

Dalla letteratura

ChatGPT batte il medico dieci a nove?

In pochi altri campi come in medicina, la disponibilità di ChatGPT e di altri large language model (LLM) ha innescato una sorta di competizione tra questi strumenti di intelligenza artificiale e il ragionamento umano. Sulle riviste scientifiche vengono pubblicati sempre più spesso i risultati di questi duelli: uno tra i più recenti è quello che ha coinvolto medici e specializzandi del Beth Israel Deaconess Medical Center, uno dei più famosi ospedali di Boston, collegato alla facoltà di medicina di Harvard¹.

Una research letter ha esposto metodi e risultati di uno studio che ha confrontato ChatGPT-4 con le prestazioni dei medici utilizzando standard di valutazione solitamente utilizzati per misurare le decisioni cliniche. Intervistato da MedPage Today, Adam Rodman - internista dell'ospedale di Boston e tra gli autori dello studio - ha sottolineato come la medicina sia costantemente alla ricerca di modi per migliorare il processo decisionale clinico, dato che «le diagnosi errate possono causare fino a 800.000 morti ogni anno negli Stati Uniti»². I LLM come il GPT-4 «sono uno degli interventi sanitari più interessanti degli ultimi 50 anni [...] anche perché non soffrono dei bias cognitivi di cui sono vittima gli esseri umani». La grande potenzialità dei LLM è dunque soprattutto nella capacità di cambiare idea, di modificare i

propri giudizi alla luce dell'esperienza e dei dati della ricerca.

I ricercatori hanno utilizzato r-IDEA, uno strumento sviluppato e validato per valutare il ragionamento clinico dei medici³. I ricercatori hanno reclutato 21 medici e 18 specializzandi, ciascuno dei quali ha lavorato su uno dei 20 casi clinici selezionati costituiti da quattro fasi sequenziali di ragionamento diagnostico. Gli autori hanno incaricato i medici di scrivere e giustificare le loro diagnosi differenziali in ogni fase. ChatGPT-4 ha ricevuto un messaggio con istruzioni identiche e ha eseguito tutti i 20 casi clinici. Alle loro risposte è stato poi assegnato un punteggio per il ragionamento clinico (punteggio r-IDEA). La ricerca ha dato risultati in certa misura sorprendenti, perché lo strumento di intelligenza artificiale ha dimostrato di poter "ragionare" in modo uguale o addirittura migliore del medico. Il LLM ha ottenuto i punteggi r-IDEA più alti, con un punteggio medio di 10 su 10, rispetto al 9 per i medici e alla media di 8 per gli specializzandi.

Sono necessari ulteriori studi per determinare come i LLM possano essere integrati al meglio nella pratica clinica, ma anche adesso potrebbero essere utili come riferimento di controllo. Ottimisticamente, Rodman ha dichiarato che se «i primi studi suggerivano che l'intelligenza artificiale potesse fare diagnosi se tutte le informazioni le fossero state fornite, quello che il nostro studio ha mostrato è che l'in-

telligenza artificiale può realmente *ragionare*, forse addirittura in modo migliore rispetto alle persone, attraverso le diverse fasi del processo. Abbiamo un'opportunità unica per migliorare la qualità e l'esperienza dell'assistenza sanitaria per i pazienti».

Bibliografia

1. Cabral S, Restrepo D, Kanjee Z, et al. Clinical reasoning of a generative artificial intelligence model compared with physicians. *JAMA Intern Med* 2024; 184: 581-3.
2. DePeau-Wilson M. ChatBot beat doctors in clinical reasoning. *MedPage Today* 2024; 1 aprile.
3. Schaye V, Miller L, Kudlowitz D, et al. Development of a clinical reasoning documentation assessment tool for resident and fellow admission notes: a shared mental model for feedback. *J Gen Intern Med* 2022; 37: 507-12.

I bambini e ChatGPT

Quasi il 95% degli adolescenti tra i 13 e i 17 anni utilizza i social media. Un rapporto pubblicato negli Stati Uniti nel 2023 ha sottolineato soprattutto gli aspetti preoccupanti di questa esposizione dei giovani al digitale, sottolineando i rischi per la salute mentale legati a un'elevata esposizione¹. Fortunatamente, nonostante sottolinei i potenziali effetti negativi, il rapporto riconosce anche le potenzialità offerte



Le immagini utilizzate per questa rubrica sono tratte da una ricerca su iStock con parola chiave "intelligenza artificiale".

dai social media per la socializzazione e la sollecitazione a una presenza online che sviluppi la creatività degli adolescenti. Poca attenzione è stata data, però, a quelle che ancora per poco possiamo definire le nuove frontiere dell'intelligenza artificiale, i large language model e il più conosciuto tra questi: ChatGPT.

Una viewpoint pubblicata sul *JAMA Pediatrics* sottolinea che la capacità dei chatbot di fornire informazioni sensibili ai giovani utenti è stata messa in discussione, portando a un'indagine dell'Information Commissioner's Office nel Regno Unito¹. I rischi – almeno potenziali – di questo “dialogo digitale” sono noti mentre mancano prove concrete sugli effetti sociali.

«Anche nelle aule scolastiche – scrivono gli autori dell'articolo uscito sul *JAMA Pediatrics* – si sta vedendo un aumento dell'utilizzo di assistenti virtuali basati sull'IA come ChatGPT di OpenAI. Mentre gli studenti riconoscono i benefici di tali strumenti nel facilitare i compiti scolastici, è evidente la necessità di una supervisione e di un equilibrio nell'uso di tali tecnologie per evitare di minare lo sviluppo delle abilità di risoluzione dei problemi».

La visione prudente dei tre autori – due psichiatri della Yale University e

un terzo collega della Cleveland Clinic – li porta a sottolineare come l'uso di assistenti virtuali basati sull'IA si stia estendendo anche nel campo clinico, con potenziali applicazioni terapeutiche per i bambini: «tuttavia è importante considerare i rischi associati all'uso di tali tecnologie, compresi possibili errori e bias nei risultati». È fondamentale – proseguono – «valutare attentamente l'impatto di tali interazioni sui bambini e garantire un equilibrio tra l'uso delle tecnologie e la promozione di relazioni umane significative. I medici e i ricercatori devono continuare a studiare gli effetti dell'IA sui bambini per fornire linee guida e informazioni precise sulla gestione di questa tecnologia in crescita».

Bibliografia

1. Department of Health and human services. Social media and youth mental health: the US Surgeon General's advisory. Maggio 2023. Disponibile su: <https://lc.cx/tOmz5u> [ultimo accesso 7 maggio 2024].
2. Pratt N, Madhavan R, Weleff J. Digital dialogue. How youth are interacting with chatbots. *JAMA Pediatr* 2024; 178: 429-30.

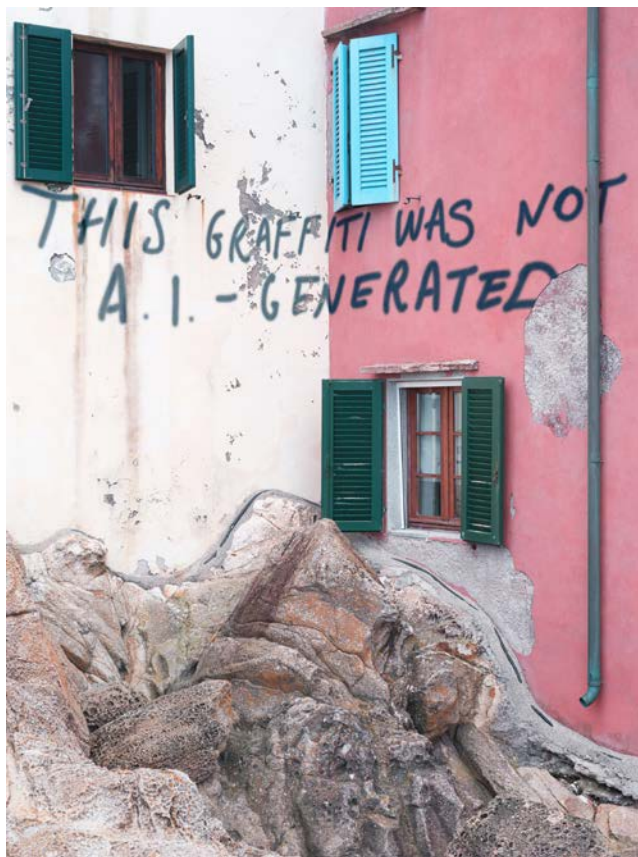
Intelligenza artificiale e cure primarie

Le applicazioni dell'intelligenza artificiale (IA) nella pratica clinica sono ormai più di una prospettiva futuribile, ma sono tutte “augurabili”? E quali sono le sfide da affrontare nell'assistenza primaria? Nell'articolo “Using artificial intelligence to improve primary care for patients and clinicians” di Urmimala Sarkar e David Bates, pubblicato sul *JAMA Internal Medicine*¹, gli autori discutono di come l'IA possa essere potenzialmente e vantaggiosamente utilizzata per compiti come l'elaborazione della posta in arrivo, la generazione di documentazione medica, l'assistenza ai pazienti tra una visita e l'altra e il supporto alle diagnosi e alle terapie.

«Mentre il potenziale per migliorare la qualità e la sicurezza dell'assistenza sanitaria in generale e dell'assistenza primaria in particolare utilizzando l'intelligenza artificiale è stato discusso per oltre due decenni, il salto tecnologico verso la disponibilità diffusa dell'intelligenza artificiale suggerisce che l'uso influenzerà presto la pratica dell'assistenza primaria su base giornaliera, se non oraria. [...] Se applicata ampiamente alle sfide dell'assistenza primaria, l'intelligenza artificiale ha il potenziale non solo per migliorare in modo incrementale l'assistenza primaria, ma anche di trasformarla. Riteniamo che l'intelligenza artificiale possa aiutare in molte delle sfide quotidiane affrontate dai medici di base nel tentativo di fornire cure di alta qualità».

Alcuni si sono chiesti se l'uso dell'IA, in particolare il suo utilizzo nella risposta ai messaggi dei pazienti, ridimensionerà ulteriormente la relazione paziente-medico. L'automazione di aspetti selezionati del lavoro di assistenza primaria, sostengono gli autori, libera tempo per la costruzione di relazioni e la cura che sono elementi essenziali. L'implementazione di strumenti di IA all'avanguardia richiede un monitoraggio rigoroso della sicurezza e dell'utilità, con aggiustamenti ripetuti che includano il contributo dei team di assistenza in prima linea, dei pazienti e delle famiglie. Inoltre, l'implementazione dell'IA, aggiungono, richiede una solida supervisione etica per evitare di perpetuare o peggiorare le disuguaglianze nell'erogazione dell'assistenza sanitaria.

«L'attuale pratica dell'assistenza primaria non è ottimale per i pazienti e non è più fattibile per i medici, e l'intelligenza artificiale ne migliorerà i risultati, l'efficienza e, si spera, la sostenibilità come percorso di carriera. Queste soluzioni di intelligenza artifi-



ciali offrono grandi vantaggi potenziali, ma solo se sviluppate attentamente tenendo conto delle esigenze dei medici in prima linea».

Nell'editoriale di commento all'articolo, pubblicato sullo stesso numero della rivista e intitolato "The perils of artificial intelligence in a clinical landscape" di Isabel Ostrer e Louise Aronson², sono puntualizzati alcuni spunti di riflessione ulteriori sulle prospettive di impiego della IA e in medicina e nei contesti di assistenza primaria in particolare. Innanzitutto, le promesse dell'IA sono indubbiamente allettanti, ma questa deve essere ben integrata nel contesto clinico per evitare di danneggiare il rapporto medico-paziente. In precedenza sono stati fatti molti sforzi per adottare nuove tecnologie nel panorama medico ma, purtroppo, questo spesso è stato fatto senza attenzione e senza un'adeguata pianificazione.

«Ciò può essere ottenuto in diversi modi. Alcuni aspetti che vale la pena considerare quando si sviluppa l'intelligenza artificiale per l'assistenza primaria comprendono la necessità di capire la lezione dai fallimenti dell'integrazione delle cartelle cliniche elettroniche, ottimizzare le cinque funzioni chiave dell'assistenza primaria identificate dall'Organizzazione mondiale della sanità e spendere risorse per miglioramenti comprovati almeno tanto quanto si spende per quelli non provati offerti dall'intelligenza artificiale», concludono gli autori.

Bibliografia

1. Sarkar U, Bates DW. Using artificial intelligence to improve primary care for patients and clinicians. *JAMA Intern Med* 2024; 184: 343-4.
2. Ostrer I, Aronson L. The perils of artificial intelligence in a clinical landscape. *JAMA Intern Med* 2024; 184: 351-2.

Alessio Malta,
in collaborazione con la Biblioteca Medica
Virtuale della Provincia di Bolzano

Intelligenza artificiale più utile: serve la Politica

L'intervento dei due studiosi della Johns Hopkins nello Health Forum del *JAMA* va diritto al punto: per una diagnosi medica non basta che un sistema di intelligenza artificiale (IA) si avvicini alle capacità cognitive di un non addetto ai lavori, né è sufficiente che abbia le stesse capacità di un medico "medio", considerato l'attuale tasso di errori

diagnostici nella pratica clinica. Serve qualcosa di più e per ottenerlo occorre che la Politica - in concreto, le istituzioni e le agenzie sanitarie - metta l'IA nelle migliori condizioni per supportare il personale sanitario¹.

«Anche se l'IA, il machine learning e i large language model (LLM) sono maturati al punto che le tecnologie stesse non costituiscono più un ostacolo al successo - scrivono gli autori - resta un punto critico, in particolare la mancanza di fonti di dati adeguate, necessarie per addestrare i sistemi di IA a raggiungere l'eccellenza nella diagnosi. Se i sistemi di IA vengono addestrati su dati errati, produrranno risultati errati. I sistemi di IA che apprendono su dati errati commetteranno generalmente gli stessi errori che commettono gli esseri umani, se non di più. Ad esempio, se i dati delle cartelle cliniche elettroniche vengono utilizzati per addestrare i sistemi di IA, questi ultimi riprodurranno (come minimo) gli attuali errori diagnostici o rispecchieranno i bias di cui sono portatori gli operatori; nel peggiore dei casi, i sistemi di IA non solo sbaglieranno spesso nelle loro raccomandazioni diagnostiche, ma comprometteranno anche la formazione clinica. Se i sistemi diagnostici basati sull'IA vengono impiegati senza un'adeguata verifica e monitoraggio dopo l'impiego, è prevedibile che la qualità della diagnosi medica diminuisca e che i pazienti ne risentano».

Che fare, allora? Serve che i National Institutes of Health, l'Agency for Healthcare Research and Quality e altre istituzioni investano nella costruzione di banche dati specifiche per il machine learning dei sistemi di IA finalizzati alla diagnosi. Alcuni database utili alla diagnosi per immagine sono già in fase di sviluppo: in radiologia, oftalmologia, patologia e dermatologia. Ma

non esistono cose del genere in altri ambiti disciplinari. «Lo sviluppo di sistemi di IA accurati dal punto di vista diagnostico richiede dati di alta qualità sia nella fase iniziale (dati demografici del paziente, sintomi, segni e risultati di laboratorio e radiografici) sia nella fase finale (diagnosi definitive, effetti del trattamento e outcome dell'assistenza)».

Gli autori sollecitano anche la Food and Drug Administration a impegnarsi a definire gli standard per i database completi e accurati necessari per formare i sistemi diagnostici di IA. «L'obiettivo dovrebbe essere quello di sfruttare i punti di forza complementari del medico e del computer per ottenere diagnosi migliori di quelle ottenute da uno solo dei due. In assenza, i rischi dell'IA per la diagnosi potrebbero oscurare i benefici, portando a esiti clinici peggiori, a cure non appropriate e a personale sanitario meno qualificato».

Bibliografia

1. Newman-Toker DE, Sharfstein JM. The role for policy in AI-assisted medical diagnosis. *JAMA Health Forum* 2024; 5: e241339.

Sistemi informatizzati: migliorano le decisioni cliniche?

I medici di medicina generale svolgono un ruolo fondamentale nella gestione della malattia renale cronica, identificando i pazienti a rischio e trattando i fattori di rischio, come l'ipertensione e il diabete. L'invio tempestivo dei pazienti alle cure nefrologiche è associato





The business of publishing and of getting published have corrupted the evidence needed for care. It is missing. It is late. It is hyped. It is noisy. Luca eloquently calls for urgent reform to abolish this travesty of science.

Victor Montori

Knowledge and Evaluation Research (KER) Unit, Mayo Clinic, Rochester

In this new book, Luca De Fiore, a longtime proponent of precision in medical evidence, casts his critical eye on academic medical publishing, and disdains what he sees.

Jerome P. Kassirer

Distinguished Professor, Tufts University School of Medicine, Editor-in-Chief Emeritus, New England Journal of Medicine

There are lots of things wrong with medical publishing. But, fortunately, some things are right, like this book by Luca De Fiore.

Steven Woloshin

Center for Medicine and media, Dartmouth University, Lisa Schwartz Foundation



Sul pubblicare in medicina

Impact factor, open access, peer review, predatory journal e altre creature misteriose

Un libro di **Luca De Fiore**
[Il Pensiero Scientifico Editore]

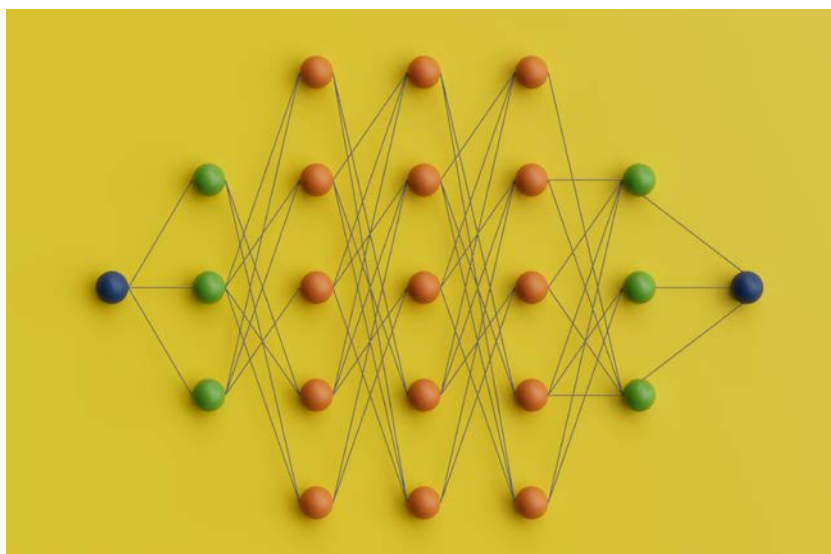
Presentazione di **Richard Smith**
[Former Editor-in-Chief, The BMJ]

Un libro di 192 pagine. € 18,00

a benefici in termini di costi, morbilità e mortalità, ma l'accesso ai servizi ambulatoriali non sempre corrisponde al fabbisogno a livello di popolazione. Alcuni studi hanno dimostrato che i medici di base non sono sempre a conoscenza delle linee guida per la gestione della malattia renale cronica (CKD) e si trovano di fronte a barriere che ne impediscono l'applicazione¹. I sistemi informatici di supporto alle decisioni cliniche mirano a colmare questa lacuna fornendo ai medici delle cure primarie informazioni specifiche sulla persona e raccomandazioni tempestive basate sull'evidenza. Lo scopo di uno studio clinico randomizzato i cui risultati sono stati appena pubblicati sul *JAMA Internal Medicine*² è stato verificare se l'utilizzo da parte del medico di medicina generale di un sistema di supporto informatizzato possa portare a una riduzione della pressione arteriosa sistolica (SBP) rispetto alla cura abituale, senza supporto informatizzato.

I ricercatori hanno coinvolto medici di cure primarie di 15 centri ospedalieri, ambulatoriali e centri assistenziali territoriali e li hanno assegnati in modo casuale, con un approccio stratificato e a coppie, all'intervento o all'assistenza abituale. L'intervento consisteva nell'utilizzo di un sistema di supporto alle decisioni cliniche basato su principi di economia comportamentale e progettato secondo i principi dello human-centered design, che forniva raccomandazioni personalizzate e basate sull'evidenza, tra cui l'inizio della terapia o l'aggiustamento del dosaggio di inibitori del sistema renina-angiotensina-aldosterone. Nel gruppo di controllo, il sistema di supporto alle decisioni cliniche non era attivato.

Tutti i pazienti di età ≥ 18 anni che hanno effettuato una visita presso un



medico di medicina generale in uno dei setting coinvolti nella ricerca nei 2 anni precedenti la prima visita durante il periodo di intervento erano idonei all'inclusione. I ricercatori hanno arruolato coloro che avevano una visita programmata una volta iniziato il periodo di studio e che soddisfacevano i criteri per malattia renale cronica di stadio 3 o 4 e per ipertensione non controllata. L'esito primario era la variazione della SBP media tra il basale e i 180 giorni in ciascun gruppo di studio. Gli esiti secondari includevano il controllo della pressione arteriosa (BP) e outcome quali la percentuale di pazienti che ricevevano un intervento in linea con le raccomandazioni fornite dal sistema informatizzato di supporto alle decisioni cliniche.

In totale, lo studio ha incluso 174 medici di base e 2026 pazienti, prevalentemente di sesso femminile (60,4%) con un'età media di 75,3 anni:

87 medici e 1029 pazienti sono stati assegnati all'intervento e 87 medici e 997 pazienti alla terapia standard. Complessivamente, 1714 (84,6%) pazienti erano in trattamento per l'ipertensione al basale. I pazienti con misurazione della SBP a 180 giorni sono stati 1623 (80,1%), di cui 815 nel gruppo di intervento e 808 nel gruppo di cura abituale. È stata notata una differenza significativa nella variazione della SBP tra i gruppi, con una variazione media della SBP di -14,6 (95% CI, da -13,1 a -16,0) mm Hg nel gruppo di intervento rispetto a -11,7 (95% CI, da -10,2 a -13,1) mm Hg nel gruppo di cura abituale ($p=,005$).

Sebbene non vi sia stata alcuna differenza significativa nella percentuale di pazienti che hanno raggiunto il controllo della BP nel gruppo di intervento (50,4%) rispetto al gruppo di controllo (47,1%), un maggior numero di pazienti tra quelli che hanno ricevuto l'intervento ha beneficiato di una terapia in linea con le raccomandazioni del supporto decisionale informatizzato. La percentuale di pazienti che hanno ricevuto un ordine per un ACE, un ARB o un diuretico tiazidico è stata significativamente maggiore nel braccio di intervento rispetto al gruppo di cura abituale ($p<,001$).

Bibliografia

1. Israni RK, Shea JA, Joffe MM, Feldman HI. Physician characteristics and knowledge of CKD management. *Am J Kidney Dis* 2009; 54: 238-47.
2. Samal L, Kilgallon JL, Lipsitz S, et al. Clinical decision support for hypertension management in chronic kidney disease: a randomized clinical trial. *JAMA Intern Med* 2024; 184: 484-92.

