

# **La Convergenza tra Evidenza e Algoritmo: Come i Modelli Linguistici di Grandi Dimensioni stanno Rimodellando il Paesaggio della Pratica Clinica Basata sull'Evidenza**

Questo documento è stato realizzato con il Gemini pro di Franco Ognibene ponendogli la domanda:

*Fai una ricerca nelle banche dati biomediche e nel web per definire come la tecnica e la cultura EBN o Evidence based clinical practice cambia con l'utilizzo delle LLM o IA.*

## Abstract

Questo report analizza il cambiamento fondamentale nel paradigma della Medicina Basata sull'Evidenza (EBM), guidato dall'integrazione dei Modelli Linguistici di Grandi Dimensioni (LLM). Viene esaminata in dettaglio l'applicazione di queste tecnologie attraverso il flusso di lavoro in cinque fasi dell'EBM, evidenziando i benefici chiave come l'accelerazione della sintesi delle evidenze e il supporto decisionale personalizzato. Parallelamente, vengono affrontate le sfide critiche, tra cui la fallibilità algoritmica, i dilemmi etici e la necessità imprescindibile della supervisione umana. L'analisi dimostra come gli LLM stiano trasformando ogni fase del processo EBM, dalla formulazione del quesito clinico alla valutazione degli esiti. Si evidenzia che, sebbene gli LLM eccellano nell'acquisizione e nella sintesi rapida di grandi volumi di letteratura scientifica, la loro capacità di valutazione critica delle evidenze rimane un limite significativo, rendendo il giudizio del clinico più cruciale che mai. Vengono inoltre esplorate le profonde implicazioni culturali, tra cui l'evoluzione del ruolo del medico da detentore di conoscenza a validatore critico di output algoritmici, e la crescente importanza della fiducia nel rapporto medico-paziente-tecnologia. Si conclude che il futuro dell'EBM risiede in una collaborazione sinergica uomo-IA, un paradigma "EBM 2.0" che ridefinisce il ruolo del clinico e richiede nuovi quadri di riferimento per la formazione, la regolamentazione e la pratica clinica. Questo nuovo modello promette di rendere l'EBM più dinamica, accessibile e personalizzata, ma solo se guidata da una rigorosa validazione, da una supervisione umana esperta e da una solida governance etica.

# **1. Introduzione: La Simbiosi tra Medicina Basata sull'Evidenza e Intelligenza Artificiale**

L'intersezione tra la pratica medica e l'innovazione tecnologica ha raggiunto un punto di svolta. L'avvento di potenti strumenti di intelligenza artificiale (IA), in particolare i Modelli Linguistici di Grandi Dimensioni (LLM), sta iniziando a rimodellare le fondamenta stesse di come la conoscenza medica viene acquisita, valutata e applicata. Al centro di questa trasformazione si trova la Medicina Basata sull'Evidenza (EBM), un approccio sistematico alla pratica clinica che per decenni ha rappresentato il gold standard per un'assistenza sanitaria di alta qualità. Questo report esplora la profonda e complessa interazione tra i principi consolidati dell'EBM e le capacità emergenti degli LLM, analizzando come questa convergenza stia modificando non solo la "tecnica" ma anche la "cultura" della pratica clinica.

## **1.1. I Pilastri della Medicina Basata sull'Evidenza: Una Rivalutazione Moderna**

Per comprendere l'impatto degli LLM, è essenziale prima riaffermare i principi fondamentali dell'EBM. Lungi dall'essere una semplice applicazione di linee guida, l'EBM è definita come l'integrazione coscienziosa, esplicita e giudiziosa di tre componenti fondamentali: la migliore evidenza di ricerca disponibile, l'esperienza clinica del medico e i valori e le preferenze uniche del paziente.<sup>1</sup> Questo triangolo concettuale sottolinea che l'evidenza da sola non è mai sufficiente per prendere una decisione clinica; essa deve essere contestualizzata dall'esperienza del professionista e allineata con ciò che conta di più per il paziente.<sup>3</sup>

La metodologia dell'EBM è strutturata attorno a un processo in cinque fasi, che fornisce una tabella di marcia per tradurre l'incertezza clinica in una pratica informata:

1. **Ask (Formulare il quesito):** Convertire il bisogno di informazioni in un quesito clinico strutturato e suscettibile di risposta.
2. **Acquire (Acquisire l'evidenza):** Ricercare con la massima efficienza le migliori evidenze disponibili per rispondere al quesito.
3. **Appraise (Valutare criticamente l'evidenza):** Valutare l'evidenza per la sua validità (vicinanza alla verità), impatto (dimensione dell'effetto) e applicabilità (utilità nella pratica clinica).
4. **Apply (Applicare l'evidenza):** Integrare l'evidenza valutata con l'esperienza clinica e i valori del paziente per prendere una decisione condivisa.
5. **Assess (Valutare la performance):** Valutare l'efficacia e l'efficienza con cui sono state

eseguite le fasi precedenti e cercare modi per migliorare in futuro.<sup>1</sup>

Un concetto centrale in questo processo, in particolare nella fase di valutazione, è la gerarchia delle evidenze. Questo costrutto classifica i diversi disegni di studio in base alla loro capacità di minimizzare il bias. Al vertice si trovano le revisioni sistematiche e le meta-analisi di studi controllati randomizzati (RCT), considerati il "gold standard" per rispondere a quesiti di terapia.<sup>1</sup> Per altre domande cliniche, come quelle relative al danno o alla prognosi, altri disegni di studio, come gli studi di coorte o caso-controllo, assumono maggiore rilevanza.<sup>3</sup> Questa comprensione sfumata che "non tutte le evidenze sono uguali" è un principio epistemologico fondamentale dell'EBM.<sup>3</sup>

## **1.2. L'Avvento dei Modelli Linguistici di Grandi Dimensioni: Un Nuovo Catalizzatore per la Gestione della Conoscenza Medica**

Parallelamente a questa consolidata tradizione medica, è emersa una tecnologia dirompente: i Modelli Linguistici di Grandi Dimensioni. Gli LLM rappresentano un progresso rivoluzionario nell'IA, caratterizzati da una capacità senza precedenti di comprendere, generare e interagire con il linguaggio umano a un livello di complessità e fluidità notevole.<sup>5</sup> Basati su architetture di deep learning come i Transformer e addestrati su enormi corpora di dati testuali, questi modelli possono svolgere una vasta gamma di compiti di elaborazione del linguaggio naturale (NLP), dalla classificazione del testo e il riassunto alla risposta a domande complesse.<sup>7</sup>

Il loro potenziale in ambito sanitario è immenso e si estende a molteplici domini. Gli LLM possono fungere da potenti sistemi di supporto alle decisioni cliniche (CDSS), assistendo i medici nella diagnosi e nelle raccomandazioni terapeutiche.<sup>6</sup> Possono accelerare la ricerca medica analizzando e sintetizzando la vasta letteratura biomedica<sup>5</sup> e migliorare la comunicazione con i pazienti, offrendo consulenze e semplificando informazioni complesse.<sup>10</sup> È fondamentale riconoscere che gli LLM non sono semplici motori di ricerca potenziati; sono strumenti capaci di sintesi, inferenza e generazione di nuova conoscenza testuale, una capacità che li posiziona in modo unico per interagire con il corpus di conoscenze dell'EBM.<sup>12</sup>

## **1.3. Tesi: L'Emergenza dell'"EBM 2.0" - Un Paradigma di Collaborazione Uomo-Macchina**

Questo report avanza la tesi che l'integrazione degli LLM nella pratica clinica non rappresenta

un semplice miglioramento incrementale degli strumenti esistenti, ma catalizza un cambiamento di paradigma fondamentale, dando origine a ciò che può essere definito "EBM 2.0". Questo nuovo paradigma è caratterizzato da una relazione simbiotica in cui l'IA si fa carico dell'elaborazione e della sintesi rapida di volumi di dati altrimenti ingestibili, mentre il ruolo del clinico umano viene elevato. Il medico non è più solo un ricercatore e un applicatore di evidenze, ma diventa un supervisore critico, un arbitro etico e un decisore contestuale di ordine superiore.<sup>14</sup>

In questo modello collaborativo, l'obiettivo non è sostituire il giudizio umano, ma aumentarlo e potenziarlo. L'IA diventa uno strumento che affina il ragionamento senza comprometterne l'essenza, un "fonendoscopio cognitivo" che permette al clinico di "ascoltare" i segnali provenienti da un universo di evidenze in continua espansione.<sup>15</sup> Questa trasformazione modifica radicalmente il tempo della pratica basata sull'evidenza. Il processo tradizionale EBM, sebbene rigoroso, è spesso lento e ad alta intensità di lavoro, una barriera significativa alla sua implementazione capillare nella pratica quotidiana.<sup>14</sup> Gli LLM, capaci di eseguire compiti come la ricerca e il riassunto della letteratura in pochi secondi<sup>14</sup>, hanno il potenziale per trasformare il flusso di lavoro EBM da un processo lineare e spesso ritardato in un ciclo rapido, iterativo e quasi conversazionale. Un clinico può ora instaurare un dialogo in tempo reale con il corpus delle evidenze, ponendo una domanda, ricevendo una sintesi, affinando la domanda sulla base della risposta e ripetendo il ciclo quasi istantaneamente. Questo collasso temporale tra l'incertezza clinica e l'intuizione informata dall'evidenza promette di rendere l'EBM più dinamica e accessibile direttamente al punto di cura.

Di conseguenza, la sfida principale per i clinici si sta spostando. Storicamente, una competenza chiave nell'EBM era la padronanza di complesse strategie di ricerca in database come PubMed.<sup>3</sup> Oggi, con gli LLM e i sistemi di generazione aumentata dal recupero (RAG) che automatizzano e ottimizzano questo processo<sup>16</sup>, il carico cognitivo primario non è più il recupero delle informazioni, ma la loro *verifica e cura*. La competenza clinica essenziale diventa la capacità di valutare criticamente l'output dell'IA per accuratezza, bias, coerenza fattuale e applicabilità al singolo paziente. La sfida non è più "trovare l'ago nel pagliaio", ma "validare l'ago che la macchina ha trovato".

## **2. La Trasformazione Digitale del Flusso di Lavoro EBM in Cinque Fasi**

L'integrazione degli LLM e dell'IA sta ridefinendo ogni anello della catena metodologica dell'EBM. L'impatto non è uniforme; alcune fasi vengono quasi completamente automatizzate, mentre altre vengono potenziate, richiedendo al contempo nuove e più sofisticate competenze da parte del clinico. Questa sezione analizza, passo dopo passo, come la tecnologia sta rimodellando il processo EBM tradizionale.

### **2.1. Formulare il Quesito Clinico: Precisione ed Espansione del PICO Potenziata dall'IA**

Il punto di partenza di ogni indagine EBM è la traduzione di un problema clinico in un quesito ben definito e ricercabile. Il framework PICO (Popolazione/Paziente, Intervento, Comparatore, Outcome) è lo strumento standard per raggiungere questa strutturazione, garantendo che tutti gli elementi chiave di una domanda siano considerati.<sup>3</sup> Tradizionalmente, questo processo richiede una riflessione deliberata da parte del clinico.

Gli LLM stanno introducendo un nuovo livello di efficienza e precisione in questa fase iniziale. Sono in grado di analizzare testi non strutturati, come note cliniche, descrizioni di casi complessi o persino trascrizioni di conversazioni con i pazienti, ed estrarre automaticamente gli elementi PICO pertinenti.<sup>17</sup> Questa capacità assicura che il quesito sia ben formulato fin dall'inizio, riducendo il rischio di ricerche infruttuose o mal indirizzate. Sistemi avanzati, come il framework 'Quicker', dimostrano una capacità ancora più sofisticata di "scomposizione della domanda" (Question Decomposition). Questi sistemi possono prendere un'ampia domanda clinica e scomporla sistematicamente in componenti PICO multipli e strutturati, ciascuno pronto per una ricerca mirata nella letteratura.<sup>18</sup> Ad esempio, una domanda generica come "Qual è il miglior trattamento per un paziente con diabete di tipo 2 e malattia renale?" può essere scomposta in quesiti PICO specifici che confrontano diversi agenti ipoglicemizzanti (Interventi) rispetto a placebo o altri farmaci (Comparatori) su esiti specifici come il controllo glicemico, la progressione della malattia renale o la mortalità cardiovascolare (Outcome) in quella specifica popolazione.

### **2.2. Acquisire l'Evidenza: Gli LLM come Bibliotecari di Ricerca**

## **Automatizzati ad Alto Rendimento**

La fase di acquisizione dell'evidenza è forse quella in cui l'impatto degli LLM è più maturo e immediatamente percepibile. La sfida tradizionale per i clinici è duplice: il volume esponenziale della letteratura medica e il tempo limitato per consultarla. L'incapacità di rimanere aggiornati è una delle barriere più citate all'adozione dell'EBP.<sup>14</sup>

Gli LLM agiscono come bibliotecari di ricerca instancabili e ultra-rapidi. Possono navigare in modo efficiente attraverso vasti database di letteratura biomedica, fornendo ai clinici ricerche, linee guida e informazioni pertinenti in una frazione del tempo richiesto dai metodi manuali.<sup>5</sup> Il processo va oltre una semplice ricerca per parole chiave. Utilizzando gli elementi PICO scomposti nella fase precedente, gli LLM possono generare stringhe di ricerca booleane complesse e ottimizzate per database come PubMed.<sup>18</sup> Inoltre, possono operare in modo iterativo: eseguono una ricerca iniziale, analizzano i risultati e, se necessario, affinano autonomamente la strategia di ricerca per migliorare la pertinenza e la completezza, emulando il comportamento di un bibliotecario esperto.<sup>18</sup> Questa automazione affronta direttamente le barriere dei vincoli di tempo e del sovraccarico di informazioni, rendendo fattibile per un clinico impegnato accedere alle evidenze più recenti quasi in tempo reale.<sup>14</sup>

### **2.3. Valutare l'Evidenza: La Frontiera dell'Analisi Critica Automatizzata e del Rilevamento dei Bias**

Questa fase rappresenta il cuore intellettuale dell'EBM. Non è sufficiente trovare studi pertinenti; è necessario valutarli criticamente per determinare la loro validità interna (rischio di bias), la dimensione e la precisione dell'effetto del trattamento (impatto) e la loro generalizzabilità al proprio contesto clinico (applicabilità).<sup>1</sup> Questo compito richiede una profonda comprensione della metodologia della ricerca, della statistica e del contesto clinico.

Qui, le capacità attuali degli LLM incontrano i loro limiti più significativi. Sebbene un LLM possa riassumere accuratamente la sezione "Metodi" di un articolo scientifico e identificare elementi riportati (ad esempio, può affermare "lo studio era randomizzato e in doppio cieco"), fatica a compiere il passo successivo: la valutazione critica della qualità di tali metodi.<sup>14</sup> Un LLM, ad esempio, non è attualmente in grado di valutare se l'occultamento dell'allocazione (concealment of allocation) è stato adeguato, se il blinding è stato mantenuto con successo o se l'analisi statistica è appropriata per i dati. Questi giudizi sfumati, che sono al centro della valutazione del rischio di bias, rimangono dominio dell'esperto umano.

Tuttavia, questo campo è una frontiera attiva di ricerca. Esistono studi esplorativi che

utilizzano gli LLM per compiti correlati, come il rilevamento di bias razziali nel testo clinico<sup>19</sup> o l'assistenza ai revisori umani nel processo di valutazione della qualità. In questo scenario, l'IA non esegue la valutazione, ma agisce come un assistente, segnalando potenziali debolezze metodologiche (ad esempio, un alto tasso di perdita al follow-up) che richiedono un'ispezione più attenta da parte dell'uomo.<sup>20</sup> Questa collaborazione uomo-macchina, dove l'IA esegue uno screening preliminare e l'uomo la validazione finale, rappresenta la direzione più promettente. L'automazione completa della valutazione critica rimane, per ora, un obiettivo distante.

Questo divario funzionale sta portando a una separazione di due compiti cognitivi che tradizionalmente erano strettamente intrecciati. Nel flusso di lavoro EBM convenzionale, un ricercatore umano legge, sintetizza e valuta criticamente l'evidenza quasi simultaneamente; l'atto stesso di riassumere un articolo implica un giudizio sulla sua qualità. Gli LLM, invece, possono eseguire la sintesi su una scala e a una velocità massicce, producendo riassunti completi di decine o centinaia di studi in pochi minuti.<sup>17</sup> Tuttavia, questi riassunti sono generati senza una vera comprensione del rigore metodologico e sono soggetti a errori e distorsioni.<sup>14</sup> Ciò crea una nuova e distinta fase nel flusso di lavoro: il clinico riceve un corpo di evidenze pre-sintetizzate e deve quindi eseguire una *pura valutazione critica* sia del riassunto dell'IA sia del materiale originale sottostante. Il carico cognitivo si sposta dal "trovare e integrare" al "validare e criticare".

## **2.4. Applicare l'Evidenza: Il Ruolo dei Sistemi di Supporto alle Decisioni Cliniche Potenziati dall'IA (AI-CDSS)**

Una volta che l'evidenza è stata acquisita e valutata, deve essere applicata. Questa fase richiede l'integrazione dei risultati della ricerca con l'esperienza clinica e i valori e le circostanze uniche del paziente. I Sistemi di Supporto alle Decisioni Cliniche (CDSS) sono strumenti informatici progettati per facilitare questo processo, e l'integrazione dell'IA li sta rendendo esponenzialmente più potenti e dinamici.<sup>23</sup>

Gli AI-CDSS possono fornire raccomandazioni basate sull'evidenza direttamente all'interno del flusso di lavoro clinico, come la cartella clinica elettronica (EHR).<sup>6</sup> Ciò che li rende rivoluzionari è la loro capacità di personalizzare queste raccomandazioni. Un sistema tradizionale potrebbe semplicemente visualizzare un avviso basato su una linea guida generale. Un AI-CDSS, invece, può analizzare i dati specifici del paziente (storia clinica, risultati di laboratorio, dati genomici, comorbilità) e contestualizzare la raccomandazione basata sull'evidenza a quel profilo individuale.<sup>24</sup>

Un esempio pratico illustra questo potenziale: un'infermiera in un reparto di emergenza sta monitorando un paziente con segni aspecifici di malessere. Un AI-CDSS integrato nell'EHR

analizza in tempo reale i parametri vitali del paziente, i risultati di laboratorio e la sua storia clinica. Il sistema, avendo precedentemente sintetizzato le ultime linee guida sulla sepsi tramite un LLM, riconosce un pattern preoccupante che potrebbe sfuggire a un operatore oberato di lavoro. A questo punto, non si limita a lanciare un allarme generico, ma suggerisce i prossimi passi attuabili e basati sull'evidenza, come l'avvio di un protocollo di screening per la sepsi, raccomandazioni sulla rianimazione con fluidi o la richiesta di una consulenza infettivologica precoce.<sup>24</sup> Questo processo armonizza la ricerca più aggiornata con la pratica clinica in tempo reale, colmando il divario tra conoscenza e azione.

Questa capacità apre la porta a una "iper-personalizzazione dell'evidenza". Una delle sfide principali dell'EBM è sempre stata l'applicazione dei risultati di un RCT, condotto su una popolazione di studio media, a un singolo paziente che può differire significativamente da quella popolazione.<sup>3</sup> Gli AI-CDSS, integrando l'evidenza a livello di popolazione con il profilo dati unico di un paziente, possono tradurre una linea guida generale in un suggerimento personalizzato. Questo va oltre la semplice "applicazione dell'evidenza" e si avvicina a una vera e propria "medicina di precisione basata sull'evidenza", sebbene sollevi anche complesse questioni sulla validità di tali estrapolazioni e sulla necessità di una supervisione clinica ancora più attenta.

## 2.5. Valutare la Performance: Audit sull'Efficacia Clinica e Algoritmica

La fase finale del ciclo EBM, la valutazione (Assess), comporta una riflessione critica sull'esito della decisione clinica e sull'efficacia del processo stesso.<sup>1</sup> Tradizionalmente, questo significa chiedersi: "La decisione basata sull'evidenza ha migliorato l'esito del paziente? Come posso migliorare la mia capacità di formulare domande o di valutare la letteratura in futuro?".

L'introduzione dell'IA aggiunge una nuova, cruciale dimensione a questo audit: la valutazione della performance dell'algoritmo stesso. Non è più sufficiente valutare la propria performance; è necessario valutare anche quella del proprio "co-pilota" digitale. Per questo motivo, stanno emergendo quadri di valutazione dedicati come MedHELM, progettati per effettuare un benchmarking sistematico delle capacità degli LLM su un'ampia gamma di compiti medici, dal calcolo di punteggi di rischio alla generazione di note cliniche e alla previsione di esiti.<sup>19</sup>

Questo crea un potenziale circolo virtuoso di feedback. I dati sugli esiti clinici reali possono essere utilizzati non solo per valutare la qualità dell'assistenza, ma anche per affinare e migliorare i modelli di IA che la supportano. Ad esempio, se si osserva che le raccomandazioni di un AI-CDSS in un determinato scenario clinico non portano ai miglioramenti attesi, questi dati possono essere utilizzati per ricalibrare l'algoritmo. Questo processo di monitoraggio e miglioramento continuo è vitale per un'implementazione responsabile e sicura dell'IA in

medicina.<sup>25</sup>

La tabella seguente riassume la trasformazione del flusso di lavoro EBM, mettendo a confronto i metodi tradizionali con quelli potenziati dall'IA in ogni fase.

---

**Tabella 1: Il Flusso di Lavoro EBM Tradizionale vs. Potenziato dall'IA**

Fase EBM	Metodo Tradizionale	Metodo Potenziato dall'IA	Benefici Chiave dell'IA	Sfide/Rischi Chiave dell'IA
<b>1. Ask (Formulare)</b>	Formulazione manuale del quesito clinico utilizzando il framework PICO. Richiede conoscenza della metodologia e tempo per la riflessione.	Estrazione automatica degli elementi PICO da note cliniche non strutturate. Scomposizione di domande complesse in quesiti multipli e ricercabili. <sup>17</sup>	<b>Velocità e Precisione:</b> Garantisce quesiti ben formulati e completi, riducendo il tempo di preparazione.	<b>Dipendenza dal Contesto:</b> L'accuratezza dell'estrazione dipende dalla qualità e chiarezza della documentazione clinica di partenza.
<b>2. Acquire (Acquisire)</b>	Ricerca manuale per parole chiave in database biomedici (es. PubMed). Richiede la costruzione di complesse stringhe di ricerca booleane.	Generazione automatica e iterativa di strategie di ricerca basate sul PICO. Recupero e aggregazione rapida di letteratura umanamente impossibile da gestire. <sup>5</sup>	<b>Efficienza e Completezza:</b> Supera il sovraccarico informativo, accedendo a un volume di letteratura umanamente impossibile da gestire. <sup>14</sup>	<b>Recupero Subottimale:</b> Può ancora mancare documenti rilevanti o recuperare articoli non pertinenti se la strategia di ricerca non è ottimizzata. <sup>16</sup>
<b>3. Appraise (Valutare)</b>	Lettura e valutazione critica manuale di ogni articolo per validità, impatto e	Riassunto dei metodi di studio. Assistenza nel segnalare potenziali rischi di bias (es. alta	<b>Supporto alla Revisione:</b> Può accelerare il processo evidenziando aree che	<b>Gap nella Valutazione Critica:</b> Incapacità di giudicare la qualità

	applicabilità. Richiede competenze avanzate in metodologia della ricerca.	perdita al follow-up). La valutazione critica finale rimane umana. <sup>14</sup>	richiedono un'analisi umana più approfondita.	metodologica. Rischio di riassunti che mascherano difetti critici dello studio.
<b>4. Apply (Applicare)</b>	Integrazione mentale dell'evidenza con l'esperienza clinica e i valori del paziente. Applicazione di linee guida generali al singolo caso.	Sistemi di Supporto alle Decisioni Cliniche (AI-CDSS) che integrano l'evidenza con i dati del paziente in tempo reale, fornendo raccomandazioni personalizzate. <sup>24</sup>	<b>Personalizzazione e Tempestività:</b> Traduce l'evidenza a livello di popolazione in suggerimenti specifici per il paziente al punto di cura.	"Black Box": Algoritmi opachi possono rendere difficile per il clinico comprendere e fidarsi della logica dietro una raccomandazione. <sup>27</sup>
<b>5. Assess (Valutare)</b>	Audit clinico manuale e auto-riflessione sulla performance e sugli esiti del paziente.	Analisi automatizzata degli esiti clinici su larga scala. Benchmarking continuo della performance dell'algoritmo (es. MedHELM) per un miglioramento iterativo. <sup>19</sup>	<b>Feedback Loop:</b> Crea un ciclo di miglioramento continuo sia per la pratica clinica che per gli strumenti di IA.	<b>Metriche di Valutazione:</b> La definizione di metriche di successo appropriate per gli algoritmi clinici è complessa e ancora in fase di sviluppo.

### **3. Performance, Affidabilità e lo Spettro della Fallibilità Algoritmica**

Mentre il potenziale degli LLM per accelerare e migliorare il processo EBM è innegabile, la loro adozione clinica sicura ed efficace dipende interamente dalla loro performance e affidabilità. L'entusiasmo per le loro capacità deve essere temperato da una valutazione rigorosa dei loro limiti e dalla consapevolezza che, come ogni strumento, possono fallire. In un dominio ad alto rischio come la medicina, comprendere la natura e la frequenza degli errori algoritmici non è un dettaglio tecnico, ma una precondizione fondamentale per un'implementazione responsabile.

#### **3.1. Benchmarking delle Capacità degli LLM nei Compiti di Ragionamento Medico**

Le prime valutazioni delle capacità mediche degli LLM hanno spesso utilizzato come benchmark il superamento di esami di abilitazione standardizzati, come lo United States Medical Licensing Examination (USMLE), con risultati spesso impressionanti.<sup>19</sup> Sebbene questi risultati dimostrino una notevole capacità di memorizzazione e recupero di conoscenze fattuali, sono un indicatore profondamente fuorviante della prontezza clinica nel mondo reale. Come sottolinea uno studio, valutare un LLM sulla base di un esame è "analogo a valutare la capacità di guida di una persona utilizzando solo un test scritto sulle regole del traffico".<sup>19</sup> La pratica clinica non è un esame a scelta multipla; è un processo dinamico di ragionamento in condizioni di incertezza, che richiede calcolo, giudizio contestuale e interazione umana.

Riconoscendo questa limitazione, la comunità di ricerca si sta muovendo verso quadri di valutazione più olistici e clinicamente rilevanti. Framework come MedHELM mirano a testare gli LLM su un'ampia gamma di compiti pratici che i medici affrontano quotidianamente. Questi includono la capacità di eseguire calcoli clinicamente rilevanti (ad esempio, calcolare un punteggio HAS-BLED da una nota clinica), generare documentazione clinica coerente, prevedere rischi come la riammissione ospedaliera e persino rilevare bias impliciti nel testo clinico.<sup>19</sup> Una revisione sistematica che ha analizzato 761 studi sulla valutazione degli LLM in medicina ha rilevato che, sebbene l'accuratezza fosse il parametro più comunemente valutato, i quadri di valutazione utilizzati erano estremamente variabili e spesso privi di standardizzazione, rendendo difficile il confronto tra i modelli e la generalizzazione dei risultati.<sup>28</sup> Questi sforzi di benchmarking sono cruciali perché forniscono un quadro più realistico dei punti di forza e di debolezza dei modelli, guidando uno sviluppo più mirato e

un'adozione più cauta.

### 3.2. Una Tipologia degli Errori dell'IA nella Sintesi delle Evidenze: dall'Incoerenza Fattuale all'"Allucinazione"

Il rischio più critico nell'utilizzo degli LLM per l'EBM è la loro tendenza a generare output che, pur essendo linguisticamente fluenti e convincenti, sono fattualmente errati o non supportati dal materiale di origine. Questo fenomeno, spesso definito "allucinazione", non è un bug occasionale ma una caratteristica intrinseca del modo in cui questi modelli funzionano, generando testo sulla base di probabilità statistiche piuttosto che su una vera comprensione logica.<sup>14</sup> In un contesto medico, dove una singola informazione errata può avere conseguenze gravi, questo rischio è inaccettabile se non gestito.

Una ricerca sistematica sulla qualità dei riassunti di evidenze mediche generati da LLM ha permesso di definire una tassonomia dettagliata di questi errori, che va oltre il concetto generico di "allucinazione"<sup>22</sup>:

- **Errori di Interpretazione (Misinterpretation Errors):** Si verificano quando c'è una discrepanza logica tra il riassunto e il documento di origine. Questi si suddividono in:
  - **Distorsione della Conclusione (Conclusion Distortion):** Il riassunto trae una conclusione significativamente diversa o opposta a quella dello studio originale. Ad esempio, un riassunto potrebbe affermare che un trattamento è efficace, mentre lo studio concludeva per un effetto trascurabile.
  - **Illusione di Certezza (Certainty Illusion):** Il riassunto rappresenta in modo errato il grado di certezza dell'evidenza. Può presentare con grande sicurezza una conclusione basata su dati di bassa qualità (ad esempio, un piccolo studio osservazionale), o, al contrario, esprimere incertezza su un'evidenza robusta proveniente da una meta-analisi.
- **Errori di Fabbricazione (Fabricated Errors):** Il riassunto include informazioni, dati o conclusioni che sono completamente assenti dal documento di origine. L'LLM "inventa" dettagli per rendere il testo più coerente o completo.
- **Errori di Attributo (Attribute Errors):** Il riassunto commette errori su dettagli non centrali ma comunque importanti, come il numero di studi inclusi in una revisione, le caratteristiche della popolazione di studio o la durata del follow-up.

Questi errori sono particolarmente insidiosi perché gli LLM sono progettati per essere convincenti. Un'affermazione fabbricata o una conclusione distorta può essere presentata con lo stesso tono autorevole di un'informazione corretta, rendendo difficile per un lettore non esperto o distratto identificarla come un errore. Gli studi indicano inoltre che la probabilità di commettere questi errori aumenta con la lunghezza e la complessità del testo da riassumere,

suggerendo che i modelli faticano a mantenere la coerenza fattuale su contesti più ampi.<sup>22</sup>

La natura stessa di questi errori rappresenta una minaccia diretta alla filosofia fondamentale dell'EBM. L'EBM è, nella sua essenza, una disciplina per la gestione dell'incertezza e la comprensione della forza dell'evidenza.<sup>2</sup> La gerarchia delle evidenze è una gerarchia di certezza. Un errore come l'"Illusione di Certezza" non si limita a riportare un fatto sbagliato; sovverte l'intero quadro epistemologico dell'EBM. Presentando una conclusione debole con la stessa sicurezza di una robusta, l'LLM erode la capacità del clinico di ponderare correttamente le evidenze, portando a decisioni potenzialmente dannose basate su un falso senso di sicurezza.

### **3.3. Il Ruolo Indispensabile dell'"Umano nel Ciclo" per la Validazione e la Supervisione**

Dati i limiti attuali e i rischi intrinseci di fallibilità algoritmica, esiste un consenso schiacciatore nella letteratura scientifica e clinica sul fatto che gli LLM non possano e non debbano essere utilizzati in modo autonomo per compiti decisionali in ambito EBM. Il modello operativo indispensabile è quello dell'"umano nel ciclo" (human-in-the-loop).<sup>27</sup> Questo principio stabilisce che la tecnologia agisce come un potente assistente, ma la supervisione, la validazione e la responsabilità finale rimangono saldamente nelle mani del professionista umano.

In questo modello, il ruolo del clinico non è quello di un utente passivo che accetta l'output dell'IA, ma quello di un supervisore attivo che lo interroga, lo verifica e lo contestualizza.<sup>14</sup> Prima di utilizzare un riassunto generato da un LLM per informare una decisione clinica, il medico ha la responsabilità di:

1. **Verificare la coerenza fattuale:** Confrontare le affermazioni chiave del riassunto con i documenti di origine per assicurarsi che non ci siano errori di interpretazione, fabbricazione o attributo.
2. **Eseguire la propria valutazione critica:** Valutare in modo indipendente la qualità metodologica degli studi originali, poiché l'IA non è in grado di farlo in modo affidabile.
3. **Contestualizzare l'evidenza:** Integrare le informazioni (una volta validate) con la propria esperienza clinica e, soprattutto, con la situazione unica del paziente.

Questo approccio collaborativo cerca di ottenere il meglio di entrambi i mondi: l'efficienza e la capacità di sintesi dell'IA, e il giudizio critico, l'etica e l'empatia dell'uomo. Come afferma una fonte, l'IA non è il pilota, ma il "co-pilota di fiducia". La decisione finale deve rimanere saldamente nelle mani del clinico.<sup>24</sup> Questo modello non solo garantisce la sicurezza del paziente, ma crea anche un paradosso interessante. La promessa degli LLM è quella di ridurre

il carico di lavoro del clinico.<sup>14</sup> Tuttavia, poiché l'output dell'IA non può essere considerato attendibile implicitamente, il clinico deve ora eseguire un compito di validazione a più livelli: il riassunto è accurato? Riflette la fonte? La fonte stessa è valida? Questo compito di validazione può essere cognitivamente molto impegnativo, specialmente quando l'IA ha sintetizzato decine di fonti. Invece di leggere e valutare cinque articoli, il clinico potrebbe dover verificare un riassunto che afferma di basarsi su cinquanta. Paradossalmente, questo potrebbe aumentare il carico del compito cognitivo più difficile dell'EBM – la valutazione critica – automatizzando solo il compito più semplice del recupero.

## **4. Rimodellare la Cultura e la Pratica della Medicina Basata sull'Evidenza**

L'integrazione dell'IA e degli LLM non è un semplice aggiornamento tecnologico; è una forza che sta rimodellando la cultura, le norme professionali e le dinamiche interpersonali della pratica medica. L'impatto si estende oltre il flusso di lavoro tecnico per toccare l'identità stessa del clinico, la natura della relazione medico-paziente e le fondamenta della fiducia su cui si basa l'assistenza sanitaria.

### **4.1. Superare le Barriere Tradizionali all'Implementazione dell'EBP: Tempo, Volume e Accesso**

La pratica basata sull'evidenza, nonostante sia riconosciuta come il gold standard, affronta da sempre significative barriere alla sua implementazione nella pratica clinica quotidiana. Tra le più importanti vi sono la mancanza di tempo, carichi di lavoro pesanti e la schiacciante difficoltà di tenersi al passo con la crescita esponenziale della letteratura medica.<sup>14</sup> I clinici spesso si trovano di fronte a un dilemma: dedicare tempo prezioso alla ricerca e alla valutazione delle evidenze o alla cura diretta del paziente.

L'IA offre soluzioni dirette e potenti a queste sfide di lunga data. Automatizzando i compiti più dispendiosi in termini di tempo, come la ricerca e la sintesi della letteratura, gli LLM liberano i medici da ore di lavoro che possono essere reinvestite nell'interazione con il paziente e nel ragionamento clinico complesso.<sup>14</sup> Inoltre, l'IA può farsi carico di una parte significativa del carico amministrativo – come la programmazione, la fatturazione, la compilazione della documentazione e la gestione delle comunicazioni di routine – che spesso consuma una parte sproporzionata della giornata di un medico.<sup>12</sup> Questa riduzione del "lavoro di documentazione" (documentation burden) è vista come un passo cruciale per combattere il burnout e riportare l'attenzione del medico sul paziente.<sup>8</sup>

Oltre alla barriera del tempo, gli LLM possono aiutare a superare le barriere della conoscenza. La letteratura scientifica è spesso scritta in un linguaggio tecnico e prevalentemente in inglese, il che può rappresentare un ostacolo per alcuni professionisti. Gli LLM, con le loro capacità di traduzione, riassunto e semplificazione del linguaggio complesso, possono rendere le evidenze scientifiche più accessibili e comprensibili a una gamma più ampia di operatori sanitari in tutto il mondo, democratizzando di fatto l'accesso alla conoscenza basata

sull'evidenza.<sup>14</sup>

## 4.2. Il Ruolo in Evoluzione del Clinico: da Detentore di Conoscenza a Valutatore Critico

L'avvento di sistemi di IA che possono accedere e sintetizzare l'intero corpus della conoscenza medica in pochi secondi costringe a una profonda rinegoziazione dell'identità professionale del medico. Per secoli, il valore di un medico è stato strettamente legato alla sua capacità di possedere e richiamare una vasta quantità di conoscenze mediche. In un mondo in cui un algoritmo può superare qualsiasi essere umano in questo compito, il valore del clinico deve necessariamente risiedere altrove.

Il ruolo del medico si sta spostando da quello di "detentore di conoscenza" a quello di "validatore critico", "interprete contestuale" e "navigatore etico". Le competenze fondamentali che definiscono un buon medico nel 21° secolo sono sempre meno legate alla memorizzazione e sempre più al pensiero critico, al ragionamento in condizioni di incertezza, all'empatia e alla capacità di integrare l'output di un sistema di IA con la propria esperienza e la narrazione unica del paziente.<sup>14</sup> Il clinico diventa il supervisore finale del sistema, colui che pone le domande giuste, valuta la qualità delle risposte dell'IA e prende la decisione finale, assumendosene la piena responsabilità.

Questa trasformazione richiede un nuovo set di competenze, spesso riassunte sotto il termine "alfabetizzazione sull'IA" (AI literacy). I futuri medici dovranno essere formati non solo in anatomia e farmacologia, ma anche nei principi di base del machine learning, nella valutazione della qualità degli algoritmi e nelle implicazioni etiche dell'IA.<sup>31</sup> I curricula delle facoltà di medicina e i programmi di formazione continua devono adattarsi rapidamente per preparare una nuova generazione di professionisti in grado di collaborare efficacemente e in sicurezza con sistemi intelligenti, comprendendone sia il potenziale che i limiti.<sup>32</sup>

Questa evoluzione potrebbe avere un impatto inaspettato sul divario di performance tra i clinici. L'IA, lungi dall'essere un grande livellatore, potrebbe amplificare le differenze di competenza esistenti. Un clinico con solide basi di EBM e pensiero critico sarà in grado di sfruttare gli LLM come un potente moltiplicatore di forza, migliorando drasticamente la propria efficienza e accuratezza. Al contrario, un clinico con scarse capacità di valutazione critica potrebbe essere più incline ad accettare acriticamente l'output di un'IA, cadendo preda di errori sottili come l'"Illusione di Certezza".<sup>22</sup> Pertanto, l'IA potrebbe rendere i migliori clinici ancora migliori, ma potenzialmente abilitare errori pericolosi da parte di coloro meno attrezzati a supervisionare la tecnologia. Questo ha profonde implicazioni per la qualità dell'assistenza, la sicurezza dei pazienti e la formazione medica.

## **4.3. Costruire la Fiducia del Clinico e del Paziente nelle Raccomandazioni Algoritmiche**

La fiducia è la valuta fondamentale dell'assistenza sanitaria e rappresenta il fattore critico per il successo dell'integrazione dell'IA nella pratica clinica.<sup>20</sup> Senza la fiducia degli operatori sanitari e dei pazienti, anche gli algoritmi più avanzati rimarranno strumenti inutilizzati. Questa fiducia non è scontata e deve essere costruita attivamente attraverso un design trasparente, una validazione rigorosa e un'implementazione attenta.

Una revisione sistematica ha identificato otto temi chiave che influenzano la fiducia degli operatori sanitari nei confronti degli AI-CDSS.<sup>20</sup> Questi temi forniscono una tabella di marcia per lo sviluppo e l'implementazione responsabili:

1. **Trasparenza del Sistema:** La necessità di modelli di IA "spiegabili" (explainable AI) che permettano ai clinici di comprendere la logica alla base di una raccomandazione, superando il problema della "scatola nera" (black box).<sup>27</sup>
2. **Formazione e Familiarità:** L'importanza di una formazione completa che permetta agli utenti di comprendere le capacità e i limiti del sistema.
3. **Usabilità del Sistema:** L'integrazione fluida nei flussi di lavoro clinici esistenti, senza aumentare il carico di lavoro o l'affaticamento da allerta (alert fatigue).
4. **Affidabilità Clinica:** La coerenza e l'accuratezza delle performance del sistema in scenari reali.
5. **Credibilità e Validazione:** La dimostrazione che il sistema è stato rigorosamente testato e validato in contesti clinici diversi.
6. **Considerazioni Etiche:** La chiarezza sulla responsabilità medico-legale, l'equità e l'aderenza agli standard etici.
7. **Design Centrato sull'Uomo:** La priorità data ad approcci che supportano la relazione medico-paziente e il processo decisionale condiviso.
8. **Personalizzazione e Controllo:** La possibilità per i clinici di adattare lo strumento alle proprie esigenze e di mantenere l'autonomia decisionale finale.

Anche la fiducia del paziente è fondamentale. Gli studi dimostrano che l'atteggiamento dei pazienti nei confronti dell'IA in sanità è fortemente correlato alla loro fiducia preesistente nei confronti dei loro medici e del sistema sanitario in generale.<sup>35</sup> Questo suggerisce che l'IA non può essere introdotta in un vuoto relazionale. Anzi, l'IA può essere utilizzata per rafforzare la relazione medico-paziente. Ad esempio, gli LLM possono aiutare i medici a redigere risposte più empatiche, dettagliate e personalizzate ai messaggi dei pazienti, migliorando la qualità della comunicazione e rafforzando il legame di fiducia.<sup>36</sup> Tuttavia, un'eccessiva dipendenza dalla tecnologia o la percezione che l'IA stia mediando o sostituendo l'interazione umana

potrebbe, al contrario, erodere questa fiducia.<sup>38</sup>

Un rischio culturale a lungo termine è l'atrofia delle competenze. Le abilità EBM, come qualsiasi altra competenza, richiedono una pratica costante. Se gli LLM forniscono costantemente risposte rapide e plausibili per l'acquisizione e la sintesi delle evidenze, i clinici potrebbero diventare meno allenati a svolgere questi compiti da soli. Con il tempo, ciò potrebbe portare a un'atrofia delle competenze EBM di base, in particolare nella ricerca della letteratura e nella valutazione critica preliminare. Questo crea una dipendenza pericolosa: gli stessi strumenti progettati per supportare l'EBM potrebbero inavvertitamente indebolire le competenze umane necessarie per supervisionarli, rendendo i clinici più vulnerabili agli errori algoritmici.

La tabella seguente sintetizza i fattori chiave per la costruzione della fiducia in un quadro pratico e attuabile per le organizzazioni sanitarie e gli sviluppatori.

---

**Tabella 2: Quadro di Riferimento per Promuovere la Fiducia nei Sistemi di Supporto alle Decisioni Cliniche Basati sull'IA**

Fattore di Fiducia	Descrizione nel Contesto di un AI-CDSS	Barriere alla Fiducia	Raccomandazioni per gli Sviluppatori	Raccomandazioni per le Organizzazioni Sanitarie
<b>Trasparenza</b>	La capacità del sistema di spiegare la logica e le evidenze alla base delle sue raccomandazioni.	Algoritmi "black box" opachi; mancanza di chiarezza sulle fonti di dati utilizzate.	Implementare funzionalità di IA spiegabile (XAI); fornire riferimenti tracciabili per ogni raccomandazione basata sull'evidenza.	Esigere la trasparenza algoritmica nei contratti di acquisto; formare i clinici su come interpretare gli output del sistema.
<b>Affidabilità e Validazione</b>	Performance accurata, coerente e riproducibile del sistema, dimostrata attraverso rigorosi test	Mancanza di studi di validazione indipendenti e nel mondo reale; performance variabile in diverse	Condurre e pubblicare studi di validazione clinica rigorosi; essere trasparenti sui limiti di performance del	Istituire un processo di validazione interno prima dell'implementazione su larga scala; monitorare

	clinici.	popolazioni di pazienti.	modello.	continuamente la performance post-implementazione.
<b>Usabilità e Integrazione</b>	Integrazione fluida e intuitiva nel flusso di lavoro clinico esistente, senza aggiungere passaggi o oneri documentali.	Interfacce utente complesse; "affaticamento da allerta" (alert fatigue); interruzione del contatto visivo con il paziente.	Progettare in co-creazione con i clinici per garantire che lo strumento si adatti al flusso di lavoro; utilizzare un design intelligente per ridurre gli avvisi non necessari.	Coinvolgere gli utenti finali nella selezione e personalizzazione degli strumenti; fornire supporto tecnico continuo e ottimizzare i flussi di lavoro.
<b>Formazione e Familiarità</b>	Conoscenza approfondita da parte degli utenti delle capacità, dei limiti e del corretto utilizzo dello strumento.	Formazione insufficiente o puramente tecnica; mancanza di comprensione dei potenziali bias dell'IA.	Creare materiali di formazione chiari e basati su casi d'uso; fornire "sandbox" per la pratica sicura.	Sviluppare programmi di formazione completi e obbligatori; istituire "super-utenti" o campioni clinici per il supporto tra pari.
<b>Etica e Responsabilità</b>	Aderenza a principi etici consolidati, protezione della privacy dei dati e chiara definizione della responsabilità in caso di errore.	Ambiguità sulla responsabilità legale; preoccupazioni sulla privacy dei dati dei pazienti; potenziale di perpetuare bias sanitari.	Incorporare principi di "privacy by design"; condurre audit sui bias; essere trasparenti sulla governance dei dati.	Stabilire quadri di governance chiari per l'uso dell'IA; definire politiche sulla responsabilità; comunicare in modo trasparente con i pazienti.
<b>Controllo e Autonomia</b>	La capacità del clinico di mantenere il controllo finale sulle decisioni,	Sistemi che operano in modo troppo prescrittivo; percezione che	Progettare il sistema come uno strumento di supporto, non di sostituzione;	Rafforzare una cultura in cui il giudizio clinico ha la precedenza;

	<p>di ignorare le raccomandazioni dell'IA e di personalizzare lo strumento.</p>	<p>l'IA stia sostituendo il giudizio clinico.</p>	<p>consentire la personalizzazione delle preferenze e delle soglie di allerta.</p>	<p>incoraggiare la valutazione critica delle raccomandazioni dell'IA.</p>
--	---	---	--	---

---

## **5. Frontiere Etiche, Regolamentari e di Responsabilità nell'EBM Potenziata dall'IA**

L'integrazione dell'IA nella pratica clinica non solleva solo questioni tecniche e culturali, ma apre anche un complesso e inesplorato territorio etico, legale e regolamentare. I quadri di riferimento esistenti, concepiti per la medicina tradizionale e per tecnologie più semplici, sono spesso inadeguati a governare la complessità, l'adattabilità e la portata dei moderni sistemi di IA. Affrontare queste sfide è essenziale per garantire che i benefici dell'IA siano realizzati in modo equo, sicuro e responsabile.

### **5.1. Navigare tra Privacy dei Dati, Bias Algoritmico e il Problema della "Scatola Nera"**

Le fondamenta su cui si costruiscono gli LLM sollevano immediate e profonde preoccupazioni etiche. Questi modelli sono addestrati su quantità massive di dati testuali, spesso raccolti da internet, che possono includere informazioni sanitarie personali.<sup>39</sup> Ciò solleva questioni cruciali sulla privacy, sul potenziale utilizzo di dati di pazienti senza un consenso esplicito e informato e sul rischio, dimostrato, di poter re-identificare individui anche da dati apparentemente anonimizzati.<sup>40</sup>

Un secondo rischio etico fondamentale è il bias algoritmico. Se i dati di addestramento di un LLM non sono rappresentativi della diversità della popolazione di pazienti – ad esempio, se sottorappresentano determinate etnie, generi o gruppi socioeconomici – il modello risultante può perpetuare e persino amplificare le disparità sanitarie esistenti.<sup>31</sup> Un algoritmo addestrato prevalentemente su dati di una popolazione potrebbe fornire raccomandazioni meno accurate o inappropriate per pazienti di un'altra, portando a un'assistenza iniqua e a esiti peggiori per i gruppi già emarginati.<sup>30</sup>

Infine, la natura intrinsecamente complessa di molti modelli di IA, spesso descritta come una "scatola nera" (black box), pone una sfida significativa alla trasparenza e all'accountability.<sup>27</sup> Se un clinico non è in grado di comprendere il "perché" un algoritmo ha formulato una particolare raccomandazione, diventa estremamente difficile fidarsi di quella raccomandazione, spiegarla al paziente e, in ultima analisi, assumersene la responsabilità. Questa opacità è in diretta tensione con i principi di trasparenza e di processo decisionale

condiviso che sono centrali sia per l'etica medica che per l'EBM.

Queste sfide spostano il baricentro del rischio etico. Nella medicina tradizionale, le preoccupazioni etiche si concentrano in gran parte sulla diade medico-paziente e sulle decisioni prese all'interno di quella relazione. L'IA introduce un rischio etico di natura sistemica. Un bias nei dati di addestramento, la provenienza incerta dei dati o la contaminazione della proprietà intellettuale non sono problemi che emergono da una singola decisione clinica; sono rischi sistematici incorporati nella progettazione, nello sviluppo e nell'implementazione della tecnologia stessa.<sup>39</sup> Un algoritmo distorto non danneggia un singolo paziente; danneggia sistematicamente un'intera fascia demografica. Questo espande l'ambito della responsabilità etica oltre il singolo clinico per includere gli sviluppatori, i curatori dei dati, gli amministratori ospedalieri e i regolatori. Il "dovere di diligenza" etico si estende ora all'intero ciclo di vita del sistema di IA.

## **5.2. Il Dilemma di Collingridge: Sincronizzare la Regolamentazione con l'Avanzamento Tecnologico**

La velocità vertiginosa con cui si sviluppano gli LLM crea una sfida formidabile per i regolatori, perfettamente incapsulata dal Dilemma di Collingridge. Questo dilemma afferma che il tentativo di controllare una tecnologia è difficile perché nelle sue fasi iniziali, quando il controllo sarebbe facile, non si sa abbastanza delle sue conseguenze sociali dannose per giustificarlo; ma quando queste conseguenze diventano evidenti, la tecnologia è così profondamente radicata nella società che il controllo è diventato costoso, lento e difficile.<sup>40</sup>

Gli LLM esemplificano questo problema di "sincronizzazione". L'architettura tecnica unica di questi modelli li differenzia sostanzialmente da altre forme di IA precedentemente regolamentate, creando ostacoli normativi unici.<sup>39</sup> La loro "natura plastica", ovvero la capacità di apprendere dinamicamente e di evolvere in base agli input degli utenti e ai contesti clinici mutevoli, rende difficile una valutazione statica della sicurezza e dell'efficacia.<sup>40</sup> Questioni come la provenienza dei dati di addestramento e la "contaminazione della proprietà intellettuale" (cioè l'uso di materiale protetto da copyright nei dati di addestramento) complicano ulteriormente il panorama.<sup>39</sup> Un quadro normativo progettato per un dispositivo medico statico è inadeguato per governare un sistema di IA in continua evoluzione.

## **5.3. Definire la Responsabilità nell'Era delle Decisioni Cliniche Assistite dall'IA**

Forse la questione legale più spinosa e irrisolta è: chi è responsabile quando una decisione clinica assistita dall'IA provoca un danno al paziente? Il panorama della responsabilità è un intreccio complesso che coinvolge potenzialmente più attori<sup>42</sup>:

- **Lo sviluppatore del software:** Potrebbe essere ritenuto responsabile secondo i principi della responsabilità da prodotto (product liability) se si può dimostrare che l'algoritmo era "difettoso". Tuttavia, definire un "difetto" in un software complesso e probabilistico è legalmente arduo.<sup>44</sup>
- **L'organizzazione sanitaria (ospedale o clinica):** Potrebbe essere ritenuta responsabile per negligenza (vicarious liability) per le azioni dei suoi dipendenti, o per non aver fornito una formazione adeguata, per non aver validato correttamente lo strumento prima dell'uso o per aver scelto un sistema inappropriato.<sup>44</sup>
- **Il clinico individuale:** Rimane responsabile secondo i principi della responsabilità professionale medica (medical malpractice). Il clinico ha il dovere di utilizzare gli strumenti a sua disposizione secondo lo standard di cura, il che include la valutazione critica delle raccomandazioni di un'IA.<sup>42</sup>

I quadri giuridici attuali non sono stati concepiti per affrontare le complessità dei sistemi di IA adattivi. Determinare se la causa di un errore risiede nell'algoritmo, nei dati di addestramento, nell'implementazione del sistema o nell'interpretazione del clinico è una sfida formidabile che i tribunali non hanno ancora affrontato in modo sistematico.<sup>43</sup> Questa incertezza legale rappresenta una barriera significativa all'adozione, poiché sia i medici che le istituzioni temono di assumersi rischi legali non definiti.

Questa incertezza crea anche un paradosso relativo allo "standard di cura". Lo standard di cura è il livello di competenza e attenzione che un medico ragionevolmente prudente fornirebbe in circostanze simili, e una deviazione da esso è la base per una causa di malasanità.<sup>42</sup> Inizialmente, un clinico potrebbe essere ritenuto responsabile per aver accettato acriticamente una raccomandazione errata dell'IA.<sup>44</sup> Tuttavia, man mano che gli strumenti di IA diventano più validati, integrati e dimostrano di migliorare gli esiti, si potrebbe raggiungere un punto in cui *non consultare* un tale strumento in uno scenario clinico appropriato potrebbe essere considerato una deviazione dallo standard di cura.<sup>44</sup> I clinici si troveranno quindi in una posizione difficile, potenzialmente responsabili sia per essersi fidati dell'IA sia per averla ignorata. Questo rende lo sviluppo di linee guida professionali chiare e di "porti sicuri" normativi (regulatory safe harbors) assolutamente critico per il futuro.

La tabella seguente riassume le principali sfide etiche e regolamentari, identificando gli stakeholder e le possibili strategie di mitigazione.

**Tabella 3: Sfide Etiche e Regolamentari degli LLM in Medicina**

Area della Sfida	Descrizione del Rischio nel Contesto EBM	Stakeholder Coinvolti	Strategie di Mitigazione Potenziali
<b>Privacy e Sicurezza dei Dati</b>	Utilizzo di dati sanitari dei pazienti per l'addestramento senza consenso esplicito; rischio di de-anonimizzazione e violazione della privacy. <sup>39</sup>	Sviluppatori, Regolatori, Organizzazioni Sanitarie, Pazienti.	Tecniche di privacy differenziale; "privacy by design"; quadri di governance dei dati robusti; consenso informato granulare per l'uso dei dati.
<b>Bias Algoritmico ed Equità</b>	Le raccomandazioni basate sull'evidenza possono essere sistematicamente distorte contro popolazioni sottorappresentate nei dati di addestramento, esacerbando le disparità sanitarie. <sup>31</sup>	Sviluppatori, Clinici, Organizzazioni Sanitarie, Comunità di Pazienti.	Audit proattivo dei bias nei set di dati e nei modelli; utilizzo di set di dati di addestramento diversificati e rappresentativi; monitoraggio post-implementazione per l'equità degli esiti.
<b>Trasparenza e Spiegabilità</b>	La natura "black box" degli LLM impedisce ai clinici di comprendere il ragionamento dietro una raccomandazione, minando la fiducia e il processo decisionale informato. <sup>27</sup>	Sviluppatori, Clinici, Regolatori.	Sviluppo e adozione di tecniche di IA Spiegabile (XAI); richiesta che i sistemi forniscano le fonti di evidenza e il "razionale" per le loro conclusioni.

<b>Responsabilità e Colpa</b>	Ambiguità su chi sia legalmente responsabile (sviluppatore, ospedale, medico) quando un errore assistito dall'IA causa un danno al paziente. <sup>42</sup>	Sistema Legale, Regolatori, Assicuratori, Sviluppatori, Organizzazioni Sanitarie, Clinici.	Sviluppo di nuovi quadri giuridici per la responsabilità dell'IA; clausole contrattuali chiare tra sviluppatori e ospedali; linee guida professionali sullo standard di cura nell'uso dell'IA.
<b>Regolamentazione e Governance</b>	La rapida evoluzione e la "natura plastica" degli LLM superano la capacità dei quadri normativi tradizionali di garantire sicurezza ed efficacia continue (Dilemma di Collingridge). <sup>40</sup>	Regolatori (es. FDA, EMA), Policymaker, Sviluppatori.	Quadri normativi agili e adattivi; requisiti di sorveglianza post-commercializzazione (post-market surveillance) continui; standardizzazione dei processi di validazione e benchmarking.
<b>Integrità dell'Informazione</b>	Rischio di "allucinazioni" e generazione di informazioni mediche plausibili ma false, che possono inquinare il processo EBM e portare a decisioni errate. <sup>22</sup>	Clinici, Sviluppatori, Pazienti.	Integrazione di meccanismi di fact-checking; sistemi di generazione aumentata dal recupero (RAG) ancorati a fonti di conoscenza verificate; enfasi sulla formazione dei clinici alla validazione critica.

## 6. Conclusione e Traiettoria Futura

La convergenza tra la Medicina Basata sull'Evidenza e i Modelli Linguistici di Grandi Dimensioni segna l'inizio di una nuova era per la pratica clinica. L'analisi condotta in questo report dimostra che l'impatto degli LLM non è superficiale né limitato a un singolo aspetto del processo EBM, ma è una forza trasformativa che ne sta rimodellando ogni fase, dalla formulazione del quesito alla valutazione degli esiti. Stiamo assistendo all'emergere di un paradigma "EBM 2.0", caratterizzato da una simbiosi uomo-macchina in cui la velocità e la capacità di sintesi dell'algoritmo si fondono con il giudizio critico, l'etica e l'empatia del clinico.

Tuttavia, questo futuro promettente non è una conclusione scontata. Il potenziale rivoluzionario degli LLM è intrinsecamente legato a sfide significative. La fallibilità algoritmica, manifestata attraverso errori sottili ma pericolosi come l'illusione di certezza e la distorsione delle conclusioni, richiede una vigilanza umana costante e competenze di valutazione critica ancora più affinate. Le complesse questioni etiche, normative e di responsabilità devono essere affrontate con urgenza per costruire un ecosistema di fiducia in cui sia i medici che i pazienti possano sentirsi sicuri.

### 6.1. Un Quadro di Riferimento Proposto per l'Integrazione Responsabile dell'IA nell'EBM

Per navigare in questo nuovo paesaggio, è necessario un approccio strategico e multifattoriale. Un'integrazione responsabile dell'IA nell'EBM deve fondarsi su tre pilastri interconnessi:

1. **Validazione Rigorosa e Continua:** L'adozione di qualsiasi strumento di IA in ambito clinico deve essere preceduta da una validazione rigorosa, che vada oltre i semplici test di conoscenza per abbracciare quadri di valutazione olistici e clinicamente rilevanti come MedHELM.<sup>19</sup> Questa validazione non può essere un evento una tantum. Data la natura adattiva di questi modelli, è essenziale un monitoraggio continuo delle performance nel mondo reale (post-deployment monitoring) per garantire che la sicurezza e l'efficacia siano mantenute nel tempo e in diverse popolazioni di pazienti.<sup>47</sup>
2. **Design Centrato sull'Uomo e Collaborativo:** La tecnologia deve essere progettata per servire e aumentare il professionista, non per sostituirlo o ostacolarlo. Questo significa dare priorità a un design "human-centric" che integri l'IA in modo fluido nei flussi di lavoro clinici, riducendo il carico cognitivo anziché aumentarlo.<sup>20</sup> Gli strumenti devono essere concepiti fin dall'inizio con il modello "human-in-the-loop" come principio fondamentale, garantendo che il controllo e l'autonomia decisionale finale rimangano

sempre nelle mani del clinico.<sup>24</sup>

3. **Barriere Etiche e di Governance:** È imperativo stabilire solide barriere di protezione etica. Ciò include l'implementazione di quadri di governance chiari per la privacy e l'uso dei dati, la conduzione di audit sistematici per identificare e mitigare i bias algoritmici e la promozione della trasparenza e della spiegabilità dei modelli.<sup>31</sup> Senza queste fondamenta etiche, la fiducia, che è l'elemento vitale per l'adozione, non potrà mai essere pienamente stabilita.

## 6.2. Raccomandazioni Chiave per Clinici, Sistemi Sanitari, Sviluppatori e Policymaker

Il cammino verso un'EBM 2.0 efficace e sicura richiede un'azione coordinata da parte di tutti gli attori dell'ecosistema sanitario. Si propongono le seguenti raccomandazioni mirate:

- **Per i Clinici:** È fondamentale abbracciare un approccio di apprendimento continuo per sviluppare una solida "alfabetizzazione sull'IA". Questo non significa diventare scienziati dei dati, ma comprendere i principi di base, i punti di forza e i limiti di questi strumenti. La competenza più critica da coltivare è l'abilità di valutazione critica, non solo della letteratura medica, ma anche degli output generati dall'IA. Il clinico del futuro deve essere un supervisore esperto di sistemi intelligenti.
- **Per i Sistemi Sanitari e gli Educatori:** Le facoltà di medicina e i programmi di formazione specialistica devono integrare urgentemente l'IA, la scienza dei dati e l'etica digitale nei loro curricula.<sup>31</sup> Le organizzazioni sanitarie devono sviluppare linee guida istituzionali chiare sull'uso appropriato degli strumenti di IA, fornire una formazione completa e continua al personale e creare una cultura che incoraggi l'uso critico e riflessivo della tecnologia, piuttosto che un'accettazione passiva.<sup>20</sup>
- **Per gli Sviluppatori:** La responsabilità inizia dalla progettazione. Gli sviluppatori devono dare priorità alla trasparenza, alla validazione clinica rigorosa e a una stretta collaborazione con i professionisti sanitari durante l'intero ciclo di vita del prodotto. I sistemi devono essere costruiti presupponendo la supervisione umana, fornendo strumenti che facilitino la verifica e la comprensione, come il collegamento diretto alle fonti di evidenza e le spiegazioni della logica delle raccomandazioni.
- **Per i Policymaker e gli Organi Regolatori:** È necessario sviluppare quadri normativi agili e adattivi, in grado di tenere il passo con l'innovazione tecnologica senza compromettere la sicurezza dei pazienti. Questi quadri devono affrontare specificamente le sfide uniche poste dagli LLM, in particolare per quanto riguarda la responsabilità legale, i diritti sui dati, la gestione dei bias e la sorveglianza post-commercializzazione. Creare "porti sicuri" normativi può incoraggiare un'adozione responsabile, fornendo al contempo meccanismi chiari per la gestione degli errori e la protezione dei pazienti.<sup>40</sup>

In conclusione, l'unione tra EBM e IA non è semplicemente una questione di nuovi strumenti, ma di una nuova mentalità. Rappresenta un'opportunità senza precedenti per realizzare la promessa dell'EBM su una scala e con una velocità prima inimmaginabili, portando un'assistenza più informata, efficiente e personalizzata a ogni paziente. Tuttavia, questa opportunità può essere colta solo attraverso un impegno collettivo verso la validazione rigorosa, la supervisione esperta e una governance etica incrollabile. Il futuro della medicina non sarà guidato dall'uomo o dalla macchina, ma dalla qualità della loro collaborazione.

## 7. Bibliografia

1. South African Society of Anaesthesiologists. Evidence-based medicine: A primer for the clinician. *Southern African Journal of Anaesthesia and Analgesia*. 2022.<sup>3</sup>
2. Moore A, Gawlicki M. Evidence Based Medicine. In: *StatPearls*. StatPearls Publishing; 2024.<sup>1</sup>
3. Guyatt G, Rennie D, Meade M, Cook D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw-Hill Education; 2015.<sup>2</sup>
4. Clinic Case Quest. Evidence-based medicine. Published 2023.<sup>4</sup>
5. Wang G, Zhang Y, Xie Q, et al. The future landscape of large language models in medicine. *J Med Internet Res*. 2025;27(1):e59069.<sup>5</sup>
6. Stanford University Human-Centered Artificial Intelligence. Holistic Evaluation of Large Language Models for Medical Applications. Published 2024.<sup>19</sup>
7. Sallam M. The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: A Scoping Review. *J Med Sci*. 2025;12(6):631.<sup>6</sup>
8. Li Y, Wang J, Chen Y, et al. The application of large language models in medicine: A scoping review. *Int J Med Sci*. 2024;21(10):2792-2802.<sup>8</sup>
9. Wikipedia. Artificial intelligence in healthcare. Accessed 2024.<sup>10</sup>
10. Foreseemed. Artificial Intelligence in Healthcare: Reshaping the Future of Medicine. Published 2024.<sup>12</sup>
11. Built In. 29 Examples of Artificial Intelligence in Healthcare. Published 2024.<sup>13</sup>
12. Titu MA, Titu A, Titu S. The Role of Artificial Intelligence in Clinical Practice. *J Pers Med*. 2024;3(1):2.<sup>9</sup>
13. Zhu Y, Li Y, Wang X, et al. Benchmarking Large Language Models in Evidence-Based Medicine. *IEEE J Biomed Health Inform*. 2024;28(1):1-1.<sup>17</sup>
14. Lin C, Chen C, Lee CC, et al. Can Large Language Models Summarize Medical Evidence? A Systematic Evaluation of GPT-3.5 and ChatGPT. *arXiv*. 2023.<sup>22</sup>
15. Kanjee Z, Crowe B, Shah A. Creation and Adoption of Large Language Models in Medicine. *JAMA*. 2023;330(7):595-596.<sup>25</sup>
16. Li Y, Wang J, Chen Y, et al. The application of large language models in medicine: A scoping review. *iScience*. 2024;27(5):109713.<sup>11</sup>
17. Ali R, Connolly ID, Farrugia A, et al. The role of artificial intelligence in advancing evidence-based clinical practice, hospital management, and health policy: a systematic review. *J Hosp Manag Health Policy*. 2024;8:14.<sup>26</sup>
18. Esteves M, Bordei P, Pires G. Evidence-based practice and artificial intelligence: A new era in healthcare. *World J Methodol*. 2023;13(6):100-112.<sup>14</sup>
19. Jia R, Liang W, Li Y, et al. Trust in Artificial Intelligence-Based Clinical Decision Support Systems Among Health Care Workers: Systematic Review. *J Med Internet Res*. 2025;27(1):e69678.<sup>20</sup>

20. Sit C, Ran K, Padmakumar K, et al. Evidence-based recommendations for the integration of artificial intelligence in the undergraduate medical curriculum: A systematic review. *PLoS Digit Health*. 2023;2(6):e0000255. <sup>31</sup>
21. Wang Z, Zhang J, Li D, et al. RAG4EBM: A Retrieval-Augmented Generation Framework for Evidence-Based Medicine. *arXiv*. 2025. <sup>16</sup>
22. Ali R, Tang O, Auda J, et al. A systematic review of large language model (LLM) evaluations in clinical medicine. *medRxiv*. 2024. <sup>28</sup>
23. Li Y, Zhang Z, Wang S, et al. From Questions to Clinical Recommendations: Large Language Models Driving Evidence-Based Clinical Decision Making. *arXiv*. 2025. <sup>18</sup>
24. Wang G, Zhang Y, Xie Q, et al. The future landscape of large language models in medicine. *Commun Med (Lond)*. 2023;3(1):141. <sup>7</sup>
25. Gutierrez J, Savarese G. The Cognitive Stethoscope: Redefining the Art of Medicine in the Age of Artificial Intelligence. *JMIR Form Res*. 2025;9(1):e76669. <sup>15</sup>
26. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and its impact on healthcare delivery: a systematic review. *BMJ Open*. 2023;13(2):e068373. <sup>23</sup>
27. Zynx Health. The Future of AI in Clinical Decision Support: Advancements and Best Practices. Published 2024. <sup>24</sup>
28. RTI Health Solutions. Evidence Synthesis and AI Tools: Finding the Right Balance. Published 2025. <sup>21</sup>
29. NEJM Group. NEJM AI. Accessed 2024. <sup>48</sup>
30. Glass Health. Glass AI. Accessed 2024. <sup>49</sup>
31. Wolters Kluwer. How medical libraries can keep pace with artificial intelligence in healthcare. Published 2024. <sup>47</sup>
32. Discoveries in Health Policy. Journal Club: NEJM Compares AI and Statistical Methods. Published 2023. <sup>27</sup>
33. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. 2024;6(6):e428-e432. <sup>39</sup>
34. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. 2024;6(6):e428-e432. <sup>39</sup>
35. The Patient Safety Learning Hub. Ethical and regulatory challenges of large language models in medicine. Published 2024. <sup>50</sup>
36. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. 2024. <sup>51</sup>
37. Thirunavukarasu AJ, Sharma M, Thibault G, et al. The role of large language models in oncology. *BMJ Oncol*. 2025;4(1):e000759. <sup>29</sup>
38. Li Y, Wang J, Chen Y, et al. The application of large language models in medicine: A scoping review. *Int J Med Sci*. 2024;21(10):2792-2802. <sup>41</sup>
39. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. 2024;6(6):e428-e432. <sup>40</sup>
40. MedTech Europe. Liability challenges in AI medical technologies. Published 2022. <sup>44</sup>
41. Price II WN, Gerke S, Cohen IG. Liability for Use of Artificial Intelligence in Medicine. In:

- The Oxford Handbook of AI Governance.* Oxford University Press; 2024. <sup>42</sup>
42. Stanford University Human-Centered Artificial Intelligence. Understanding Liability Risk from Healthcare AI. Published 2024. <sup>45</sup>
43. Amato M, Ciliberti R, D'Errigo P, et al. AI and professional liability assessment in healthcare. A revolution in legal medicine? *J Public Health Res.* 2024;13(1):jphr.2023.1080. 52
44. Stanford University Human-Centered Artificial Intelligence. Understanding Liability Risk from Healthcare AI. Published 2024. <sup>46</sup>
45. Amato M, Ciliberti R, D'Errigo P, et al. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. *J Public Health Res.* 2023;12(4):jphr.2023.956. <sup>43</sup>
46. Ahsan M, Ekarat J, Sy A, et al. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC Med Educ.* 2025;25(1):1187. <sup>32</sup>
47. Ahsan M, Ekarat J, Sy A, et al. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC Med Educ.* 2025;25(1):1187. <sup>53</sup>
48. Khan RA, Jawaid M. Systematic Review: The Importance of Artificial Intelligence in Medical Education. *J Contemp Comm Pract.* 2025;2(1):1019. <sup>54</sup>
49. Parra-Perez C, Fernandez-Aleman JL, Toval A. AI in the Health Sector: Systematic Review of Key Skills for Future Health Professionals. *JMIR Med Educ.* 2025;11(1):e58161. <sup>33</sup>
50. Masters K, Cheema AR, Ahmad S, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach.* 2024;46(4):379-393. <sup>55</sup>
51. Blanchard M, Rdesinski S. Integrating artificial intelligence into medical education: a roadmap informed by a survey of faculty and students. *Teach Learn Med.* 2025;1-12. <sup>34</sup>
52. Council of Europe. Potential impact of AI on the doctor-patient relationship. Accessed 2024. <sup>38</sup>
53. Performance Health Partners. How AI Enhances Doctor-Patient Communication. Published 2024. <sup>36</sup>
54. Liaw W, Krist AH, Tong ST, et al. The Promise of AI: Enhancing Care, Equity, and Physician-Patient Interaction. *Ann Fam Med.* 2025;23(2):100-102. <sup>30</sup>
55. Miller LE, Lee C, Lignou S. How Will Artificial Intelligence Affect Patient-Clinician Relationships? *AMA J Ethics.* 2020;22(5):E436-441. <sup>56</sup>
56. UC San Diego Health. Study Reveals AI Enhances Physician-Patient Communication. Published 2024. <sup>37</sup>
57. Largent EA, Miller JE, Buniak L, et al. Expectations of healthcare AI and the role of trust: understanding patient views on how AI will impact cost, access, and patient-provider relationships. *J Am Med Inform Assoc.* 2025;32(5):795-801. <sup>35</sup>
58. IBM. Optimizing clinical trial site performance: A focus on three AI capabilities. Published 2024. <sup>57</sup>
59. Sharma A, Kumar A, Sharma R, et al. Artificial Intelligence (AI) Applications in Drug

Discovery and Drug Delivery: Revolutionizing Personalized Medicine. *Pharmaceutics*. 2024;16(11):1604.<sup>58</sup>

## Bibliografia

1. Evidence-Based Medicine - StatPearls - NCBI Bookshelf, accesso eseguito il giorno ottobre 14, 2025, <https://www.ncbi.nlm.nih.gov/books/NBK470182/>
2. What Is Evidence-Based Medicine? | Users' Guides to the Medical Literature - JAMAEvidence, accesso eseguito il giorno ottobre 14, 2025, <https://jamaevidence.mhmedical.com/content.aspx?bookid=847§ionid=69031458>
3. The principles of evidence-based medicine, accesso eseguito il giorno ottobre 14, 2025, <https://www.sajaa.co.za/index.php/sajaa/article/view/2914/3189>
4. Principles of evidence-based medicine | ClinCaseQuest, accesso eseguito il giorno ottobre 14, 2025, <https://clincasequest.hospital/evidence-based-medicine/>
5. Revolutionizing Health Care: The Transformative Impact of Large ..., accesso eseguito il giorno ottobre 14, 2025, <https://www.jmir.org/2025/1/e59069/>
6. Large Language Models in Healthcare and Medical Applications: A Review - MDPI, accesso eseguito il giorno ottobre 14, 2025, <https://www.mdpi.com/2306-5354/12/6/631>
7. Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine - PMC - PubMed Central, accesso eseguito il giorno ottobre 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11751657/>
8. Large Language Models in Medicine: Applications, Challenges, and Future Directions, accesso eseguito il giorno ottobre 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12163604/>
9. The Integration of Artificial Intelligence into Clinical Practice - MDPI, accesso eseguito il giorno ottobre 14, 2025, <https://www.mdpi.com/2813-0464/3/1/2>
10. Artificial intelligence in healthcare - Wikipedia, accesso eseguito il giorno ottobre 14, 2025, [https://en.wikipedia.org/wiki/Artificial\\_intelligence\\_in\\_healthcare](https://en.wikipedia.org/wiki/Artificial_intelligence_in_healthcare)
11. The application of large language models in medicine: A scoping review - PubMed, accesso eseguito il giorno ottobre 14, 2025, <https://pubmed.ncbi.nlm.nih.gov/38746668/>
12. Artificial Intelligence (AI) in Healthcare & Medical Field, accesso eseguito il giorno ottobre 14, 2025, <https://www.foreseemed.com/artificial-intelligence-in-healthcare>
13. AI in Healthcare: Uses, Examples & Benefits | Built In, accesso eseguito il giorno ottobre 14, 2025, <https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare>
14. New evidence-based practice: Artificial intelligence as a barrier ..., accesso eseguito il giorno ottobre 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10789101/>
15. Medicine Beyond Machines: Viewpoint on the Art of Thinking in the Age of AI, accesso eseguito il giorno ottobre 14, 2025, <https://formative.jmir.org/2025/1/e76669>
16. Enhancing LLM Generation with Knowledge Hypergraph for Evidence-Based

- Medicine, accesso eseguito il giorno ottobre 14, 2025,  
<https://arxiv.org/html/2503.16530v1>
- 17. Benchmarking Large Language Models in Evidence-Based Medicine, accesso eseguito il giorno ottobre 14, 2025, <https://pubmed.ncbi.nlm.nih.gov/39437276/>
  - 18. (PDF) From Questions to Clinical Recommendations: Large ..., accesso eseguito il giorno ottobre 14, 2025,  
[https://www.researchgate.net/publication/391776315\\_From\\_Questions\\_to\\_Clinical\\_Recommendations\\_Large\\_Language\\_Models\\_Driving\\_Evidence-Based\\_Clinical\\_Decision\\_Making](https://www.researchgate.net/publication/391776315_From_Questions_to_Clinical_Recommendations_Large_Language_Models_Driving_Evidence-Based_Clinical_Decision_Making)
  - 19. Holistic Evaluation of Large Language Models for Medical Applications | Stanford HAI, accesso eseguito il giorno ottobre 14, 2025,  
<https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications>
  - 20. Trust in Artificial Intelligence-Based Clinical Decision Support ..., accesso eseguito il giorno ottobre 14, 2025, <https://www.jmir.org/2025/1/e69678>
  - 21. Potential of AI Tools in Evidence Synthesis | RTI-HS, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.rtihs.org/updates-and-events/evidence-synthesis-ai-tools>
  - 22. Evaluating large language models on medical evidence ..., accesso eseguito il giorno ottobre 14, 2025, <https://PMC.ncbi.nlm.nih.gov/articles/PMC10449915/>
  - 23. Determinants of implementing artificial intelligence-based clinical decision support tools in healthcare: a scoping review protocol | BMJ Open, accesso eseguito il giorno ottobre 14, 2025,  
<https://bmjopen.bmj.com/content/13/2/e068373>
  - 24. Advancements in AI Driven Clinical Decision Support Systems | Zynx Health, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.zynxhealth.com/insights/ai-clinical-decision-support-advancements/>
  - 25. Creation and Adoption of Large Language Models in Medicine - PubMed, accesso eseguito il giorno ottobre 14, 2025,  
<https://pubmed.ncbi.nlm.nih.gov/37548965/>
  - 26. Integrating artificial intelligence in clinical practice, hospital management, and health policy: literature review - Nasef, accesso eseguito il giorno ottobre 14, 2025, <https://jhmp.americanmedicalgroups.org/article/view/9575/html>
  - 27. Journal Club: NEJM Compares AI and Traditional Statistics - Discoveries in Health Policy, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.discoveriesinhealthpolicy.com/2023/09/journal-club-nejm-compares-ai-and.html>
  - 28. (PDF) A systematic review of large language model (LLM) evaluations in clinical medicine, accesso eseguito il giorno ottobre 14, 2025,  
[https://www.researchgate.net/publication/389659170\\_A\\_systematic\\_review\\_of\\_large\\_language\\_model\\_LLM\\_evaluations\\_in\\_clinical\\_medicine](https://www.researchgate.net/publication/389659170_A_systematic_review_of_large_language_model_LLM_evaluations_in_clinical_medicine)
  - 29. Large language models in oncology: a review, accesso eseguito il giorno ottobre 14, 2025, <https://bmjoncology.bmj.com/content/4/1/e000759>
  - 30. The Promise of AI: Enhancing Care, Equity, and Physician-Patient Interaction, accesso eseguito il giorno ottobre 14, 2025,

<https://www.annfammed.org/content/promise-ai-enhancing-care-equity-and-physician-patient-interaction>

31. An evidence-based approach to artificial intelligence education for medical students: A systematic review | PLOS Digital Health - Research journals, accesso eseguito il giorno ottobre 14, 2025,  
<https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000255>
32. (PDF) Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions - ResearchGate, accesso eseguito il giorno ottobre 14, 2025,  
[https://www.researchgate.net/publication/394877649\\_Integrating\\_artificial\\_intelligence\\_into\\_medical\\_education\\_a\\_narrative\\_systematic\\_review\\_of\\_current\\_applications\\_challenges\\_and\\_future\\_directions](https://www.researchgate.net/publication/394877649_Integrating_artificial_intelligence_into_medical_education_a_narrative_systematic_review_of_current_applications_challenges_and_future_directions)
33. AI in the Health Sector: Systematic Review of Key Skills for Future Health Professionals - JMIR Medical Education, accesso eseguito il giorno ottobre 14, 2025, <https://mededu.jmir.org/2025/1/e58161/PDF>
34. Full article: Integrating artificial intelligence into medical education: a roadmap informed by a survey of faculty and students, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.tandfonline.com/doi/full/10.1080/10872981.2025.2531177?scroll=top&needAccess=true>
35. Expectations of healthcare AI and the role of trust: understanding patient views on how AI will impact cost, access, and patient-provider relationships - Oxford Academic, accesso eseguito il giorno ottobre 14, 2025,  
<https://academic.oup.com/jamia/article/32/5/795/8046745>
36. How AI Enhances Doctor-Patient Communication - Performance Health Partners, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.performancehealthus.com/blog/how-ai-enhances-doctor-patient-communication>
37. Study Reveals AI Enhances Physician-Patient Communication - UC San Diego Health, accesso eseguito il giorno ottobre 14, 2025,  
<https://health.ucsd.edu/news/press-releases/2024-04-15-study-reveals-ai-enhances-physician-patient-communication/>
38. V. Potential impact of ai on the doctor-patient relationship - Human Rights and Biomedicine, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.coe.int/en/web/human-rights-and-biomedicine/potential-impact-of-ai-on-the-doctor-patient-relationship>
39. Ethical and regulatory challenges of large language models in ..., accesso eseguito il giorno ottobre 14, 2025, <https://pubmed.ncbi.nlm.nih.gov/38658283/>
40. Ethical and regulatory challenges of large language models in medicine - ResearchGate, accesso eseguito il giorno ottobre 14, 2025,  
[https://www.researchgate.net/publication/380058006\\_Ethical\\_and\\_regulatory\\_challenges\\_of\\_large\\_language\\_models\\_in\\_medicine](https://www.researchgate.net/publication/380058006_Ethical_and_regulatory_challenges_of_large_language_models_in_medicine)
41. Large Language Models in Medicine: Applications, Challenges, and Future Directions, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.medsci.org/v22p2792.htm>

42. Liability for Use of Artificial Intelligence in Medicine - University of Michigan Law School Scholarship Repository, accesso eseguito il giorno ottobre 14, 2025, [https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1569&context=book\\_chapters](https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1569&context=book_chapters)
43. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review - PMC, accesso eseguito il giorno ottobre 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10711067/>
44. Liability challenges in AI medical technologies - MedTech Europe, accesso eseguito il giorno ottobre 14, 2025, [https://www.medtecheurope.org/wp-content/uploads/2022/10/medtech-europe\\_liability-challenges-in-ai-medical-technologies\\_document-paper\\_13-october-2022.pdf](https://www.medtecheurope.org/wp-content/uploads/2022/10/medtech-europe_liability-challenges-in-ai-medical-technologies_document-paper_13-october-2022.pdf)
45. Understanding Liability Risk from Healthcare AI, accesso eseguito il giorno ottobre 14, 2025, <https://hai-production.s3.amazonaws.com/files/2024-02/Liability-Risk-Healthcare-AI.pdf>
46. Understanding Liability Risk from Healthcare AI | Stanford HAI, accesso eseguito il giorno ottobre 14, 2025, <https://hai.stanford.edu/policy-brief-understanding-liability-risk-healthcare-ai>
47. How medical libraries can keep pace with artificial intelligence in healthcare, accesso eseguito il giorno ottobre 14, 2025, <https://www.wolterskluwer.com/en/expert-insights/how-medical-libraries-can-keep-pace-with-artificial-intelligence-in-healthcare>
48. NEJM AI - NEJM Group, accesso eseguito il giorno ottobre 14, 2025, <https://www.nejmgroup.org/featured/nejm-ai/>
49. Glass | AI Diagnosis & Clinical Decision Support (CDS), accesso eseguito il giorno ottobre 14, 2025, <https://glass.health/>
50. Ethical and regulatory challenges of large language models in, accesso eseguito il giorno ottobre 14, 2025, [https://www.pslhub.org/learn/commissioning-service-provision-and-innovation-in-health-and-care/digital-health-and-care-service-provision/288\\_artificial-intelligence/ethical-and-regulatory-challenges-of-large-language-models-in-medicine-23-april-2024-r11836/](https://www.pslhub.org/learn/commissioning-service-provision-and-innovation-in-health-and-care/digital-health-and-care-service-provision/288_artificial-intelligence/ethical-and-regulatory-challenges-of-large-language-models-in-medicine-23-april-2024-r11836/)
51. Ethical and regulatory challenges of large language models in medicine. | Semantic Scholar, accesso eseguito il giorno ottobre 14, 2025, <https://www.semanticscholar.org/paper/Ethical-and-regulatory-challenges-of-large-language-Ong-Chang/59325da69a3d47a30becc723000b2a44a3c9dbcd>
52. AI and professional liability assessment in healthcare. A revolution in legal medicine? - PMC, accesso eseguito il giorno ottobre 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10800912/>
53. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions - PubMed, accesso eseguito il giorno ottobre 14, 2025, <https://pubmed.ncbi.nlm.nih.gov/40849650/>
54. Systematic Review: The Importance of Artificial Intelligence in Medical Education,

- accesso eseguito il giorno ottobre 14, 2025,  
<https://www.jccpractice.com/article/systematic-review-the-importance-of-artificial-intelligence-in-medical-education-1019/>
55. A scoping review of artificial intelligence in medical education: BEME Guide No. 84, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.tandfonline.com/doi/full/10.1080/0142159X.2024.2314198>
56. How Will Artificial Intelligence Affect Patient-Clinician Relationships? - AMA Journal of Ethics, accesso eseguito il giorno ottobre 14, 2025,  
<https://journalofethics.ama-assn.org/article/how-will-artificial-intelligence-affect-patient-clinician-relationships/2020-05>
57. Optimizing clinical trial site performance: A focus on three AI capabilities | IBM, accesso eseguito il giorno ottobre 14, 2025,  
<https://www.ibm.com/think/topics/clinical-trial-optimization-ai>
58. Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine - PMC - PubMed Central, accesso eseguito il giorno ottobre 14, 2025,  
<https://pmc.ncbi.nlm.nih.gov/articles/PMC11510778/>