Dobbiamo fidarci dell'abilità diagnostica dell'intelligenza artificiale? - 3

Data 28 giugno 2026
Categoria Medicinadigitale

L'intelligenza artificiale non "indovina" la diagnosi e non possiede doti divinatorie. Tuttavia in questa serie di tre articoli, che avranno pubblicazione settimanale, mostriamo come, di fronte a un caso clinico complesso, possa aiutare il medico ad ampliare la diagnosi differenziale e a individuare gli accertamenti più appropriati.

In una pillola recente questa testata ha recensito un articolo del JAMA dal titolo provocatorio: "ChatGPT Health non convince JAMA" (https://www.pillole.org/public/aspnuke/news.asp?id=8983) in cui, tra l'altro, venivano avanzati dubbi sulle capacità dell'AI di elaborare corrette diagnosi differenziali.
Abbiamo voluto testare a nostra volta ChatGPT con un caso clinico sviluppato in una serie di tre pillole, pubblicate settimanalmente. Nelle prime due invitiamo i lettori a elaborare le possibili diagnosi differenziali. Nelle terza vedremo come "pensa" ChatGPT e sveleremo il mistero traendo anche alcune conclusioni.

Avvertenzaimportante: si tratta di un caso simulato, che non corrisponde a nessun paziente reale. Sebbene costruito per scopi didattici, riproduce una sequenza diagnostica realistica e clinicamente plausibile.

La prima parte è stata pubblicata domenica 14 giugno 2026 (https://www.pillole.org/public/aspnuke/news.asp?id=9106), la seconda domenica 21 giugno 2026
(https://www.pillole.org/public/aspnuke/news.asp?id=9107).

A questo punto abbiamo proposto il caso a ChatGPT. Questo il suo ragionamento:

Non è possibile stabilire una diagnosi definitiva in base agli elementi dati. Noto però che ogni specialista aveva spiegato il pezzo di sua competenza:
reumatologo → artrite sieronegativa;
ematologo → anemia da infiammazione;
neurologo → disturbo cognitivo lieve;
gastroenterologo → non patologie significative.

Ma nessuna di queste diagnosi spiegava tutto. Da queste considerazioni parte il mio primo ragionamento: probabilmente esiste una malattia sistemica unica che collega tutti i reperti.
A quel punto non penso più in termini di organi ma costruisco una lista mentale di categorie:
malattie infettive croniche;
malattieinfiltrative;
malattiegranulomatose;
vasculiti;
sindromilinfoproliferative;
malattie autoimmuni atipiche.

Da qui escono le prime ipotesi:

- Whipple: ipotesi che a mio avviso è la più probabile: esordisce quasi sempre con artralgie/artriti migranti sieronegative che precedono di anni i sintomi gastrointestinali. Successivamente compaiono calo ponderale, diarrea/alvo irregolare, linfoadenopatia mesenterica, anemia da disordine cronico e, nel 10-40% dei casi, manifestazioni neurologiche e cognitive (ansia, insonnia, deficit di memoria, alterazioni del SNC). L'uso di immunosoppressori (cortisone e metotrexato) per la presunta artrite può aver accelerato la progressione sistemica della malattia.

- Sarcoidosi: può causare artrite sieronegativa, linfoadenopatie (anche addominali), sintomi neurologici aspecifici (neurosarcoidosi) ed elevazione degli indici di flogosi; tuttavia la TC addome mostra solo "lieve" ingrossamento linfonodale e non ci sono menzioni di sintomi respiratori o di linfoadenopatie ilari polmonari (sebbene non sia stata prescritta una TC torace).

- Vasculite sistemica ANCA negativa: le vasculiti possono colpire contemporaneamente articolazioni, sistema nervoso (centrale o periferico) e l'apparato gastrointestinale, sostenendo un'anemia da infiammazione cronica. Però la puntura lombare e la RM encefalo negative per flogosi attiva riducono la probabilità di una neurovasculite franca.

- Linfoma intestinale primitivo o a localizzazione occulta a lenta evoluzione: può provocare infiammazione, calo ponderale, linfoadenopatia mesenterica e sintomi neurologici paraneoplastici. Contro: l'andamento è estremamente indolente per essere un linfoma non trattato nell'arco di 5 anni, e l'aspirato midollare è normale.

Ipotesi che ritengo meno probabili:

- Amiloidosi: potrebbe spiegare il coinvolgimento multisistemico, ma l'assenza di proteinuria significativa, un ecocardiogramma completamente normale (senza il classico ispessimento miocardico "brillante") e una progressione così lenta senza un danno d'organo florido la rendono un'ipotesi decisamente remota.

- Castleman: disordine linfoproliferativo raro che mima perfettamente una malattia autoimmune o un linfoma, causando PCR e VES alte, anemia da disordine cronico e linfoadenopatie. Tuttavia, di solito si accompagna a linfoadenopatie ben più vistose e diffuse (soprattutto mediastiniche o ascellari) rispetto a quel "lieve e aspecifico" ingrossamento dei linfonodi mesenterici visto alla TC.

- Endocardite subacuta: potrebbe spiegare l'infiammazione cronica, le artralgie migranti (da immunocomplessi) e i sintomi neurologici/vertigini (da microemboli). Però un ecocardiogramma completamente negativo e l'assenza di picchi febbrili classici tolgono quasi tutto il peso a questa ipotesi.

Il passo successivo è individuare degli esami ad alto rendimento: quale esame può discriminare tra le ipotesi principali? In ordine di importanza chiederei:

- Whipple: Gastroscopia con biopsie duodenali + PCR
- Sarcoidosi: PET-TC, ACE, biopsia
- Endocardite: Emocolture + ripetere ecocardiogramma
- Linfoma: PET-TC + biopsia
- Amiloidosi: Biopsia tessutale

Questa è stata la risposta di ChatGPT. Ma ora è arrivato il momento di svelare il mistero: il paziente consulta un secondo gastroenterologo che effettua una EGDS e ricerca di Tropheryma whipplei su biopsia e/o PCR. Viene posta diagnosi di malattia di Whipple.

Che cosa dimostra questo caso?

Non dimostra che ChatGPT sia infallibile. Non dimostra nemmeno che l'intelligenza artificiale sia in grado di sostituire il ragionamento clinico o l'esperienza di uno specialista. Dimostra però qualcosa di diverso e forse altrettanto interessante. Di fronte a un quadro complesso, in cui diversi specialisti avevano interpretato correttamente singoli aspetti del problema senza riuscire a ricondurli a una spiegazione unitaria, ChatGPT ha adottato un approccio diverso: ha considerato simultaneamente un numero molto elevato di ipotesi diagnostiche appartenenti a discipline differenti, le ha organizzate secondo un criterio di probabilità e ha individuato gli accertamenti più utili per escludere le diverse possibilità.
La diagnosi corretta non è emersa perché l'intelligenza artificiale possieda una sorta di intuizione clinica superiore, ma perché dispone di una base di conoscenze estremamente ampia che le consente di richiamare rapidamente anche patologie rare o poco frequenti che il singolo medico, comprensibilmente, può non avere presenti in quel momento.

Il vero valore aggiunto non è quindi l'individuazione della risposta giusta al primo tentativo, bensì la capacità di costruire una diagnosi differenziale estesa, mantenere aperte più piste contemporaneamente e suggerire gli esami con il più alto rendimentodiagnostico.
Naturalmente l'ultima parola spetta sempre al clinico, che deve interpretare il contesto, valutare il paziente e decidere quali ipotesi meritino realmente di essere esplorate.
Casi come questo suggeriscono che l'intelligenza artificiale potrebbe rappresentare un utile strumento di supporto soprattutto nelle situazioni più complesse, quando si può correre il rischio non tanto di formulare una diagnosi sbagliata, quanto interrompere prematuramente l'iter diagnostico.
Ogni specialista ha fatto benissimo il suo lavoro nel proprio settore: il reumatologo ha curato l'artrite, l'ematologo l'anemia, il neurologo il deficit cognitivo e il primo gastroenterologo ha escluso i tumori. Il problema è che la medicina moderna è così specializzata che i medici tendono a lavorare in "ambienti chiusi" che non vanno oltre il proprio campo di competenze.
Inoltre, subentra il bias di ancoraggio: una volta etichettato il paziente come "reumatico", tutti i sintomi successivi vengono visti come complicanze di quella prima diagnosi (ad esempio l'ematologo che dice "anemia da infiammazione cronica legata alla patologia reumatica").
L'AI non ha questi bias. Non si "innamora" della prima diagnosi. Guarda i dati in modo agnostico e simultaneo, collegando l'articolazione del reumatologo con il linfonodo del gastroenterologo e l'insonnia del neurologo.
Un medico, nel corso della sua carriera, può non incontrare mai un caso di Malattia di Whipple. La memoria umana è guidata dall'esperienza e dalla frequenza (la regola clinica: "Se senti rumore di zoccoli, pensa ai cavalli, non alle zebre"). Ma a volte, come in questo caso, si tratta proprio di una zebra. L'AI ha immagazzinato una mole enorme di dati derivanti dall'intera letteratura medica mondiale; per lei, una malattia rara ha la stessa dignità di memorizzazione (non di probabilità si badi bene, ma di archiviazione) di un'influenza. Non la dimentica in un cassetto. In altre parole impedisce l'interruzione prematura dell'iter diagnostico. L'AI eccelle nel dire al medico: "Sei sicuro di voler chiudere il caso? Guarda che se unisci questi tre puntini, potrebbe esserci anche quest'altra opzione".
Usare l'AI per ampliare la diagnosi differenziale nei casi complessi non è un segno di debolezza clinica, ma una forma di prudenza diagnostica - allo stesso modo in cui si consulta un collega o si cerca un riferimento in letteratura.

Notabene
Abbiamo proposto lo stesso caso anche ai concorrenti di ChatGPT (Claude, Google Gemini, Copilot) e abbiamo ottenuto risposte del tutto sovrapponibili.

RenatoRossi