’impiego di sistemi generativi può migliorare l’efficienza operativa senza ridurre la qualità clinica, ad esempio nella produzione di documenti frutto dello *scribing* (l’interpretazione e la raccolta di dati automatica da una conversazione con il paziente durante la visita medica).

Un tema ulteriore riguarda i *bias*. Gli LLM apprendono da enormi quantità di dati che riflettono inevitabilmente le disuguaglianze presenti nella società e nei sistemi sanitari. Recenti evidenze mostrano che sistemi di IA possono produrre raccomandazioni differenti in base a caratteristiche socioeconomiche o demografiche, amplificando disparità già esistenti. Non dobbiamo poi dimenticare che affidarsi sistematicamente a un sistema di intelligenza artificiale genera un rischio a cui non sempre si presta la dovuta attenzione: la progressiva delega cognitiva e il

“ Uno dei problemi più discussi è quello delle “allucinazioni”: risposte formulate in modo plausibile ma contenenti errori fattuali.

“ La medicina non consiste esclusivamente nell'applicazione di conoscenze scientifiche, ma richiede soprattutto l'esercizio del giudizio clinico.

conseguente effetto “*deskilling*”. La medicina non consiste esclusivamente nell'applicazione di conoscenze scientifiche, ma richiede soprattutto l'esercizio del giudizio clinico. Il medico combina dati biologici, contesto sociale, storia individuale, valori del paziente e aspetti relazionali che non possono essere completamente tradotti in un algoritmo. Un utilizzo acritico dell'IA potrebbe pertanto determinare una semplificazione della complessità umana e, in ultima analisi, una riduzione della capacità di affrontare e risolvere problemi clinici. Ad esempio, alcuni studi hanno evidenziato che l'esposizione a strumenti di IA per il supporto alla diagnosi di adenoma ha reso gli endoscopisti meno abili nel riconoscere correttamente la diagnosi dopo la rimozione di tale supporto tecnologico.

Alla luce di queste considerazioni, la domanda iniziale richiede una risposta sfumata. L'IA non è una tecnologia monolitica: parlare di “affidabilità dell'IA” senza distinguere tra approcci predittivi e generativi rischia di generare confusione. Oggi l'IA predittiva dispone di una base scientifica più solida

e di percorsi di validazione più maturi. L'IA generativa, invece, mostra enormi potenzialità ma necessita ancora di robuste evidenze cliniche, maggiore trasparenza e standard condivisi di valutazione.

Per rendere questi strumenti realmente affidabili e integrati nella pratica clinica saranno necessari almeno cinque interventi: studi prospettici e randomizzati su esiti clinici reali; maggiore trasparenza sui dati di addestramento e sugli aggiornamenti dei modelli; sistemi di monitoraggio continuo delle prestazioni e degli errori; integrazione con fonti scientifiche verificabili e aggiornate; mantenimento della supervisione umana nelle decisioni cliniche ad alto impatto. Alcuni di questi interventi sono parte integrante della normativa disponibile a livello europeo (tra cui l'AI Act, la legge italiana per l'intelligenza artificiale, il regolamento europeo sui dispositivi medici). Gli altri dovrebbero essere argomenti sui quali gli Ordini dei Medici provinciali, le associazioni medico-scientifiche nazionali, le istituzioni sanitarie dovrebbero esprimersi e prendere posizione.

L'obiettivo ultimo, in definitiva, non dovrebbe essere sostituire il professionista sanitario, ma sviluppare strumenti che amplifichino le sue capacità mantenendo al centro sicurezza, responsabilità e qualità delle cure.

Le opinioni espresse in questo articolo dal Dott. Eugenio Santoro sono esclusivamente personali e non riflettono necessariamente la posizione ufficiale dell'Istituto di Ricerche Farmacologiche Mario Negri IRCCS.