

Rischio cardiaco: l'Intelligenza artificiale fornisce risultati incoerenti

Nonostante la capacità segnalata di ChatGpt di superare gli esami medici, una nuova ricerca indica che non sarebbe saggio fare affidamento su di esso per alcune valutazioni sanitarie, ad esempio se un paziente con dolore toracico deve essere ricoverato in ospedale

In uno studio che ha coinvolto migliaia di casi simulati di pazienti con dolore toracico, ChatGpt ha fornito conclusioni incoerenti, restituendo diversi livelli di valutazione del rischio cardiaco per gli stessi identici dati del paziente. Inoltre, il sistema di intelligenza artificiale generativa non è riuscito a eguagliare i metodi tradizionali utilizzati dai medici per giudicare il rischio cardiaco di un paziente. I risultati sono stati pubblicati sulla rivista Plos One.

“ChatGpt non agiva in modo coerente”, ha affermato l'autore principale, il dottor Thomas Heston, ricercatore presso l'Elson S. Floyd College of Medicine della Washington State University. “Forniti gli stessi identici dati, ChatGpt darebbe un punteggio di rischio basso, successivamente un rischio intermedio e, occasionalmente, arriverebbe a dare un rischio elevato”. Gli autori ritengono che il problema sia probabilmente dovuto al livello di casualità incorporato nell'attuale versione del software, ChatGpt4, che lo aiuta a variare le sue risposte per simulare il linguaggio naturale. Questa stessa casualità, tuttavia, non funziona bene per gli usi sanitari che richiedono una risposta unica e coerente, ha detto Heston. “Abbiamo scoperto che c'era molta variazione, e quella variazione nell'approccio può essere pericolosa - ha commentato. - Può essere uno strumento utile, ma penso che la tecnologia stia andando molto più veloce della

nostra comprensione, quindi è di fondamentale importanza fare molta ricerca, specialmente in queste situazioni cliniche ad alto rischio”.

I dolori toracici sono disturbi comuni nei pronto soccorso, che richiedono ai medici di valutare rapidamente l'urgenza delle condizioni di un paziente. Alcuni casi molto gravi sono facili da identificare dai loro sintomi, ma quelli a basso rischio possono essere più complicati, soprattutto quando si determina se qualcuno deve essere ricoverato in ospedale per osservazione o mandato a casa e ricevere cure ambulatoriali.

Attualmente, i professionisti medici utilizzano spesso una delle due misure che vanno sotto gli acronimi Timi e Heart per valutare il rischio cardiaco. Heston ha paragonato queste scale a calcolatrici, dove ciascuna utilizza una manciata di variabili tra cui sintomi, storia di salute ed età. Al contrario, una rete neurale AI come ChatGpt può valutare rapidamente miliardi di variabili, il che significa che potrebbe potenzialmente analizzare una situazione complessa in modo più rapido e approfondito.

Per questo studio, Heston e il collega Dr. Lawrence Lewis della Washington University di St. Louis hanno prima generato tre set di dati di 10.000 casi randomizzati e simulati ciascuno. Un set di dati conteneva le sette variabili della scala Timi, il secondo set includeva le cinque variabili della scala Heart e un terzo aveva

44 variabili di salute randomizzate.

Sui primi due set di dati, ChatGpt ha fornito una valutazione del rischio diversa dal 45% al 48% delle volte sui singoli casi rispetto a un punteggio Timi o Heart fisso. Per l'ultimo set di dati, i ricercatori hanno eseguito i casi quattro volte e hanno scoperto che ChatGpt spesso non era d'accordo con se stesso, restituendo diversi livelli di valutazione per gli stessi casi il 44% delle volte.

Nonostante i risultati negativi di questo studio, Heston vede un grande potenziale per l'IA generativa nell'assistenza sanitaria, con ulteriori sviluppi. Ad esempio, supponendo che gli standard di privacy possano essere soddisfatti, intere cartelle cliniche potrebbero essere caricate nel programma e, in un contesto di emergenza, un medico potrebbe chiedere a ChatGpt di fornire rapidamente i fatti più pertinenti su un paziente. Inoltre, per i casi difficili e complessi, i medici potrebbero chiedere al programma di generare diverse possibili diagnosi.

“ChatGpt potrebbe essere eccellente nel creare una diagnosi differenziale e questo è probabilmente uno dei suoi maggiori punti di forza”, ha affermato Heston.

E.T.

• Heston TF, Lewis LM. ChatGpt provides inconsistent risk-stratification of patients with atraumatic chest pain. *PlosOne* 2024. <https://doi.org/10.1371/journal.pone.0301854>