

12. ottobre

Diffidate di ChatGPT: a volte esagera

La fiducia è la sola cura conosciuta per la paura.

Lena Kellogg Sadler

La fiducia. Una porta aperta in una stanza senza muri.

vinkweb, Twitter

Il leone e il vitello giaceranno insieme.

Ma il vitello non dormirà molto.

Woody Allen

Un studio pubblicato oggi (8 ottobre di chi scrive) ha scoperto che se **ChatGPT** venisse utilizzato al pronto soccorso, potrebbe suggerire radiografie e antibiotici non necessari ad alcuni pazienti e ricoverarne altri che non necessitano di cure ospedaliere.



Nel report :

Williams CYK et al

**Evaluating the use of large language models
to provide clinical recommendations in the Emergency Department.**

Nat Commun. 2024 Oct 8;15(1):8236.

I ricercatori hanno affermato che, sebbene il modello possa essere sollecitato in modi che ne rendano le risposte più accurate, non è ancora in grado di reggere il giudizio clinico di un medico.



*"Questo è un messaggio prezioso per i medici, che non devono fidarsi ciecamente di questi modelli", ha affermato il post-dottorato **Chris Williams***

"ChatGPT può rispondere a domande di esami medici e aiutare a redigere note cliniche, ma al momento non è progettato per situazioni che richiedono più considerazioni, come quelle in un pronto soccorso".

Di recente, **Williams** ha dimostrato che ChatGPT, un modello linguistico di grandi dimensioni (LLM) che può essere utilizzato per la ricerca sulle applicazioni cliniche dell'intelligenza artificiale, è risultato leggermente più efficace degli esseri umani nel determinare quale tra due pazienti in pronto soccorso fosse in condizioni di salute più acute, una scelta semplice tra il paziente A e il paziente B.

Con lo studio attuale, Williams ha sfidato il modello AI a svolgere un compito più complesso: fornire le raccomandazioni che un medico fa dopo aver inizialmente esaminato un paziente al pronto soccorso. Ciò include decidere se ricoverare il paziente, sottoporlo a raggi X o altre scansioni o prescrivere antibiotici.

Per ciascuna delle tre decisioni, il team ha compilato un set di 1.000 visite al pronto soccorso da analizzare da un archivio di oltre 251.000 visite.

I set avevano lo stesso rapporto di risposte "sì" e "no" per le decisioni su ricovero, radiologia e antibiotici che si vedono nel pronto soccorso dell'UCSF Health.

Utilizzando la piattaforma AI generativa sicura dell'UCSF, che ha ampie protezioni della privacy, i ricercatori hanno inserito le note dei medici sui sintomi di ogni paziente e sui risultati degli esami in ChatGPT-3.5 e ChatGPT-4.

Quindi, hanno testato l'accuratezza di ogni set con una serie di prompt sempre più dettagliati. Nel complesso, i modelli AI tendevano a raccomandare i servizi più spesso del necessario. ChatGPT-4 era l'8% meno accurato dei medici residenti e ChatGPT-3.5 era il 24% meno accurato.

Williams sostiene che la tendenza dell'intelligenza artificiale a prescrivere troppi farmaci potrebbe essere dovuta al fatto che i modelli vengono addestrati su Internet, dove i siti di consulenza medica legittimi non sono progettati per rispondere a domande mediche di emergenza, ma piuttosto per indirizzare i lettori a un medico che può farlo.

"Questi modelli sono quasi calibrati per dire, 'chiedi consiglio medico', il che è abbastanza corretto da una prospettiva di sicurezza pubblica generale", ha detto. "Ma sbagliare dalla parte della cautela non è sempre appropriato nell'ambiente del pronto soccorso, dove interventi non necessari potrebbero causare danni ai pazienti, mettere a dura prova le risorse e portare a costi più elevati per i pazienti".

Ha affermato che modelli come ChatGPT avranno bisogno di framework migliori per valutare le informazioni cliniche prima di essere pronti per il pronto soccorso. Le persone che progettano tali framework dovranno trovare un equilibrio tra l'assicurarsi che l'IA non sfugga qualcosa di serio, e l'impedire che inneschi esami e spese non necessari.

Ciò significa che i ricercatori che sviluppano applicazioni mediche dell'intelligenza artificiale, insieme alla più ampia comunità clinica e al pubblico, devono valutare dove tracciare tali confini e quanto sia opportuno peccare di prudenza.

"Non esiste una soluzione perfetta", ha affermato, "ma sapendo che modelli come ChatGPT presentano queste tendenze, abbiamo il compito di riflettere su come vogliamo che funzionino nella pratica clinica".

Allen)