# EBM per l'AI: prove di efficacia per l'intelligenza artificiale in medicina

**Data** 11 giugno 2026
**Categoria** Medicinadigitale

Un recente editoriale sostiene che l'intelligenza artificiale in medicina si sta diffondendo molto rapidamente, ma che le prove del suo reale valore clinico sono ancora insufficienti

L'editoriale, pubblicato su Nature Medicine, affronta un tema di grande attualità: il rapporto tra innovazione tecnologica, medicina e prove scientifiche. Il testo parte da una constatazione evidente: gli strumenti basati sull'intelligenza artificiale stanno entrando sempre più rapidamente nei sistemi sanitari, dai modelli predittivi agli strumenti di supporto alle decisioni cliniche, fino ai Large Language Models (LLM) usati anche dai cittadini per ottenere informazioni mediche e consigli in autonomia. Tuttavia questa diffusione non è ancora accompagnata da prove sufficientemente solide del loro effettivo valore per pazienti, professionisti sanitari e organizzazioni sanitarie.
Il punto più importante sollevato dal testo riguarda la **differenza tra prestazione tecnica e impatto clinico reale** . Molti strumenti di AI vengono valutati attraverso metriche tecniche come sensibilità, specificità, calibrazione o capacità discriminativa, indicatori certamente utili, perché mostrano se un sistema è in grado di produrre risultati corretti dal punto di vista computazionale. Tuttavia, l'editoriale sottolinea che tali metriche non bastano a dimostrare che uno strumento migliori davvero la cura dei pazienti. Un algoritmo può ottenere risultati eccellenti in uno studio osservazionale retrospettivo, ma fallire nella pratica clinica se è difficile da interpretare, se interrompe il lavoro dei medici o se non viene usato correttamente
In ambito sanitario, non basta che una tecnologia sia nuova o sofisticata: deve dimostrare di produrre benefici clinici concreti. La medicina si fonda su un principio di responsabilità verso il paziente. Ogni intervento, farmaco, dispositivo o procedura deve essere valutato non solo in base alla sua efficacia teorica, ma anche in base alla sua sicurezza, alla sua efficienza e alle possibili conseguenze indesiderate. L'editoriale ricorda che, nel caso dei farmaci, esistono percorsi progressivi di valutazione e controllo prima che un beneficio clinico venga accettato. Per l'AI medica, invece, norme e standard condivisi sono ancora insufficienti o frammentari. Si corre il rischio di **adozioneprematura** . Se ospedali, aziende sanitarie o professionisti adottano strumenti di AI sulla base dell'entusiasmo o di promesse commerciali, senza evidenze adeguate, possono investire risorse in tecnologie dal valore incerto. Questo non è un problema solo economico: può avere effetti concreti sulla qualità dell'assistenza. Uno strumento mal integrato può aumentare il carico di lavoro, generare falsi allarmi, creare dipendenza da risultati automatizzati o introdurre nuovi errori nei processi decisionali. L'editoriale invita quindi a non confondere la disponibilità di una tecnologia con la sua reale utilità clinica.
Secondo gli autori, e noi condividiamo, il settore dell'AI medica ha bisogno di un quadro valutativo condiviso per collegare ogni promessa di valore clinico al tipo di prova necessario. Se si afferma che uno strumento migliora gli esiti dei pazienti o l'efficienza del sistema sanitario, servono studi solidi, prospettici e comparativi. Se invece si parla solo di prestazioni analitiche, possono bastare validazioni robuste nel contesto e nella popolazione di riferimento. Il principio centrale è quello della **proportionalevidence** : più forte è la promessa fatta da uno strumento di AI, più forti devono essere le prove che la sostengono. Non tutti gli strumenti devono necessariamente affrontare studi randomizzati complessi e costosi, ma non è scientificamente corretto considerarli affidabili solo perché mostrano buone prestazioni tecniche.
Particolarmente importante è anche il riferimento al **monitoraggio dopo l'implementazione** . Gli strumenti di AI, infatti, non sono statici: le loro prestazioni possono cambiare nel tempo a causa dell'evoluzione dei dati, delle popolazioni, delle pratiche cliniche o degli aggiornamenti del modello. Per questo motivo, l'editoriale sostiene che il controllo post-distribuzione non dovrebbe essere considerato un'aggiunta opzionale, ma una responsabilità istituzionale. Questo punto è decisivo perché, in medicina, la sicurezza non può essere garantita solo al momento dell'introduzione di una tecnologia: deve essere verificata continuamente.
Nel complesso, l'editoriale assume una posizione prudente ma non contraria all'AI. Il suo messaggio non è che l'intelligenza artificiale non debba essere usata in medicina, ma che il suo valore debba essere dimostrato con la stessa serietà richiesta ad altre innovazioni sanitarie. L'AI può certamente contribuire a migliorare diagnosi, prognosi, gestione dei pazienti e organizzazione dei servizi, ma solo se viene valutata nel contesto reale in cui opera. La tecnologia, da sola, non produce automaticamente progresso: il progresso nasce quando una tecnologia migliora in modo misurabile e sicuro la vita delle persone.

**Conclusioni**

L'adozione della IA procede più velocemente della verifica della sua utilità clinica reale. L'editoriale sottolinea la necessità di superare l'era dell'entusiasmo e delle promesse per entrare in una fase più matura, fondata su evidenze proporzionate, criteri condivisi e responsabilità collettiva. Senza un legame chiaro tra ciò che l'AI promette e ciò che dimostra, il rischio è che i sistemi sanitari adottino strumenti affascinanti ma non necessariamente utili. La vera sfida, la nostra responsabilità, non è solo costruire modelli più potenti, ma dimostrare che questi modelli producono valore clinico reale, sicuro e sostenibile.

**Giampaolo Collecchia e Riccardo De Gobbi**

**Riferimentobibliografico**

Show us the evidence for the value of medical AI. Nat Med 32, 1163 (2026). https://doi.org/10.1038/s41591-026-04389-4