Medicina e chatbot: perché milioni di persone chiedono consigli sanitari all'IA (e perché può essere rischioso)

Sempre più cittadini e professionisti sanitari utilizzano i modelli linguistici di intelligenza artificiale per interpretare sintomi, analizzare referti o decidere se rivolgersi al Pronto Soccorso. Ma studi recenti mostrano che questi strumenti, pur eccellendo nei test teorici, possono fallire proprio nelle situazioni cliniche più critiche

(Fonte: https://www.corriere.it/ 10 marzo 2026)

Negli ultimi anni, **i modelli linguistici di grandi dimensioni (LLM)** sono diventati uno dei principali punti di accesso alle informazioni sanitarie. **Milioni di persone utilizzano chatbot con IA generativa, come ChatGPT,** per interpretare sintomi, capire il significato di esami medici o decidere se rivolgersi al **Pronto Soccorso**. Anzi OpenAI sta lanciando ChatGPT Salute «un'esperienza dedicata che integra in modo sicuro le informazioni sanitarie con l'intelligenza di ChatGPT, per aiutare a sentirsi più informati, preparati e sicuri nella gestione della propria salute», come recita il claim online.

Questa tendenza è favorita da due fattori: **la disponibilità immediata delle risposte e la difficoltà di accesso ai servizi sanitari in molti Paesi**. Per molti utenti, parlare con un chatbot rappresenta ormai il primo passo prima di contattare un medico in carne e ossa.

La fiducia nell'IA cresce in tutto il mondo

Un segnale di questa trasformazione arriva dallo studio pubblicato su ***AI & Society*** («Who lets AI take over? Cross-national variations in willingness to delegate socially important roles to artificial

intelligence»). La ricerca, condotta dalla Bournemouth University su circa **31.000 adulti in 35 paesi**, ha analizzato quanto le persone siano disposte a delegare all'intelligenza artificiale ruoli socialmente importanti.

I risultati mostrano un livello sorprendentemente alto di fiducia: **il 45% degli intervistati a livello globale** ha dichiarato che si affiderebbe all'IA per svolgere il ruolo del proprio medico; nel Regno Unito **il 41%** sarebbe disposto a utilizzare l'intelligenza artificiale per servizi di consulenza psicologica; oltre **tre quarti degli intervistati** parlerebbero con un chatbot come se fosse un compagno o un amico; circa **un quarto degli adulti britannici** affiderebbe all'IA il compito di insegnare ai propri figli.

Secondo la psicologa **Ala Yankouskaya**, responsabile dello studio, «con il rapido sviluppo e la massiccia diffusione dell'intelligenza artificiale sempre più persone ripongono fiducia in questi strumenti e iniziano a considerarli in ruoli sempre più importanti della vita quotidiana».

Perché sempre più persone si rivolgono ai chatbot

La crescente fiducia nell'IA è legata anche a **problemi strutturali dei sistemi sanitari.** In molti paesi ottenere un appuntamento con uno specialista o con uno psicologo può richiedere settimane o mesi.

In questo contesto **la possibilità di ottenere risposte immediate da un chatbot diventa estremamente attraente**. «Se qualcuno soffre di depressione, non vuole aspettare mesi per un appuntamento, quindi può rivolgersi all'intelligenza artificiale», ha osservato Yankouskaya. Tuttavia, ha aggiunto, «**quando ho testato personalmente alcuni strumenti ho trovato il linguaggio molto vago e confuso**. Gli sviluppatori evitano di entrare nei dettagli delle diagnosi, quindi questi sistemi non possono sostituire il colloquio con un professionista sanitario».

Quando la fiducia diventa un boomerang

Gli strumenti di IA generativa sono inoltre **progettati per adattare il tono delle risposte a quello dell'utente,** creando una conversazione empatica e non giudicante. Questa caratteristica può aumentare il senso di fiducia e di sicurezza percepita.

Fin troppo, anche. **I genitori di Adam Raine, 16 anni,** hanno intentato la prima causa per omicidio colposo contro OpenAI, sostenendo che ChatGPT abbia rafforzato i pensieri suicidari del figlio e fornito informazioni utili alla pianificazione del gesto. **Per la famiglia di Jonathan Gavalas, 36 anni,** l'uomo si sarebbe tolto la vita su istigazione di Gemini, il chatbot di Google, per non essere riuscito a portare a termine una serie di missioni reali assegnate dal chatbot. Così ha denunciato Google per omicidio colposo. In Florida, nel 2024, **i genitori di un quattordicenne** hanno citato in giudizio Character.AI accusando la piattaforma di aver favorito il suicidio del figlio dopo settimane di conversazioni digitali. Già nel 2022, **in Belgio, un uomo aveva perso la vita dopo lunghe interazioni con l'app Chai,** basata sul modello «Eliza», che secondo la moglie aveva alimentato pensieri deliranti e autolesivi.

Il test clinico su ChatGPT Health

Mentre cresce la fiducia del pubblico, **gli studi scientifici iniziano a valutare la sicurezza di questi strumenti**. Uno dei primi lavori indipendenti su questo tema è stato da poco pubblicato su ***Nature Medicine*** («ChatGPT Health performance in a structured test of triage recommendations»).

Si tratta di un vero «crash test». I ricercatori della Icahn School of Medicine at Mount Sinai hanno creato **60 scenari clinici realistici** che coprivano **21 specialità mediche**, dalle condizioni lievi alle emergenze. Tre medici indipendenti hanno stabilito il livello corretto di urgenza per ogni caso utilizzando le linee guida di **56 società scientifiche**.

Gli scenari sono stati poi testati in **16 varianti diverse**, per un totale di **960 interazioni con il sistema di IA**. L'obiettivo era verificare se un chatbot sanitario fosse in grado di dire chiaramente agli utenti quando era necessario recarsi al Pronto Soccorso.

Emergenze mancate e segnali ignorati

I risultati dello studio hanno mostrato un quadro contrastante. **Il sistema riconosceva correttamente molte emergenze «da manuale»**, come ictus o gravi reazioni allergiche. Tuttavia nelle situazioni più ambigue — che nella pratica clinica sono molto comuni — le prestazioni peggioravano.

In **oltre la metà dei casi che richiedevano cure di emergenza**, il sistema suggeriva di restare a casa o di prenotare una visita medica.

Secondo l'autore principale dello studio, **Ashwin Ramaswamy**, questo era proprio il punto centrale della ricerca: «Volevamo rispondere a una domanda molto semplice ma fondamentale: se qualcuno sta vivendo una vera emergenza medica e si rivolge a ChatGPT Health per chiedere aiuto, il sistema gli dirà chiaramente di andare al pronto soccorso?».

I ricercatori hanno anche osservato che il chatbot spesso **riconosceva i segnali di pericolo nelle proprie spiegazioni**, ma continuava comunque a rassicurare l'utente.

«Il sistema funzionava bene nelle emergenze più evidenti, come ictus o gravi reazioni allergiche», ha spiegato Ramaswamy. «Ma **aveva difficoltà nelle situazioni più sfumate**, dove il rischio non è immediatamente evidente. E sono proprio questi i casi in cui il giudizio clinico è più importante».

Il rischio anche per i medici

La crescente diffusione degli LLM nella pratica clinica solleva quindi un problema meno discusso: **il rischio che anche i professionisti sanitari sviluppino una fiducia eccessiva** nelle risposte generate dall'intelligenza artificiale.

Secondo **Girish Nadkarni**, direttore del Windreich Department of Artificial Intelligence and Human Health del Mount Sinai, uno dei risultati più preoccupanti dello studio riguarda proprio **il modo in cui il sistema gestisce le situazioni di rischio suicidario**.

«Ci aspettavamo una certa variabilità», ha spiegato Nadkarni. «Ma ciò che abbiamo osservato

andava oltre una semplice incoerenza. **Gli avvisi del sistema erano quasi invertiti rispetto al rischio clinico,** comparendo più spesso negli scenari a rischio minore che nei casi in cui una persona spiegava esattamente come intendeva farsi del male».
Secondo il ricercatore, nella pratica clinica reale questo tipo di errore potrebbe avere conseguenze molto gravi.

Le difficoltà dell'interazione tra utenti e IA

Ulteriori elementi di complessità emergono dallo studio pubblicato su **Nature Medicine** («Reliability of LLMs as medical assistants for the general public: a randomized preregistered study»*).* I ricercatori hanno analizzato **come le persone utilizzano concretamente i modelli linguistici per interpretare sintomi e decidere se cercare assistenza medica**. Nel trial randomizzato, che ha coinvolto oltre 1.200 partecipanti, i volontari dovevano valutare una serie di vignette cliniche utilizzando diversi strumenti informativi, tra cui modelli linguistici e ricerche online tradizionali. I risultati hanno mostrato un fenomeno inatteso: mentre **i modelli linguistici valutati da soli** mostravano una buona capacità di identificare correttamente le condizioni cliniche, **gli utenti che utilizzavano l'IA non prendevano decisioni più accurate** rispetto a chi utilizzava una normale ricerca su Internet. Secondo gli autori, il problema non è tanto la conoscenza medica dei modelli quanto **la difficoltà degli utenti nel formulare le domande corrette e interpretare le risposte generate dall'IA**, un limite che può ridurre significativamente il potenziale beneficio di questi strumenti nella pratica quotidiana.

L'accuratezza dei modelli ChatGPT nel triage sanitario

Un'altra analisi, pubblicata su ***Communications Medicine*** («Evaluating the accuracy of ChatGPT model versions for giving care-seeking advice»), ha esaminato in modo sistematico **22 versioni diverse di ChatGPT** utilizzando **45 casi clinici realistici**. Ogni caso è stato testato **dieci volte**, per un totale di **9.900 valutazioni**. I risultati mostrano che l'accuratezza complessiva dei modelli rimane **intorno al 70%**, con il modello più performante che raggiunge **il 74% di raccomandazioni corrette**. Tuttavia i ricercatori non hanno osservato miglioramenti significativi nelle versioni più recenti dei modelli.
Lo studio evidenzia inoltre una tendenza sistematica dei chatbot a **raccomandare cure più urgenti del necessario**, un fenomeno noto come *overtriage*. Questo comportamento **riduce il rischio di sottovalutare emergenze**, **ma può anche portare a un aumento delle visite inutili in ospedale.** I modelli hanno mostrato le maggiori difficoltà nell'identificare correttamente i casi in cui **l'autogestione dei sintomi (self-care)** sarebbe stata sufficiente.
Un altro aspetto rilevante riguarda la **variabilità delle risposte**: lo stesso modello può fornire raccomandazioni diverse quando viene interrogato più volte con lo stesso caso clinico, segno che le risposte generate non sono sempre stabili.

Il confronto con medici e pazienti

Uno studio pubblicato su **The Lancet Digital Health** («The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study») **ha confrontato direttamente le prestazioni dell'intelligenza artificiale con quelle di medici e cittadini.** Il modello GPT-3 ha dimostrato una buona capacità di identificare le possibili diagnosi e di suggerire quando cercare assistenza medica. Tuttavia le sue prestazioni restavano inferiori a quelle dei medici, che continuavano a mostrare una maggiore accuratezza sia nella diagnosi sia nelle decisioni di triage.

Una tecnologia potente ma da usare con cautela

In conclusione: l'intelligenza artificiale potrebbe diventare uno strumento prezioso per la medicina del futuro, ma solo se utilizzata con consapevolezza dei suoi limiti. Il medico e informatico **Isaac Kohane**, della Harvard Medical School, ha osservato che l'uso crescente di questi strumenti da parte dei pazienti rende necessario un controllo molto più rigoroso.

«Gli LLM sono diventati il primo punto di riferimento dei pazienti per i consigli medici», ha dichiarato. «Ma sono meno sicuri proprio agli estremi clinici, dove il giudizio distingue tra emergenze mancate e allarmi inutili».

Per questo motivo, ha aggiunto, «quando milioni di persone utilizzano un sistema di intelligenza artificiale per decidere se hanno bisogno di cure di emergenza, **la posta in gioco è estremamente alta. Le valutazioni indipendenti dovrebbero essere la norma, non un'eccezione**».

ChatGpt (e altre AI) sempre più utilizzate per «consulti» sanitari anche in Italia