quotidianosanità.it

Mercoledì 03 SETTEMBRE 2025

Algoritmi in corsia: tutela, paternalismo, nuova medicina difensiva?

Gentile Direttore,
la crescente e significativa integrazione dei Sistemi Artificiali Intelligenti (Artificial Intelligence Systems, AISs) nei processi clinici (Clinical Decision Support Systems, CDSSs; Diagnostic Decision Support Systems, DDSSs), ripropone e attualizza due tensioni classiche della bioetica: l'eterodirezione delle scelte del paziente (paternalismo) e le condotte dell'operatore sanitario per scopi di autotutela (medicina difensiva).

Queste tensioni si ibridano. Modelli predittivi e strumenti di supporto decisionale possono spingere a rifugiarsi verso l'accettazione di percorsi standardizzati nella prospettiva di un uso "difensivo" degli AISs. Delegando decisioni o accumulando test per poter documentare - se richieste - perizia, prudenza e diligenza.

Lo slippery slope (piano scivoloso) delle possibili rivendicazioni medico-legale tende sempre più a spostare l'attenzione verso un uso paternalistico e difensivo degli algoritmi. Ad esempio, spingendo verso una sempre maggiore priorità degli AISs per condividere o alleggerire responsabilità. Oppure moltiplicando esami biomedici guidati da alert algoritmici per documentare prudenza. Per quanto alcune recenti ricerche suggeriscono che un'assistenza basata su IA, se ben progettata, può anche ridurre l'overtreatment e migliorare l'accuratezza terapeutica.

Ecco la necessità di nuove piste operative per una governance della salute anti-paternalista e anti-difensiva.

In ambiente sanitario sempre più digitalizzato, la pretesa correttezza ascritta ai sistemi di IA può essere percepita come "oggettiva". Il rischio maggiore è che pazienti e medici tenderebbero a sospendere un giudizio critico conformandosi all'output del sistema. Questo significa, principalmente, erodere i fondamentali della relazione di cura, l'autodeterminazione informata, soprattutto quando l'opacità tecnica degli AISs rende difficile spiegare davvero come l'algoritmo abbia inciso sulla raccomandazione clinica.

L'explicability (spiegabilità) rappresenta un principio di riferimento, fondamentale, tra i nuovi principi applicati all'IA. Consentendo di "spiegare sia i processi tecnici di un sistema di IA, sia le decisioni umane collegate". Con l'explicability il sistema fornisce chiarimenti sulle cause delle decisioni (output); rende conto dei meccanismi di elaborazione e interpretazione dei dati; consente di valutare appropriatezza o inadeguatezza dei risultati prodotti.

Secondo L. Floridi, l'explicability deve essere inteso come un principio più ampio che indica una combinazione di intellegibility (risposta all'interrogativo epistemologico: come funziona?) e accountability (risposta all'interrogativo etico: chi è responsabile nel modo in cui funziona?). Eppure, "un sistema trasparente non è necessariamente un sistema spiegabile. La trasparenza dovrebbe essere intesa come una caratteristica passiva del sistema, mentre la spiegabilità richiederebbe uno sforzo attivo, da parte

del sistema, per rendersi comprensibile. In questo senso, garantire trasparenza non equivale automaticamente a garantire spiegabilità e viceversa".

Comunque, una spiegabilità sufficiente è condizione necessaria per un rapporto uomo-macchina che rafforzi, non sostituisca, il giudizio clinico né tantomeno la relazione di cura. È questo l'orientamento della human-centered medicine che in ambito bioetico e legale, sulla base di ragionevoli e condivise riflessioni, si va sviluppando in merito all'uso dell'IA in sanità.

Nel 2025 l'UE ha pubblicato il documento "Interplay between the Medical Regulation (MDR) & In vitro Diagnostic Medical Devices Regulation (IVDR) and the Artificial Intelligence Act (AIA)". Approvato dall'Artificial Intelligence Board (AIB) e dal Medical Device Coordination Group (MDCG). Il documento non ha valore giuridicamente vincolante ma certamente ha una sua propria specificità e rilevanza bioetica.

Dal documento si evince che per l'IA "ad alto rischio" incorporata in dispositivi medici sono previsti requisiti integrati su data governance, trasparenza, human oversight (attuazione di meccanismi progettati per bilanciare la tensione tra il concetto di "controllo" e quello di "autonomia" nella interazione tra gli esseri umani e sistemi di IA), documentazione tecnica unificata, monitoraggio post-market compresi meccanismi per rilevare bias. Tali obblighi, se ben implementati a livello di fornitore e deployer, possono spostare l'uso dell'IA in una pratica tracciabile e spiegabile. A supporto della relazione di cura, dell'autonomia dell'operatore sanitario, dell'autodeterminazione del paziente, del consenso informato e condiviso.

Come si rileva, il rapporto tra paternalismo algoritmico e medicina difensiva coinvolge i fondamentali della bioetica, della deontologia e, certo non ultimo, del biodiritto. Implicazioni bioetiche che possono essere ragionevolmente sintetizzate nei seguenti ambiti: consenso realmente informato, co-decisione responsabile, formazione anti-paternalistica.

Con il consenso realmente informato, il paziente deve sapere se, dove e come l'IA incide sul percorso, con una comunicazione proporzionata e comprensibile (incluse limitazioni, dati utilizzati, condizioni d'uso). La disclosure - intesa come divulgazione, rivelazione o comunicazione di informazioni in modo trasparente e completo - diventa parte integrante della relazione di cura e fonda il ricorso agli AISs.

Direttamente correlato al consenso è la co-decisione responsabile, ovvero la spiegabilità adeguata allo scopo (fit-for-purpose explainability). Una condizione assolutamente necessaria che riduce l'asimmetria nel rapporto triadico medico-paziente-sistema di IA che ha già superato da tempo quello classico, duale, medico-paziente. In questa nuova configurazione il terzo attore, rappresentato dall'IA e dalle tecnologie digitali di supporto decisionale, riformula profondamente le dinamiche di fiducia e responsabilità.

Per ultima, ma non ultima, la formazione anti-paternalista da sviluppare sia sul piano deontologico che bioetico. Competenze di lettura critica dell'output, gestione dell'incertezza e comunicazione del rischio consentono di preservare sia l'autodeterminazione del paziente che l'autonomia professionale. Paternalismo che è direttamente correlato all'automation complacency. Vale a dire quel fenomeno psicologico/cognitivo caratterizzato da un eccesso di affidamento e fiducia verso i sistemi automatizzati. Anche sottovalutando i rischi di errore o malfunzionamento.

In definitiva, i sistemi di IA offrono certamente maggiore efficienza nei processi diagnostico/terapeutici. Tuttavia, si sollevano questioni delicate. Tra queste, l'affievolirsi della vigilanza critica, l'impoverimento della relazione di cura e la possibilità di esiti meno soddisfacenti in termini di benessere globale della persona. Ecco la necessità della cooperazione. Integrando la tecnologia nella pratica clinica ma senza accantonare la centralità della competenza e della responsabilità umana nel processo decisionale. Perché il supporto decisionale si traduca in un autentico miglioramento della cura, è indispensabile bilanciare

l'elaborazione algoritmica dei dati con l'expertise situata ed esperienziale degli operatori sanitari, che custodisce sapere relazionale e discernimento etico. È questa la prospettiva per un'integrazione tecnologica che sia realmente rispettosa della dignità della persona. Che possa realmente favorire ciò che costituisce il nucleo irrinunciabile della cura: competenza, qualità delle relazioni, promozione della dimensione umana, intenzionalità condivisa.

E la macchina non conosce intenzionalità. Rileva P. Benanti: "l'intenzionalità è una proprietà che un computer, per quanto sofisticato, non può avere eseguendo un programma: la macchina è capace di μῆτις (ingegnosità pratica, *ndr*) ma il νοῦς (intelligenza intuitiva e creativa, *ndr*) sembra sfuggire a qualsiasi capacità computazionale".

Lucio Romano
Componente Commissione Scientifica - Centro Interuniversitario di Ricerca Bioetica (CIRB)