

## Intelligenza generativa artificiale in medical education: ragionamento clinico artificiale vs ragionamento clinico umano

### Artificial Generative Intelligence in Medical Education: Artificial Clinical Reasoning vs Human Clinical Reasoning

Rosa Cera\*

#### Riassunto

La finalità principale del presente contributo è di illustrare le potenzialità dell'utilizzo dell'intelligenza generativa artificiale (GenAI) in *medical education*. In particolare, l'autore persegue quattro specifici obiettivi: illustrare le potenzialità di GenAI e nello specifico di LLM (Large Language Model) e GPT-4 (quarta generazione della serie GPT, modello linguistico di grandi dimensioni multimodale) per lo sviluppo del curriculum in *medical education* (integrazione di contenuti di conoscenza, personalizzazione degli obiettivi di apprendimento, utilizzo di strumenti didattici innovativi come i pazienti virtuali); documentare il contributo di GenAI nel ragionamento clinico e la necessità di fare riferimento all'intelligenza ibrida, un misto tra le due, dove entrambe svolgono compiti epistemici chiaramente delineati e complementari; effettuare una chiara distinzione tra compiti epistemici del *clinical decision support systems* (CDSS) e quelli invece propri dell'essere umano, oltre a sottolineare l'importanza del contesto *embedded* nella elaborazione diagnostica; progettare un *teaching framework* di ragionamento clinico.

**Parole chiave:** intelligenza artificiale generativa, *medical education*, ragionamento clinico, compiti epistemici, intelligenza ibrida, *teaching framework*.

#### Abstract

The main purpose of this paper is to illustrate the potential of using generative artificial intelligence (GenAI) in medical education. In particular, the author pursues four specific objectives: to illustrate the potential of GenAI and in

---

\* Ricercatrice in Pedagogia Generale e Sociale presso l'Università degli Studi di Foggia. E-mail: [rosa.cera@unifg.it](mailto:rosa.cera@unifg.it).

particular of LLM (Large Language Model) and GPT-4 (4th generation of GPT series, large multimodal language model) for the development of the curriculum in medical education (integration of knowledge contents, personalization of learning objectives, use of innovative teaching tools such as virtual patients); to document the contribution of GenAI in clinical reasoning and the need to refer to hybrid intelligence, a mix between the two, where both perform clearly delineated and complementary epistemic tasks; to make a clear distinction between epistemic tasks of clinical decision support systems (CDSS) and those specific to humans, in addition to underlying the importance of the embedded context in diagnostic processing; to design a teaching framework for clinical reasoning.

**Keywords:** generative artificial intelligence, medical education, clinical reasoning, epistemic tasks, hybrid intelligence, teaching framework

*First submission: 29/08/2024, accepted: 03/12/2024*

## 1. LLMs e GPT-4 in medical education

I Large Language Model (LLMs) sono modelli di apprendimento automatico, Chatgpt è uno dei più recenti, gestiscono vasti volumi di dati, mostrano un linguaggio umano avanzato e sono in grado di svolgere diverse attività come l'elaborazione del linguaggio naturale, la traduzione, il riassunto e la correzione grammaticale (Safranek *et al.*, 2023). Gli LLM sono considerati modelli di *generative artificial intelligence* (GenAI) in quanto dato un prompt, un blocco di testo, sono in grado di prevedere la parola successiva che seguirà il prompt stesso, generando così una lunga sequenza di testo coerente e grammaticalmente corretto. Al di là del completamento automatico di frasi e concetti, gli LLM offrono alcune specifiche opportunità in *medical education* tra cui (Abd-Alrazaq *et al.*, 2023): *sviluppo del curriculum; metodologie di insegnamento; piani di studio personalizzati e materiali di apprendimento; valutazioni; assistenza nella scrittura; ricerca medica e revisione della letteratura; monitoraggio dei programmi*. Riguardo allo *sviluppo del curriculum*, GPT-4, quarta generazione di modello linguistico di grandi dimensioni multimodale, offre ai docenti la possibilità di individuare le lacune nei contenuti di conoscenza, fornendo approfondimenti su vari argomenti medici (Wang *et al.*, 2023). Con l'introduzione delle nuove tecnologie, il modo stesso di organizzare le conoscenze è, infatti, cambiato, in quanto non più considerata come un *corpus* unico e validato, ma come in continuo divenire perché fluida e aperta (Corsi *et al.*, 2023a).

Una conoscenza fluida che attraverso il confronto continuo con altri saperi va oltre i propri confini, al fine di acquisire un valore trans-disciplinare idoneo a interpretare e comprendere la complessità propria della contemporaneità. Sempre attraverso GPT-4 è, inoltre, possibile, *personalizzare gli obiettivi di apprendimento*, in modo da soddisfare i diversi bisogni formativi, promuovendo così esperienze personalizzate e adattive. In altre parole, LLM e GPT-4 durante la pianificazione del “Corso di Laurea in Medicina e Chirurgia” possono offrire supporto riguardo al contenuto di ogni corso di studio, agli obiettivi di apprendimento e alle metodologie di insegnamento. LLM, oltre a consentire una maggiore integrazione di contenuti di conoscenza su specifici argomenti, può, ad esempio, anche creare scenari di pazienti virtuali realistici, porre domande, interpretare le risposte e fornire feedback, aiutando così gli studenti nel ragionamento clinico e nella *decision making* (Grunhut *et al.*, 2022). GPT-4 può a sua volta facilitare le discussioni, nel lavoro a piccolo gruppo, generando domande stimolanti e incoraggiando le interazioni peer-to-peer. Inoltre, sempre GPT-4 può promuovere la pratica clinica attraverso la proposta di casi studio interattivi, facilitando così non solo il processo diagnostico, ma anche le opzioni di trattamento e le considerazioni etiche. La pratica clinica come le attività teoriche traggono, quindi, reali vantaggi nell'utilizzo di LLM, personalizzando ad esempio le attività sulla base dei ritmi e degli stili di apprendimento di ogni studente. Riguardo a ciò, GPT-4 può, ad esempio, offrire riassunti e schede sintetiche relative ad argomenti specifici su cui gli studenti ritengono di dover migliorare. L'utilizzo di LLM e GPT-4 può ritornare utile sia ai docenti che agli studenti anche nei processi di *valutazione*, innescando ad esempio meccanismi di feedback, al fine di individuare i progressi compiuti nel raggiungimento degli obiettivi di apprendimento (Gilson *et al.*, 2023). Un ulteriore supporto di LLM e GPT-4 consiste nella *scrittura* e nella stesura di documenti, attraverso la selezione di frasi e linguaggio adeguati. Nella *ricerca medica e nella revisione della letteratura*, LLM è, infatti, considerato strumento prezioso nel raccogliere in modo rapido le informazioni necessarie per l'elaborazione dei rapporti e degli articoli di ricerca. Per quanto attiene al *monitoraggio* e alla *revisione dei programmi*, LLM come GPT-4 possono fornire ai responsabili della progettazione informazioni relative all'efficacia dei programmi stessi attraverso l'analisi di dati raccolti da varie fonti, come i feedback degli studenti o i risultati di test (Fig. 1).

### 1.1 GenAI e attività curriculari

Le opportunità offerte da LLM e GPT-4 in *medical education* potrebbero corrispondere a specifiche attività riferibili allo sviluppo del curriculum medico

(Ahn, 2023): *self-learning* con testo dinamico, permettendo agli studenti, attraverso domande mirate e materiale di apprendimento specifico, di acquisire specifici contenuti di conoscenza e sviluppare il ragionamento clinico; elaborazione di saggi e articoli, sintetizzando i documenti, riscrivendo un paragrafo e scrivendo un saggio per intero facendo riferimento solo a un elenco di parole chiave; trasformazione dei metodi di valutazione promuovendo negli studenti l'utilizzo del pensiero critico, al fine di comprendere e di interpretare quanto proposto da LLM e GPT-4, oltre a facilitare le valutazioni formative attraverso maggiori feedback; interazione con chatbot di pazienti simulati permettendo agli studenti di migliorare le capacità di comunicazione, il recupero delle informazioni cliniche, l'abilità nella risoluzione dei problemi e il ragionamento clinico. In ambito medico, LLM è quindi capace di eseguire ragionamenti e di interagire con gli studenti iscritti ai primi anni di medicina, i quali nel porre domande riguardo alle cause e alle relative conseguenze di alcune malattie possono ricevere risposte congrue e logiche. In particolare, GPT-4 è considerato utile per generare casi simulati dei pazienti e valutazioni didattiche per integrare l'educazione medica tradizionale (Eysenbach, 2023). Un modello questo più avanzato in quanto in grado di generare risposte dettagliate alle domande poste dagli studenti a seguito della maggiore quantità di dati inglobati (Fig. 1). Inoltre, con il supporto della GenAI, i docenti possono fornire feedback tempestivi e personalizzati a un ampio gruppo di studenti durante le lezioni, in modo da prevedere i risultati di apprendimento (Gratani *et al.*, 2023).

Fig. 1 - LLM in medical education



## 2. Per una intelligenza ibrida nel ragionamento clinico

Tra le abilità che gli studenti di medicina sono chiamati ad acquisire, il ragionamento clinico è una delle più importanti, consiste nel raccogliere e analizzare i dati dei pazienti per poi effettuare una diagnosi e quindi prendere decisioni in tempi, quasi sempre, brevi sui trattamenti da adottare. Il ragionamento clinico è considerato come un modo di pensare strettamente connesso al contesto e un saper prendere decisioni nel guidare le azioni pratiche (Higgs, Jensen, 2018). In altre parole, la capacità di saper prendere decisioni del clinico non può prescindere dalla considerazione dell'influenza che il contesto esercita nel modo in cui ogni paziente vive, racconta e interpreta i sintomi della propria malattia. Questo implica attività cognitive e metacognitive che possono essere sia consapevoli sia inconsapevoli, ma sempre connesse a ogni singola e unica situazione vissuta dai pazienti e unicamente interconnesse ai fattori contestuali. Il ragionamento clinico può, quindi, essere considerato come un insieme di processi di risoluzione dei problemi, basato sulla diagnosi differenziale, sul ragionamento intuitivo, sul ragionamento analitico e sull'inferenza bayesiana. Allo stesso tempo, è stato dimostrato che LLM, pur essendo capace di dividere un compito in piccole fasi di ragionamento e quindi completare l'attività passo-passo, non è ancora in grado di garantire la precisione e l'affidabilità del ragionamento clinico umano. La causa di tale incapacità è dovuta alle allucinazioni (intese queste ultime come risultati a cui giunge LLM che a volte non sono coerenti con i dati di input inseriti nel sistema) imprevedibili che rendono quindi impossibile assicurare la razionale accuratezza logica e fattuale di un pensiero (Savage *et al.*, 2024). Per questa ragione, alcuni studiosi invece di proporre la sostituzione del ragionamento umano con quello artificiale hanno pensato a un genere di intelligenza ibrida, un misto tra le due, dove entrambe svolgono compiti epistemici chiaramente delineati e complementari (van Baalen *et al.*, 2021). Affinché gli studenti di medicina imparino, però, a utilizzare l'intelligenza ibrida nel ragionamento clinico è necessario che sappiano distinguere le attività epistemiche svolte dal *clinical decision support systems* (CDSS - sistemi di supporto alle decisioni cliniche) da quelle svolte dall'essere umano.

I compiti epistemici sono compiti di prestazione, in cui si fa ricorso a conoscenze appropriate affinché un'attività possa essere svolta nel migliore dei modi. Diversi sono i compiti epistemici che gli algoritmi di apprendimento automatico possono svolgere nel ragionamento clinico (Boon, 2020): abbinare i dati di input (ad esempio, un insieme di dati come segni clinici e sintomi) con casi simili nel loro database; interpretare i dati di input come appartenenti a una categoria specifica, definita dagli esseri umani; diagnosticare un set di dati di input come probabilmente appartenenti a una certa classe e da ciò dedurre altre

proprietà del target; strutturare grandi quantità di dati per trovare modelli, correlazioni e relazioni causali; effettuare calcoli e statistiche; simulare un processo dinamico complesso. Il CDSS può, quindi, fornire un utile supporto nel ragionamento statistico, induttivo e deduttivo effettuando, ad esempio, confronti tra le informazioni che i medici hanno su un paziente con le informazioni riguardanti gruppi di altri pazienti presenti nel database. Attraverso questo confronto potrebbe essere più facile effettuare una diagnosi e prevedere i risultati di uno specifico trattamento.

### *2.1 Compiti epistemici attribuibili a CDSS e compiti epistemi propri del clinico*

Nonostante il contributo offerto da CDSS, il ragionamento clinico non consiste solo nel classificare i sintomi e abbinarli tra loro, un compito epistemico essenziale è quello di saper interagire con i sintomi e la malattia del paziente. Compito quest'ultimo non certo in grado di essere svolto dal CDSS, ma attribuibile solo all'intelligenza umana, attraverso la quale è possibile rivedere ed elaborare i dati prima di inserirli nel CDSS, al fine di individuare le informazioni importanti e valutare i risultati.

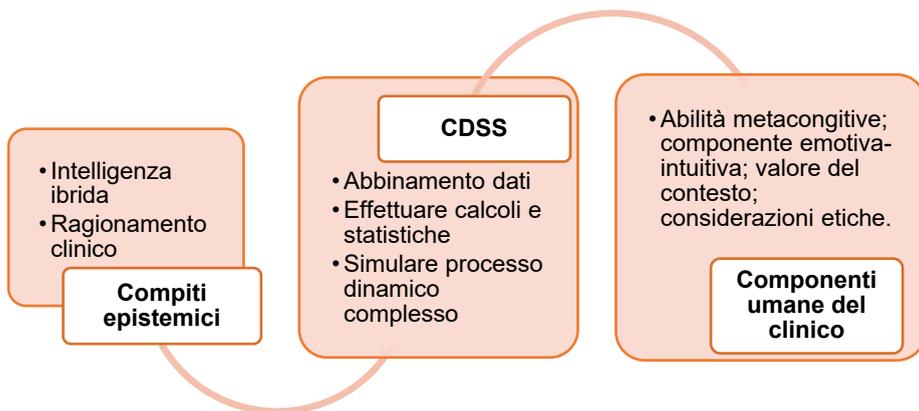
Al di là dei compiti epistemici svolti dal CDSS nel ragionamento clinico, lo studente di medicina dovrebbe anche imparare a conoscere gli aspetti del ragionamento propri dell'essere umano. Aspetti questi ultimi che riguardano alcuni specifici compiti epistemici in cui la componente emotiva e profondamente intuitiva, l'abilità metacognitiva, le considerazioni etiche e il valore del contesto rivestono un ruolo cruciale. Nel ragionamento clinico, le componenti in precedenza menzionate contribuiscono, infatti, a rendere il procedimento diagnostico più riflessivo e interpretativo, in quanto più incline a considerare la dimensione umana della malattia. In altre parole, il ragionamento riflessivo partendo dal particolare e riflettendo sulle singole situazioni consente di ricercare un concetto, una teoria; mentre l'utilizzo di tecniche narrative facilita l'integrazione logica delle informazioni disponibili.

L'acquisizione di competenze utili per lo svolgimento di compiti epistemici riguardanti il ragionamento clinico non può, quindi, consistere solo nell'imparare a utilizzare GenAI, ma anche nell'acquisire consapevolezza delle proprie conoscenze e competenze (van Baalen, Bonn, 2017). Un'abilità quest'ultima riferibile all'ambito metacognitivo al quale andrebbe associata anche la capacità di lavorare in team. L'elaborazione di una diagnosi non potendo, infatti, essere mai il frutto del ragionamento di una singola persona ma il risultato del confronto e dell'interazione epistemica tra esperti, pone in evidenza l'essenzialità dell'imparare a collaborare durante le fasi diagnostiche o di trattamento.

In sintesi, i compiti epistemici propri dell'essere umano nel ragionamento clinico, facendo affidamento sulle abilità metacognitive, sulle pratiche narrative

unitamente alle capacità cognitive, alla componente emotiva-intuitiva e al riconoscimento del valore assoluto del contesto, dimostrano la loro unicità ed esclusività rispetto al CDSS (Fig.2).

Fig. 2 - Intelligenza ibrida e ragionamento clinico: compiti epistemici di CDSS e componenti umane del clinico



### 3. Abilità metacognitive, componente emotiva-intuitiva, contesto, dimensione etica: GenAI e ragionamento clinico umano

Il ragionamento clinico, al di là dall'essere pensato come processo puramente razionale e cognitivo, si distingue per la complessità del procedimento diagnostico. Una complessità questa che necessita sia del supporto di GenAI sia del pensiero umano per poter essere compresa e gestita. Al di là del notevole contributo che GenAI offre nel ragionamento clinico, illustrato in parte in precedenza, permangono differenze significative tra il ragionamento clinico umano e quello artificiale. GenAI è ad esempio priva della componente emotiva e profondamente intuitiva presente invece nel processo decisionale umano, il quale interpreta i dati all'interno di un contesto più ampio. Un'abilità quest'ultima che GenAI deve ancora imparare a replicare con precisione. In altre parole, mentre ci sono sostanziali somiglianze nell'utilizzo dei metodi statistici e nell'analisi dei dati tra il processo decisionale clinico e il ragionamento di GenAI, gli aspetti umani dell'intuizione, dell'empatia e delle considerazioni etiche pongono in evidenza le caratteristiche uniche di ciascun compito epistemico. Inoltre, è sempre e solo unica responsabilità umana quella di verificare l'affidabilità e l'accuratezza delle informazioni inserite in GenAI. Un compito questo di fondamentale importanza affinché i consigli generati da GenAI possano

essere in linea con gli attuali standard medici ed etici, rispettare l'autonomia del paziente e garantire un trattamento equo. La risoluzione di problemi clinici necessita, infatti, non solo di conoscenze mediche specifiche e di abilità cognitive, ma anche di consapevolezza e di capacità di monitoraggio dei propri processi di pensiero.

In tal senso, l'abilità metacognitiva è ciò che consente agli umani, a differenza di GenAI, di monitorare il ragionamento diagnostico e per questa ragione è considerata essenziale nella *medical education*. L'abilità metacognitiva che distingue, quindi, il ragionamento umano è ciò che promuove negli studenti l'apprendimento auto-regolato e come tale favorisce la conoscenza della cognizione, l'analisi dei compiti (inclusa la definizione degli obiettivi e la pianificazione strategica), il monitoraggio, la valutazione e l'adattamento (Cera *et al.*, 2013). Tutti compiti questi che non sempre GenAI è in grado di svolgere in modo riflessivo e con estrema esattezza.

Le componenti emotiva e intuitiva distinguono altresì il ragionamento clinico umano da quello eseguito da GenAI, per il significato stesso attribuito alla definizione di intuizione. Intesa quest'ultima come preconsua, olisticamente associativa e carica di affetti (Sinclair, Ashkanasy, 2005), una dimensione quella dell'intuizione dove a rivestire un ruolo significativo non è solo il "brain" (aspetti cognitivi), ma anche il "body" (aspetti emozionali, sensoriali, somatici e viscerali). Attraverso l'intuizione diventa, quindi, possibile aprire il ragionamento clinico a uno spettro di creatività multilivello, promuovendo la ricerca di soluzioni creative ai problemi diagnostici. Soluzioni queste ancora difficilmente attribuibili a GenAI.

Il contesto, a sua volta, rappresenta, a seconda del significato ad esso attribuito, un ulteriore elemento assente nell'interpretazione dei dati clinici da parte di GenAI a differenza di quanto avviene invece nel ragionamento clinico umano. Attraverso la lettura e interpretazione del contesto è possibile comprendere quanto accade a un paziente, in quanto nessun sintomo può essere decontestualizzato poiché perderebbe la sua stessa significazione. Il significato del contesto è, quindi, determinato da ipotesi relative alla natura della conoscenza e al modo di conoscere. Inoltre è l'epistemologia del ragionamento clinico a determinare come il contesto è concettualizzato e interpretato. La sola applicazione della conoscenza teorica come linea guida basata sull'evidenza non può, ad esempio, essere considerata sufficiente nel ragionamento clinico. A volte, è anche necessario fare ricorso alla *wisdom* (saggezza) umana, in quanto in grado di ritagliare le caratteristiche salienti e ambigue di una situazione e di individuare il modo di agire più appropriato per ogni specifica circostanza. In altre parole, la padronanza del contesto assume particolare pregnanza nel ragionamento clinico, in quanto capacità in grado di navigare all'interno delle dimensioni epistemologiche delle situazioni, in modo da tollerare le incertezze e da

interpretare criticamente quanto accade nel contesto stesso (Koufidis *et al.*, 2022).

Legate al contesto sono anche le considerazioni etiche che dovrebbero essere presenti in ogni ragionamento clinico e che invece sono, per lo più assenti, in GenAI. I pazienti meritano, infatti, trasparenza sul come i medici prendono decisioni sul loro trattamento in base al ragionamento clinico e i medici, a loro volta, dovrebbero essere in grado, sulla base del tipo di formazione ricevuta, di offrire spiegazioni riguardo alle loro decisioni (Shamy *et al.*, 2023). Altresì è utile ricordare come, ancora oggi, permangano numerosi problemi etici associati all'utilizzo di GenAI nel processo decisionale clinico. Alcuni problemi etici riguardano, ad esempio, la privacy dei pazienti, oppure la capacità dei medici di individuare i dati distorti spesso forniti dalla stessa GenAI. In genere, la persona che utilizza l'intelligenza artificiale non conosce i passaggi intermedi o i dati che l'intelligenza artificiale utilizza nel formulare le sue opinioni finali. Il paziente non è, quindi, in grado di conoscere quali fattori sono stati soppesati, sottovalutati o addirittura omessi dalla GenAI nel formulare la sua diagnosi. Questo potrebbe generare dubbi riguardo alla possibilità che i pregiudizi umani abbiano infettato gli algoritmi di IA. Inoltre, non potendo conoscere tutti i processi attraverso cui GenAI giunge alle sue conclusioni, non è neanche possibile sapere in quale punto del processo è stato commesso un errore, e quindi quando un medico potrebbe intervenire per eliminare l'errore stesso.

#### **4. Un modello di teaching framework di ragionamento clinico: GenAI, pazienti virtuali e componenti umane**

LLM in *medical education*, avvalendosi dell'utilizzo di pazienti virtuali al fine di promuovere il ragionamento diagnostico, utilizza simulazioni di scenari clinici. Attraverso questo gioco di simulazioni, gli studenti nell'emulare il ruolo di clinico raccolgono storie di malattie, conducono un esame fisico, elaborano diagnosi e prevedono terapie. In altre parole, i pazienti virtuali sono utilizzati per insegnare e valutare la qualità dei colloqui, il procedimento diagnostico e il trattamento delle condizioni mediche. Le attività che i pazienti virtuali possono svolgere sono progettate sulla base degli obiettivi didattici da conseguire e dei compiti epistemici da eseguire. Sempre gli stessi pazienti virtuali possono essere utili sia nelle valutazioni formative che in quelle sommative, al fine di verificare le abilità acquisite dagli studenti nel ragionamento clinico. Molto dipende, però, dal modo in cui i pazienti simulati sono stati progettati, dal tipo di attività che sono in grado di svolgere e dal modo in cui gli stessi strumenti virtuali sono utilizzati.

Il *framework* di ragionamento clinico, qui proposto, comprende sia compiti epistemici svolti facendo affidamento sull'intelligenza umana sia compiti epistemici eseguibili attraverso l'utilizzo di GenAI. In particolare, il quadro didattico proposto si esplicita attraverso alcune specifiche fasi: *role-modelling*, in cui il medico esperto presenta un identico caso clinico, utilizzando sia un paziente virtuale sia reale; pratica clinica in cui gli studenti, in piccolo gruppo, analizzano un identico caso clinico su paziente virtuale e reale; individuazione di possibili ipotesi diagnostiche; formulazione della diagnosi; valutazione formativa e sommativa. In queste diverse fasi del ragionamento clinico, gli studenti svolgono compiti epistemici avvalendosi sia delle competenze proprie dell'essere umano, come l'intuizione, sia della GenAI.

Nel *role-modeling*, gli studenti non si limitano, ad esempio, solo a osservare l'esperto, ma cercano anche di sollecitare il pensiero attivo facendo appello alla loro componente empatica ed emotiva, ricercando l'interazione con l'esperto stesso. In questa prima fase, gli studenti fanno affidamento a componenti esclusivamente umane.

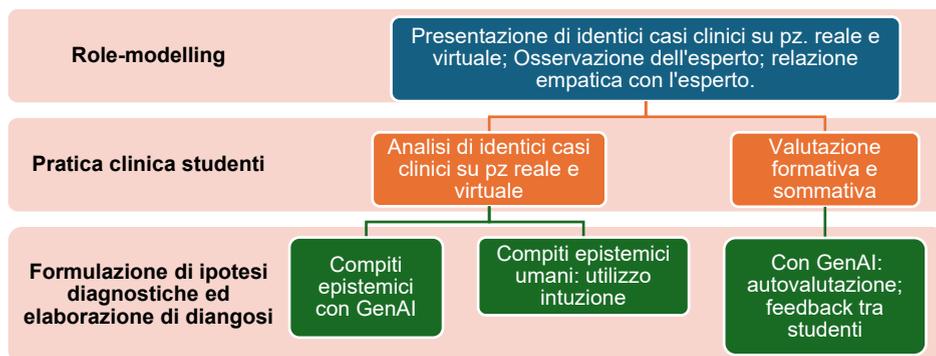
Nella fase di pratica, in cui sono oggetto di analisi due uguali casi clinici, gli studenti, in piccolo gruppo, analizzano entrambi, di cui uno riferibile a paziente virtuale e un altro a uno reale. In questa fase, come in quella precedente, è necessario che lo studio dei casi venga fatto su pazienti virtuali e reali, in modo che gli studenti si rendano conto di come il contesto influenza l'analisi stessa. L'unicità e la singolarità del contesto in cui ogni paziente vive e con cui si relaziona influenza il modo in cui ogni persona interpreta, intende e comprende la propria malattia. Questa la ragione per la quale, il contesto rappresenta una variabile non trascurabile nell'elaborazione delle ipotesi diagnostiche.

Nelle fasi di analisi dei casi clinici e di formulazione di ipotesi diagnostiche, gli studenti svolgono, quindi, compiti epistemici avvalendosi sia della GenAI e in particolare dell'LLM in quanto in grado di abbinare un insieme di dati come segni clinici e sintomi con casi simili nel database, sia dell'intuizione umana, al fine di formulare alcune possibili ipotesi di diagnosi. Anche nella fase successiva di elaborazione di diagnosi, gli studenti si avvalgono dell'aiuto sia della GenAI, in quanto capace di diagnosticare un set di dati (segni clinici) come probabilmente appartenenti a una certa classe, sia dell'intuizione umana cercando di non dimenticare la dimensione etica della diagnosi stessa.

Nella fase finale di valutazione formativa e sommativa, gli studenti possono altresì utilizzare sia la GenAI, la quale attraverso interviste e domande mirate promuove l'auto-valutazione e quindi l'utilizzo del pensiero critico, sia il feedback tra gli studenti stessi, i quali facendo ricorso alle abilità metacognitive possono monitorare la qualità del loro ragionamento diagnostico (Fig. 3). In

questo caso, le tecnologie contribuiscono nel rendere la valutazione un'esperienza più significativa per gli studenti, introducendo nuove prospettive di azione e di relazione, promuovendo la riflessione sui propri compiti e l'autovalutazione sulla propria traiettoria individuale e professionale (Giannandrea, Winstone, 2024).

Fig. 3 - Teaching framework di ragionamento clinico: GenAI e componenti umane



## 5. Potenzialità e limiti di GenAI in medical education

GenAI ha scientificamente dimostrato di avere notevoli potenzialità in *medical education*, in particolare in una delle più importanti attività, il ragionamento clinico (Cabral *et al.* 2024; Krishnan *et al.*, 2023). Un'attività questa la cui complessità necessita di una revisione del modo in cui finora è stato insegnato il ragionamento clinico. I risultati di alcuni studi empirici hanno, ad esempio, quantificato i miglioramenti nelle prestazioni degli studenti quando utilizzano strumenti basati sull'intelligenza artificiale. A tale riguardo, l'obiettivo dello studio di Alegría-Bernal, Fernández-Delgado e Andía-Alegría (2024) è di valutare l'impatto di GenAI sull'apprendimento nella pratica clinica così come percepito dagli studenti. I risultati di questo studio hanno dimostrato come la maggior parte degli intervistati ha indicato di essere pienamente d'accordo con le affermazioni che valutano i benefici dell'IA, tra cui la varietà di casi clinici (66,6%), il feedback immediato e personalizzato (46,7%) e la motivazione e l'interesse nell'apprendimento (58,3%). Questi risultati evidenziano l'utilità percepita dell'IA nel facilitare l'applicazione e il trasferimento della conoscenza (41,6%) e nel migliorare la soddisfazione dell'apprendimento (58,3%). Inoltre, gli intervistati hanno identificato la riflessione e il miglioramento della propria pratica (41,6%), la stimolazione della creatività e del pensiero critico

(66,6%) e la facilitazione del lavoro collaborativo (50%) come benefici significativi.

Allo stesso tempo bisogna anche segnalare come alcuni studi hanno posto in evidenza come l'utilizzo di ChatGPT non sempre sia un affidabile supporto clinico. Lo studio di Saban e Dubovi (2024) ha, ad esempio, valutato la capacità del chatbot di fornire risposte di alta qualità agli scenari clinici di triage confrontando le sue risposte con quelle degli infermieri. ChatGPT ha mostrato una certa indecisione nel determinare accuratamente il livello di urgenza; è stato impreciso nelle sue raccomandazioni di test di laboratorio e di imaging. In particolare, tendeva a sovrastimare suggerendo test non necessari e contemporaneamente non richiedendo test necessari. Con GPT-4 il livello di affidabilità è sicuramente migliorato, ciononostante sarebbe consigliabile che gli studenti non facessero esclusivo affidamento su GenAI nella loro pratica clinica. Alla luce di quanto in precedenza illustrato, sarebbe utile che gli studenti imparassero, ad esempio, a fare maggiore affidamento sulle proprie componenti umane come l'intuizione, la sensibilità e la riflessione, al fine di poter leggere e interpretare l'analisi dei dati diagnostici. Questo il motivo per il quale nel *teaching framework* proposto, gli studenti sono chiamati ad analizzare e diagnosticare identici casi clinici, sia attraverso pazienti virtuali sia attraverso quelli reali, in modo che possano prendere atto dell'impossibilità di poter sostituire per intero l'intelligenza umana con quella artificiale. Come impossibile risulta, alla fine, essere anche l'estrapolazione completa del paziente dal proprio contesto, in quanto ognuno ha una propria storia di malattia che va letta e interpretata nel contesto sociale e culturale di appartenenza. In questo senso, solo una pedagogia in grado di trascendere i propri confini per arrivare dall'iperuranio greco all'intelligenza artificiale (Corsi *et al*, 2023b), nella propria poliedricità può contribuire alla formulazione di un ragionamento clinico in cui componenti umane e artificiali si integrano nel promuovere il passaggio da una medicina esclusivamente basata sulle evidenze a una medicina basata su decisioni condivise con il paziente e tra professionisti.

## 6. Conclusioni

In *medical education*, il ragionamento clinico tende alla promozione dell'approccio olistico al paziente, in cui ad essere oggetto di analisi non sono solo i sintomi della malattia, ma anche gli aspetti fisiologici, psicologici, sociali, emozionali e comportamentali di ogni singola persona. Il paziente necessita, quindi, di essere co-costruttore del ragionamento clinico, in modo da promuovere un genere di procedimento diagnostico dal valore esclusivamente oli-

stico (Rahaman *et al*, 2023). Un valore questo che verrebbe meno se si utilizzasse solo GenAI, la quale elaborando i dati (sintomi) in modo unicamente sintetico e cognitivo non consentirebbe agli studenti di cogliere le esperienze di apprendimento significativo del ragionamento clinico. Inoltre, sarebbe conveniente ricordare come il significato stesso del contesto vari a seconda della natura della conoscenza e del sapere, in quanto è l'epistemologia del ragionamento clinico a determinarne la concettualizzazione (Koufidis *et al.*, 2022). Ogni evidenza empirica necessita, quindi, di essere interpretata attraverso il significato attribuito al contesto. Significato del contesto e relazione con il paziente rappresentano, infatti, due dimensioni principali del ragionamento clinico, alle quali lo studente dovrebbe approcciarsi facendo leva sulla propria *agency*. In questo modo, lo studente non farebbe esclusivo riferimento al *machine learning* (apprendimento automatico supportato dalla tecnologia abilitante del *key enabling technology*) e imparerebbe a elaborare un proprio pensiero critico e ad auto-regolare il proprio apprendimento (Costa, 2022). La capacità dello studente di relazionarsi al paziente contribuirebbe, così, a evitare un prosciugamento delle qualità umane per fini unicamente prestazionali (d'Aniello, 2022). Alla luce del ruolo imprescindibile ed esclusivo svolto dalle abilità proprie dell'essere umano nel ragionamento clinico, sarebbe improponibile pensare a un procedimento diagnostico esclusivamente basato sull'utilizzo delle tecnologie.

## Riferimenti bibliografici

- Abd-Alrazaq A., AlSaad R., Alhuwail D., Ahmed A., Healy P.M., Latifi S., Aziz S., Damseh R., Alabed Alrazak S. and Sheikh J. (2023). Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Medical Education*, 9, e48291. DOI: 10.2196/48291.
- Ahn S. (2023). The impending impacts of large language models on medical education. *Korean Journal of Medical Education*, 35(1): 103-107. DOI: 10.3946/kjme.2023.253.
- Alegria-Bernal C.M., Fernández-Delgado J.C. and Andía-Alegria F.S. (2024). Generative Artificial Intelligence in Clinical Practice: Undergraduate Experience. *Evolutionary Studies in Imaginative Culture*, 8.2(53): 532-542. DOI: 10.70082/esiculture.vi.1868.
- Boon M. (2020). How scientists are brought back into science – the error of empiricism. In: Bertolaso M., Sterpetti, F., editors, *A Critical Reflection on Automated Science – Will Science Remain Human. Springer Series Human Perspectives in Health Sciences and Technologie* (pp. 43-66). Dordrecht, the Netherlands: Springer.

- Cabral S., Restrepo D., Kanjee Z., Wilson P., Crowe B., Abdunnour R-E. and Rodman A. (2024). Clinical Reasoning of a Generative Artificial Intelligence Model Compared With Physicians. *Jama Internal Medicine*, 184(5): 581-583. DOI: 10.1001/jamainternmed.2024.0295.
- Cera R., Mancini M. and Antonietti A. (2013). Relationships between Metacognition, Self-efficacy and Self-regulation in Learning. *Educational Cultural and Psychological Studies*, 7: 115-141. Testo diponibile alla pagina: <https://www.ledonline.it/index.php/ECPS-Journal/article/view/511/500>.
- Corsi M., Stramaglia M., Guerra P. and Farina T. (2023a). Humani nihil a me alienum puto. La pedagogia “intera”. Editoriale. *Education Science and Society – Open Access*, 14(2): 15-22. DOI: 10.3280/ess2-2023oa16890.
- Corsi M., Rossi P.G., Giannandrea L. and Winstone N. (2023b). Didattica universitaria, innovazione e inclusione. Valutazione e feedback. Editoriale. *Education Science and Society – Open Access*, 14(1): 9-14. Testo disponibile alla pagina: <https://journals.francoangeli.it/index.php/ess/article/view/16109/2351>.
- Costa M. (2022). Formazione, innovazione e modelli pedagogici per la formazione dei lavoratori. In: Galimberti A., Muschitiello A., a cura di. *Pedagogia e lavoro: le sfide tecnologiche*. Fano (PU): Aras Edizioni.
- d’Aniello F. (2022). Le character skills tra riflessioni critiche e opportunità pedagogica. In: Di Vita, A., a cura di. *Orientare nelle transizioni scuola-università-lavoro promuovendo le character skills* (pp. 29-42). Lecce: PensaMultimedia.
- Eysenbach G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Medical Education*, 9, e46885. DOI: 10.2196/46885.
- Giannandrea L., Winstone N. (2024). Valutazione e tecnologie. In: Rivoltella P.C. e Rossi P.G., a cura di, *Tecnologie per l’educazione – 2/ED* (pp. 177-188). Milano: Pearson.
- Gilson A., Safranek C.W., Huang T., Socrates V., Chi L., Taylor R.A. and Chartash D. (2023). How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9: 1-9. DOI: 10.2196/45312.
- Gratani F., Capolla L.M., Giannandrea L., Screpanti L. and Scaradozzi D. (2023). Facilitating feedback at university using AI-based techniques. In A.A., *HELMeTO* (pp. 157-159) – *BOOK OF ABSTRACTS*. STUDIUM s.r.l.
- Grunhut J., Marques O. and Wyatt A.T.M. (2022). Needs, challenges, and applications of artificial intelligence in medical education curriculum. *JMIR Medical Education*, 8(2): 1-5. DOI: 10.2196/35587.
- Higgs J., Jensen G.M. (2018). Clinical reasoning: challenges of interpretation and practice in the 21st century. In: Higgs J., Jensen G.M, Loftus S. and Christensen N., editors, *Clinical reasoning in the health professions. 4th Edition*, (pp. 3-11). Amsterdam: Elsevier.
- Koufidis C., Manninen K., Nieminen J., Wohlin M. and Silén C. (2022). Representation, interaction and interpretation. Making sense of the context in clinical reasoning. *Medical Education*, 56: 98-109. DOI: 10.1111/medu.14545.

- Krishnan G., Singh S., Pathania M., Gosavi S., Abhishek S., Parchani A and Dhar M. (2023). Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. *Frontiers in Artificial Intelligence*, 6, 1227091. DOI: 10.3389/frai.2023.1227091.
- Rahman N.F.A., Davies N., Suhaimi J., Idris F., Mohamad S.N.S. and Park S. (2023). Transformative learning in clinical reasoning: a meta-synthesis in undergraduate primary care medical education. *Education for Primary Care*, 34(4): 211-219. DOI: 10.1080/14739879.2023.2248070.
- Saban M., Dubovi I. (2024). A comparative vignette study: Evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing. *Leading Global Nursing Research*, 00: 1-11. DOI: 10.1111/jan.16146.
- Safranek, C.V., Sidamon-Eristoff, A.E., Gilson, A. and Chartash, D. (2023). The Role of Large Language Models in Medical Education: Applications and Implications. *JMIR Medical Education*, 9: 1-12. DOI: 10.2196/50945.
- Savage T., Nayak A., Gallo R., Rangan E. and Chen J.K. (2024). Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *npj Digital Medicine*, 7(20): 1-7.
- Shamy M., Dewar B. and Fedyk M. (2023). Ethical evaluation in acute stroke decision-making. *Journal of Evaluation in Clinical Practice*, 30(5): 749-755. DOI: 10.1111/jep.13927.
- Sinclair M., Ashkanasy N.M. (2005). Intuition: Myth or a Decision-making Tool?. *Management Learning*, 36(3): 353-370. DOI: 10.1177/1350507605055351
- van Baalen S., Bonn M. (2017). Evidence-based medicine versus expertise – knowledge, skills and epistemic actions. In: Bluhm R., editor, *Knowing and Acting in Medicine* (pp. 21-38). London, UK: Rowman & Littlefield.
- van Baalen S., Bonn M. and Verhoef P. (2021). From clinical decision support to clinical reasoning support systems. *Journal of Evaluation in Clinical Practice*, 27: 520-528. DOI: 10.1111/jep.13541.
- Wang L.K.P., Paidisetty P.S. and Cano A.M. (2023). The next paradigm shift? ChatGPT, artificial intelligence, and medical education. *Medical Teacher*, 45(8): 925. DOI: 10.1080/0142159X.2023.2198663.