

Valutazione di casi clinici simulati in chirurgia toracica utilizzando ChatGPT

**Maurizio Cortale, Stefano Lovadina,
Alessia Arbore, Marina Troian**

Introduzione

L'intelligenza artificiale (AI) sta trasformando il panorama dell'educazione e della pratica medica. I recenti progressi nei modelli linguistici di grandi dimensioni, come ChatGPT, hanno dimostrato la capacità di questi sistemi di generare un testo coerente che imita accuratamente il linguaggio umano¹. Ciò ha portato a un crescente interesse per l'uso di queste tecnologie nell'insegnamento e nell'apprendimento clinico². In particolare, la simulazione di casi clinici complessi potrebbe offrire ai medici in formazione l'opportunità di rafforzare le loro capacità diagnostiche e terapeutiche in un ambiente virtuale sicuro e controllato. Tuttavia, sono necessarie ulteriori ricerche per valutare i vantaggi e gli svantaggi dell'integrazione dell'AI nella formazione medica.

In questo studio, abbiamo esplorato l'utilità di ChatGPT 4 nella simulazione di una serie di casi clinici in chirurgia toracica. Lo scopo era valutare quanto accuratamente ChatGPT potesse generare simulazioni di casi che rispecchiassero fedelmente gli scenari del mondo reale, nonché creare quiz mirati per testare le conoscenze.

Materiali e metodi

Sono stati creati casi clinici simulati semplici e complessi utilizzando il modello linguistico ChatGPT (versione 4) di Open AI. Per i casi semplici, sono state fornite a ChatGPT brevi descrizioni di sintomi e segni di condizioni comuni come pneumotorace, dolore toracico ed emottisi. ChatGPT ha quindi generato descrizioni complete di casi clinici sulla base di queste informazioni. Per i casi complessi, è stato utilizzato un approccio di *role play* in cui lo sperimentatore interagiva con ChatGPT nei panni di un chirurgo toracico e ChatGPT rispondeva nei panni del paziente. Sono stati discussi casi che richiedevano valutazioni diagnostiche dettagliate e decisioni terapeutiche complesse. Inoltre, nell'ipotesi di voler testare la preparazione teorica dei chirurghi ChatGPT è stato utilizzato per generare 20 quiz a scelta multipla focalizzati su argomenti rilevanti per la chirurgia toracica. Alla fine 6 chirurghi toracici con diversi gradi di esperienza hanno valutato l'accuratezza, il realismo e il valore educativo delle simulazioni dei casi clinici e hanno completato il quiz generato da ChatGPT. È stata utilizzata la scala Likert a 5 punti per valutare questi parametri.

Risultati

I chirurghi hanno assegnato punteggi elevati per l'accuratezza, il realismo e il valore educativo delle simulazioni di casi clinici generate da ChatGPT (Tabella 1) riferendo che i casi rispecchiavano fedelmente scenari della pratica clinica reale. Le simulazioni avrebbero inoltre fornito l'opportunità di testare le proprie capacità decisionali e di esplorare opzioni diverse. Sarebbe anche stato apprezzato il grado di dettaglio delle simulazioni. I quiz generati da ChatGPT sono stati a loro volta considerati accurati e utili per testare le loro conoscenze. Essi avrebbero migliorato infatti la loro capacità di ragionamento clinico e identificato aree di ulteriore apprendimento. I chirurghi hanno tuttavia anche sottolineato l'importanza della supervisione umana nell'utilizzo di ChatGPT. Hanno infatti osservato che, nonostante le sue capacità, ChatGPT non può replicare completamente la complessità delle interazioni reali con i pazienti. Pertanto, i suoi consigli devono essere interpretati con cautela e comunque integrati con il giudizio clinico umano.

Discussione

Questo studio ha dimostrato che i moderni modelli linguistici di AI come ChatGPT possono generare simulazioni di casi clinici altamente accurate e realistiche per l'educazione e la formazione in chirurgia toracica. La possibilità di creare una vasta gamma di

scenari clinici in modo efficiente rappresenta un notevole vantaggio di questi sistemi. Inoltre, le simulazioni di *role play* e i quiz mirati possono offrire ai chirurghi in formazione preziose opportunità per mettere in pratica le proprie capacità decisionali e di consolidare le conoscenze. Sebbene i sistemi AI come ChatGPT continuino a migliorare, non possono ancora replicare appieno la complessità delle situazioni cliniche reali. Pertanto, il loro uso deve essere attentamente monitorato e i loro consigli interpretati criticamente alla luce dell'esperienza clinica.

Conclusioni

I modelli linguistici di AI rappresentano strumenti promettenti per migliorare la simulazione di casi clinici e migliorare l'educazione e la formazione in chirurgia toracica. Tuttavia, il loro uso efficace richiede una supervisione attenta da parte di educatori umani esperti. Un approccio equilibrato, che sfrutti i punti di forza dell'IA e dell'*expertise* umana, potrà condurre a miglioramenti significativi nell'apprendimento e nella pratica anche della chirurgia toracica.

Bibliografia di riferimento

- BOMMASANI R, HUDSON DA, ADELI E, ET AL. *On the opportunities and risks of foundation models*. arXiv: 2108.07258.
- SALLAM M, *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns*. Healthcare 2023;11:887.

MEDICO	ACCURATEZZA	REALISMO	VALORE EDUCATIVO
M1	4	4	3
M2	4	4	4
M3	4	4	3
M4	4	4	4
M5	4	4	3
M6	4	4	4

Tabella 1. Scala Likert a 5 punti.