# ChatGPT un po’ scarso nelle diagnosi differenziali

**Data** 05 novembre 2023
**Categoria** Medicinadigitale

Secondo una nuova ricerca le prestazioni del chatbot sono discrete in termini di accuratezza diagnostica complessiva ma scarse nelle diagnosi differenziali.

Una delle sfide principali dei sistemi come chatGPT è dimostrare la propria efficacia in ambiti complessi come la diagnostica.
I ricercatori dell'Innovation in Operations Research Center del Massachusetts General Hospital di Boston hanno addestrato ChatGPT su 36 scenari clinici pubblicati dal manuale clinico Merck Sharpe & Dohme
e ne hanno confrontato l'accuratezza su diagnosi differenziali, test diagnostici, diagnosi finale e gestione in base all'età del paziente, al sesso e alla gravità del caso.
I ricercatori hanno rilevato una accuratezza complessiva, nei 36 casi clinici, del 71,7%.
ChatGPT ha formulato possibili diagnosi e ha posto le diagnosi finali e le decisioni sulla gestione delle cure. Sono stati assegnati punteggi in cieco per le risposte corrette.
L’accuratezza del chatbot è stata valutata sulla diagnosi differenziale, i test diagnostici, la diagnosi finale e la gestione in un processo clinico strutturato.
ChatGPT si è dimostrato migliore nel fare una diagnosi finale, nella quale ha avuto una precisione del 77%.
Ha ottenuto prestazioni più basse nell'effettuare diagnosi differenziali, dove ha raggiunto un’accuratezza solo del 60%, e nelle decisioni di gestione clinica, con prestazioni inferiori al 68%.
La difficoltà di ChatGPT con la diagnosi differenziale non stupisce perché è un ambito nel quale i medici sono veramente esperti e apportano il massimo valore. Nelle fasi iniziali della cura del paziente il curante dispone di Informazioni spesso indistinte, variabili, dove prevalgono elementi soggettivi e qualità tacite, come intuizione, occhio clinico, empatia per interpretare il volto del paziente, il suo modo di parlare, la sequenza espositiva, il detto e il non detto, la sofferenza, i dettagli soggettivi. I dati osservati inoltre assumono significato diverso in funzione del contesto clinico, delle loro caratteristiche, della cronologia di comparsa, della frequenza nella malattia ipotizzabile. Ad esempio una epatomegalia associata a splenomegalia in bevitore orienta verso una cirrosi, in presenza di giugulari e fibrillazione verso uno scompenso cardiaco.
Prima che strumenti come ChatGPT possano essere presi in considerazione per l’integrazione nell’assistenza clinica, sono sicuramente necessari ulteriori miglioramenti.

**Giampaolo Collecchia, Riccardo De Gobbi, Roberto Fassina**

**Bibliografia**
Rao A et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res 2023;25:e48659) doi: 10.2196/48659

Per approfondire:
Collecchia G, De Gobbi R, Fassina R, Ressa G, Rossi RL: La Diagnosi Ritrovata. Il Pensiero Scientifico Editore
http://pensiero.it/catalogo/libri/professionisti/la-diagnosi-ritrovata

Collecchia G. De Gobbi R.: Intelligenza Artificiale e Medicina Digitale. Una guida critica. Il Pensiero Scientifico Ed. Roma 2020
http://pensiero.it/catalogo/libri/pubblico/intelligenza-artificiale-e-medicina-digitale

© 2004 - 2026 Pillole - Registrazione T. di Roma n. 2/06 | stampato il 13/11/2026 alle ore 21:53 | Pagina 1 di 1