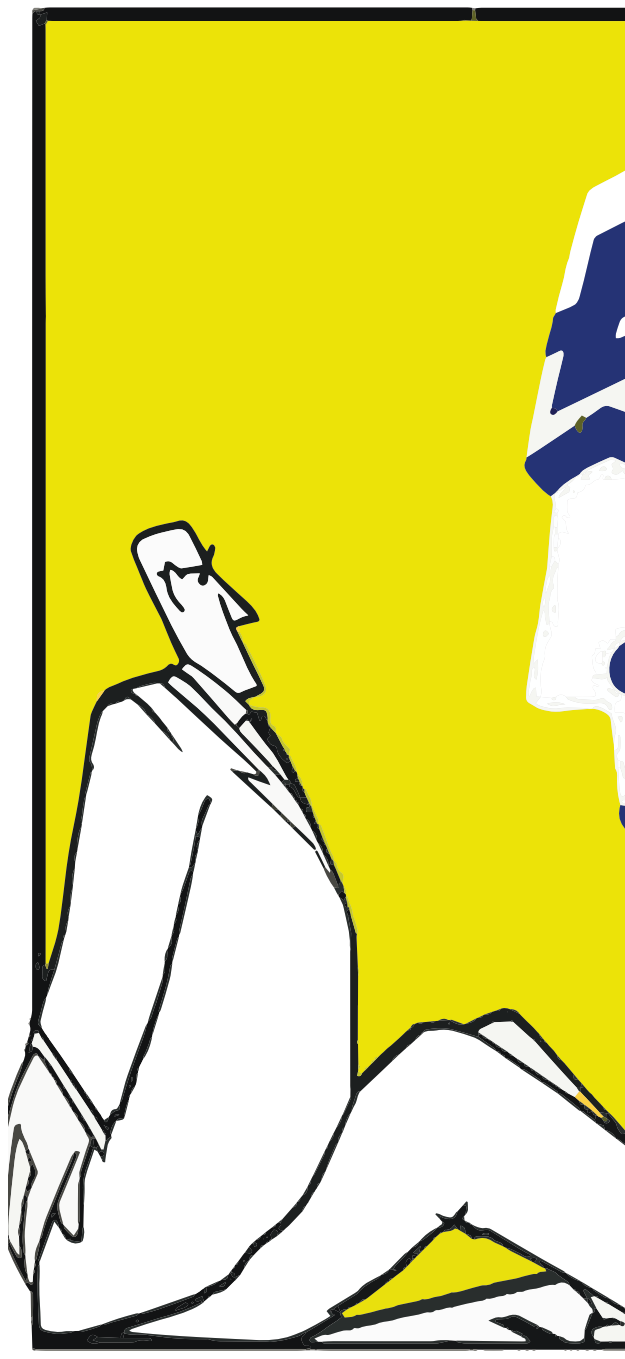
L'Intelligenza Artificiale in medicina non sostituisce il *medico, lo responsabilizza

Tra dati incompleti, bias nascosti e opacità, solo la responsabilità dei medici può trasformare l'algoritmo in un vero alleato



La medicina sta cambiando a una velocità enza precedenti

ACCANTO ALLA COMPETENZA del medico, oggi ci sono strumenti digitali capaci di elaborare in pochi istanti una mole impressionante di dati. Parliamo degli algoritmi di Intelligenza Artificiale (IA), in particolare quelli basati su modelli di machine learning e deep learning, che vengono addestrati su grandi dataset per riconoscere pattern, formulare ipotesi diagnostiche e supportare le decisioni terapeutiche. Promettono diagnosi più rapide, terapie più mirate e trattamenti personalizzati per ogni paziente.

Ma se la tecnologia corre, la nostra capacità di comprenderla arranca. Come ha scritto recentemente il filosofo Luciano Floridi su *Le Letture* – inserto culturale del *Corriere della Sera* – «questa rivoluzione digitale non è stato un processo lento e graduale: è arrivata come uno tsunami in pochi anni. Ha prodotto una tale accelerazione da superare di gran lunga la nostra capacità di comprenderla e governarla». È il paradosso dell'IA: uno strumento pensato per aiutare il medico può diventare una nuova fonte di incertezza. In altri termini, uno strumento pensato per portare chiarezza può finire per complicare le decisioni cliniche, finendo per generare nuovi dubbi invece di sciogliere quelli esistenti.

È proprio in questa zona d'ombra fatta di grandi potenzialità e di altrettante incertezze, che si gioca la sfida dell'IA applicata alla salute. Per questo motivo capire i punti di forza e i limiti di questi strumenti non è più un esercizio accademico, ma una necessità concreta. Da questa consapevolezza dipendono la qualità delle cure, la sicurezza dei pazienti e la fiducia nella relazione medico-paziente. È il motivo per cui dobbiamo chiederci: quanto l'IA in medicina è davvero affidabile? E quali fragilità rischiano di trasformare un aiuto prezioso in un ostacolo per la cura? Sono interrogativi che non riguardano solo i medici, ma ciascuno di noi: perché dalla loro risposta dipende il futuro stesso della medicina.

CARLO SBIROLI

Già direttore Uoc di Ginecologica Oncologia Istituto Nazionale dei Tumori "Regina Elena IFO", Roma

Past president Aogoi

Una medicina affidabile, ma non infallibile

DELINEATO IL QUADRO GENERALE, resta una domanda importante che ci dobbiamo porre: *quanto possiamo davvero fidarci dell'IA in medicina?* La risposta non risiede solo nella potenza degli algoritmi, ma soprattutto nella qualità dei dati che li alimentano.

Un algoritmo può essere sofisticato quanto si vuole, ma la sua capacità di fornire previsioni corrette dipende in primo luogo dai *dati* con cui è stato addestrato. Se questi non rappresentano in modo adeguato la popolazione a cui il modello viene applicato, il rischio di errore cresce. Nella pratica clinica ciò significa che un sistema può funzionare molto bene in un contesto, ma fallire in un altro, semplicemente perché i pazienti o le condizioni di partenza non corrispondono a quelli presenti nel suo set di addestramento.

È quanto ha mostrato il gruppo di Wu (2025): un algoritmo addestrato su pazienti di un singolo ospedale perdeva gran parte della sua accuratezza quando applicato in strutture con protocolli e popolazioni differenti. Un fenomeno simile è stato osservato anche da Joshi e collaboratori (2025) con un modello per la diagnosi precoce del carcinoma ovarico: calibrato su dati prevalentemente europei, risultava poco affidabile su pazienti asiatici o afroamericani.

Episodi come questi mettono in luce un concetto importante da tener presente: *in medicina il contesto clinico conta quanto, se non più, dell'algoritmo stesso*. Quando i dati sono limitati o non rappresentativi, l'affidabilità delle previsioni ne risente e il rischio di indicazioni fuorvianti aumenta. L'IA, dunque, può rappresentare un valido supporto, ma non è infallibile. La sua efficacia dipende dalla disponibilità di dataset ampi, diversificati e di qualità, e soprattutto dalla capacità critica del medico nell'interpretarli e integrarli nella pratica clinica.

Ma anche quando i dati sono solidi, resta un altro problema: spesso non siamo in grado di ricostruire il percorso logico che porta un modello alla sua conclusione. È in questi casi che emergono le maggiori criticità, legate alla mancanza di trasparenza dei sistemi intelligenti.

Opacità: quando l'IA non spiega

QUESTA MANCANZA DI TRASPARENZA prende il nome di *opacità*: una delle questioni più discusse dell'IA in medicina. Si tratta della difficoltà, se non addirittura dell'impossibilità, di comprendere il processo con cui un algoritmo giunge a un risultato, che può sembrare attendibile, ma il processo che lo ha generato resta oscuro. È come ricevere una diagnosi scritta in una lingua straniera non conosciuta: possiamo leggerla, ma non capirla davvero, né verificarla.

Il problema diventa evidente soprattutto nei sistemi più sofisticati, come quelli basati sull'apprendimento profondo (deep learning). Prendiamo l'esempio di un software che analizza un'ecografia ovarica: può indicare con alta probabilità se una massa è benigna o maligna, ma non dice quali caratteristiche siano state decisive per giungere a questo risultato, lasciando il medico senza una spiegazione comprensibile. Per il clinico significa trovarsi di fronte a un verdetto che non può essere discusso né confrontato con altri dati.

Questa opacità mina due pilastri della medicina moderna: il consenso informato e la responsabilità clinica. Come spiegare a un paziente che la sua diagnosi o la sua terapia si basano su un calcolo il cui funzionamento resta oscuro persino al medico? E come assumersi la piena responsabili-

tà di una decisione clinica se parte del processo resta invisibile? È una frattura che rischia di incrinare il rapporto di fiducia medico-paziente.
La comunità scientifica è consapevole del problema. Studi come quello di Sivaraman e collaboratori (2023) hanno mostrato come l'uso di sistemi opachi in oncologia complica la validazione clinica e aumenta il rischio di interpretazioni fuorvianti. Per questo la comunità scientifica spinge verso lo sviluppo di sistemi di *IA spiegabile*, capaci di accompagnare le previsioni con elementi di trasparenza e motivazione.
L'opacità, insomma, non è solo un problema tecnico da risolvere in laboratorio. È una questione etica e culturale che tocca il concetto stesso di medicina. Ignorarla significa accettare che la cura diventi un atto cieco, guidato da modelli che non possiamo comprendere né contestare. E tutto questo ci porta a un'altra insidia: anche quando un algoritmo appare estremamente preciso, il rischio è quello di cadere nell'illusione della precisione.

Illusione della precisione e interpretabilità

UNO DEI RISCHI più ingannevoli dell'IA in medicina, infatti, è proprio l'*illusione della precisione*. Grafici curati, numeri precisi fino al decimale, tabelle dettagliate: tutto sembra dare un'impressione di verità assoluta. Ma non sempre questi numeri raccontano tutta la storia. Se i dati di partenza sono sbagliati o incompleti, anche le previsioni migliori possono essere fuorvianti.
Da qui nasce la questione dell'*interpretabilità*. Non basta sapere che un algoritmo ha attribuito una certa probabilità a una diagnosi: è necessario capire quali variabili ha considerato, su quali dati ha costruito il suo calcolo, e in che modo ha collegato le informazioni per giungere alla sua conclusione. Senza questa trasparenza, anche il modello più evoluto rimane una "scatola nera" (sistema opaco che non mostra come arriva ai suoi risultati).
Un contributo importante su questo tema è arrivato recentemente da Dario Amodei, fondatore e CEO di Anthropic. In un saggio dal titolo *The Urgency of Interpretability* (2025) scrive che i sistemi di IA generativa «non sono progettati come il software tradizionale, ma "coltivati" attraverso enormi quantità di dati, sviluppando così «comportamenti spesso imprevedibili e difficili da decifrare». Da qui il suo avvertimento : «l'interpretabilità è una corsa contro il tempo. I modelli evolvono sempre più velocemente e stanno diventando più intelligenti, ma se non impariamo a capirli rischiamo di rimanere indietro». Un richiamo quanto mai attuale in medicina, dove ogni decisione può cambiare la vita di un paziente. Lasciare che questi strumenti restino incomprensibili non è solo un problema tecnico, ma un rischio etico che non possiamo permetterci.
Ma che cosa significa davvero per un medico, "interpretabilità"? Non è un tecnicismo da laboratorio, ma la possibilità di capire perché un algoritmo propone un certo risultato: quali dati hanno contato di più e in quale direzione. Se un sistema segnala un "alto rischio", deve anche mostrare i fattori che hanno portato a quella conclusione. Solo così smette di essere una scatola nera e diventa un alleato verificabile.
Alcuni studi recenti lo dimostrano. Il gruppo di Guha (2025), in oncologia ginecologica ha sperimentato strumenti capaci di spiegare le proprie valutazioni: invece di restituire un semplice punteggio, l'algoritmo evidenziava con mappe visive e indicatori clinici le caratteristiche che avevano spinto verso la diagnosi di benignità o malignità di una lesione ovarica. Il medico poteva così confrontare la logica del modello con la propria esperienza e con la specificità del caso.
Lo stesso criterio è stato adottato da Chen e colleghi (2025) in uno studio sul carcinoma ovarico sieroso di alto grado. Analizzando oltre 300 pazienti, il modello non solo stimava il rischio di recidiva, ma segnalava anche i parametri clinici, istopatologici e laboratoristici più determinanti. In questo modo, un numero astratto si trasformava in uno strumento clinico contestualizzato. Interpretabilità, in definitiva, non significa "aprire la pancia" di un algoritmo, ma restituire al medico il controllo: sapere quali fattori hanno orientato una previsione, discuterli, confrontarli con la storia del paziente.

> Uno dei rischi più ingannevoli dell'IA in medicina, infatti, è proprio l'illusione della precisione. Grafici curati, numeri precisi fino al decimale, tabelle dettagliate: tutto sembra dare un'impressione di verità assoluta

Oltre la trasparenza: criticità che non si possono ignorare

SE L'OPACITÀ DEGLI ALGORITMI è il lato più evidente del problema, esistono insidie più sottili, quasi invisibili, che rischiano di erodere la fiducia nella medicina digitale. Sono insidie silenziose, che non fanno notizia ma che possono cambiare il modo in cui la medicina viene praticata ogni giorno.
La prima riguarda il cosiddetto *clinician-in-the-loop*, cioè la presenza costante del medico nel processo decisionale dell'IA. Non si tratta di un tecnicismo, ma di un vero cambio di paradigma: il medico non deve limitarsi a eseguire ciò che "dice la macchina", bensì restare protagonista del processo, interagendo con l'algoritmo in ogni fase, dalla domanda iniziale alla verifica del risultato. Studi come quello di Kostick-Quenet e colleghi (2022) mostrano che questo approccio riduce gli errori e aumenta l'affidabilità diagnostica. In questo modo, l'IA non sostituisce, ma affianca il sapere clinico, riportando la responsabilità nelle mani di chi cura.
Un secondo rischio, meno evidente ma altrettanto concreto, è il *lock-in* tecnologico. Quando una struttura sanitaria adotta una piattaforma, il pericolo è quello di rimanere "prigioniera" di quel sistema e dei suoi aggiornamenti, anche se nel frattempo emergono soluzioni più avanzate. Maleki Varnosfaderani e Forouzanfar (2024) hanno messo in guardia da questa trappola: scelte iniziali poco lungimiranti possono trasformare un'opportunità in un vincolo strutturale, riducendo la libertà dei clinici e l'accesso a strumenti migliori, anche quando questi sarebbero più adatti a un caso specifico o a una nuova esigenza clinica.
Un'altra insidia, meno visibile ma importante per la qualità delle cure, è l'*overfitting* (sovradattamento). È la situazione in cui un modello "impara troppo bene" dai dati di addestramento, fino a memorizzarne dettagli e imperfezioni. Risultato: quando incontra nuovi casi, fallisce. È come uno studente che ripete a memoria la lezione senza aver davvero compreso la materia. Per evitarlo servono procedure rigorose, come la validazione su dati esterni, che permettono di capire se l'algoritmo funziona davvero nel mondo reale (Kernbach e coll., 2022).
Tre criticità diverse, ma unite da un filo comune: la potenza dell'IA può diventare vulnerabilità, se non viene gestita con attenzione. E quando queste fragilità - lock-in, sovradattamento, esclusione del clinico - si combinano con dati incompleti o distorti, il problema non è più solo tecnico. In questi casi l'apparente neutralità della macchina maschera il peso delle scelte umane e delle condizioni iniziali. Si trasforma in un'ingiustizia silenziosa, che assume le sembianze dell'oggettività. È proprio in questa ambiguità che si radica uno dei temi più delicati dell'IA in medicina: i bias.

Bias e disuguaglianze: quando l'algoritmo sbaglia senza che ce ne accorgiamo

UNO DEI RISCHI PIÙ INSIDIOSI dell'IA in medicina sono i *bias*, ovvero quelle distorsioni invisibili che si insinuano nei modelli e finiscono per influenzare diagnosi e trattamenti. A differenza degli errori clinici tradizionali, qui non c'è nulla di apertamente scorretto: il risultato può apparire coerente, persino sofisticato, e proprio per questo è più difficile da riconoscere, e da mettere in discussione.
Negli ultimi anni la letteratura scientifica ha iniziato a documentare casi sempre più numerosi. Il team di Smiley (2023), ad esempio, ha mostrato come un algoritmo di diagnostica radiologica fosse meno accurato nelle donne afroamericane rispetto a quelle caucasiche, semplicemente perché era stato addestrato su dati sbilanciati. Un problema simile è stato rilevato da Wang e Yang (2024) in ambito cardiovascolare: alcuni modelli predittivi tendevano a sopravvalutare il rischio

INTELLIGENZA ARTIFICIALE IN MEDICINA

nei pazienti più giovani e a sottovalutarlo in quelli anziani, con effetti concreti sulla scelta delle terapie preventive.
Questi non sono casi isolati. Al contrario, mostrano come i *bias* siano spesso radicati nel funzionamento stesso degli algoritmi, e se non riconosciuti possono amplificare, anziché ridurre, le disuguaglianze già presenti nel sistema sanitario. L'IA, pur essendo uno strumento potente, rischia così di diventare un moltiplicatore di errori, soprattutto quando i dati di partenza riflettono storture sociali o lacune cliniche.
Per questo, il tema dei *bias* ci obbliga a spostare lo sguardo dal piano tecnico a quello etico. Comprendere i limiti dell'IA significa anche ridefinire le responsabilità della pratica clinica, i criteri di trasparenza e le forme di tutela del paziente. Le implicazioni etiche non possono essere considerate un corollario teorico: sono parte integrante dell'azione medica quotidiana, e devono guidare ogni scelta che coinvolge strumenti algoritmici.

Un'etica operativa per il medico digitale

L'USO DELL'IA IN MEDICINA solleva questioni che vanno oltre la tecnica. Chi è responsabile di una decisione clinica presa con il supporto di un algoritmo? E come garantire trasparenza e tutela del paziente quando i processi digitali non sono sempre interpretabili?
A questi interrogativi hanno cominciato a rispondere società scientifiche e istituzioni sanitarie. La *American Medical Association* (AMA) ha ribadito l'importanza della supervisione umana. Il *Royal College of Physicians* ha richiamato alla tracciabilità delle decisioni algoritmiche. Anche il *Consiglio d'Europa*, la *World Medical Association* e la *Commissione Europea* – con l'AI Act e le linee guida dell'EHDS – sottolineano la necessità di coniugare innovazione, sicurezza e rispetto dei diritti del paziente.
Da questo confronto emergono alcuni principi condivisi: mantenere una supervisione clinica attiva, validare scientificamente gli algoritmi, garantire tracciabilità e accessibilità dei processi, prevenire ogni forma di discriminazione. Non si tratta di norme astratte, ma di strumenti per aiutare i medici a governare con consapevolezza un ecosistema tecnologico sempre più centrale nella pratica clinica.
L'etica dell'IA, in questo senso, non limita la medicina: la qualifica. È solo attraverso una responsabilità condivisa che l'IA può diventare un alleato, e non un sostituto, della relazione terapeutica.

CONCLUSIONI

L'IA STA TRASFORMANDO LA MEDICINA. Ma, per quanto potente, resta pur sempre un algoritmo: elabora dati, non conosce il contesto, non assume conseguenze delle decisioni, non compie scelte etiche. La sua efficacia clinica dipende dalla nostra capacità di interpretarla, verificarne le basi, metterne in discussione i risultati quando necessario.
Per affrontare questa sfida, serve una nuova cultura professionale: una cultura che, accanto al sapere medico tradizionale, sviluppi competenze capaci di riconoscere un errore algoritmico, valutarne le implicazioni etiche e discuterle apertamente con colleghi e pazienti. Una cultura digitale, sì, ma soprattutto critica, riflessiva, capace di dialogo.

OCCORRE ANCHE UN NUOVO LINGUAGGIO. Dire a una paziente "lo dice l'algoritmo" non è medicina: è deresponsabilizzazione. Le decisioni automatiche devono essere spiegate, discusse, condivise. Il consenso informato va ripensato: l'uso dell'IA deve essere illustrato in modo chiaro e comprensibile, accessibile a ogni paziente. Perché la fiducia non nasce dall'obbedienza cieca, ma dalla comprensione.
C'È POI UN'URGENZA che non possiamo ignorare. Come ha osservato Luciano Floridi, "le tecnologie digitali si sviluppano a una velocità che supera di gran lunga la nostra capacità di comprenderle e governarle". È questo squilibrio a rendere indispensabile un'assunzione di responsabilità collettiva. Non possiamo permettere che strumenti destinati a orientare le scelte cliniche restino opachi e incomprensibili.
Questo impegno deve coinvolgere tutti: i ricercatori, chiamati a investire nella comprensibilità degli algoritmi; le aziende, nella trasparenza; i legislatori, in regole chiare e condivise. Perché, per riprendere ancora Floridi, il rischio è quello di diventare "ospiti inconsapevoli di un mondo che non capiamo più".
La medicina del futuro non potrà mai essere guidata solo dalla tecnica: *deve restare fondata sulla responsabilità, sulla comprensione e sulla fiducia.*

> La medicina del futuro non potrà mai essere guidata solo dalla tecnica: deve restare fondata sulla responsabilità, sulla comprensione e sulla fiducia.

PER SAPERNE DI PIÙ

American Medical Association (2025). *Augmented Intelligence in Medicine – AMA Policy Statement*. https://www.ama-assn.org/practice-management/digital-health/augmented-intelligence-medicine.

Amodei D. The urgency of interpretability [Internet]. Anthropic Research Paper. 2025 [cited 2025 Aug 22]. Available from: https://www.darioamodei.com/post/the-urgency-of-interpretability

Chen Z, Ouyang H, Sun B, Ding J, Zhang Y, Li X. Utilizing explainable machine learning for progression-free survival prediction in high-grade serous ovarian cancer: insights from a prospective cohort study. *Int J Surg.* 2025;111(5): 3224-3234. doi:10.1097/JS9.0000000000002288. PMID: 39878156; PMCID: PMC12165540.

Council of Europe (2024). *Framework Convention on Artificial Intelligence and Human Rights*. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-council-europe.

Council of Europe (2024). *Report on the Application of AI in Healthcare*. https://www.coe.int/en/web/human-rights-and-biomedicine/-/report-on-the-application-of-ai-in-healthcare.

Floridi L. L'era digitale richiede responsabilità. *La Lettura* (Corriere della Sera).

2025, 20 Luglio; p.7.

European Commission (2024). *AI Act: Regulation on Artificial Intelligence*. https://www.nature.com/articles/s41746-024-01232-3.

European Commission (2025). *EHDS: European Health Data Space Regulation*. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en.

Guha S, Kodipalli A, Fernandes SL, Dasar S.. Explainable AI for Interpretation of Ovarian Tumor Classification Using Enhanced ResNet50. *Diagnostics* (Basel). 2024 Jul 19;14(14):1567. doi: 10.3390/diagnostics14141567. PMID: 39061704; PMCID: PMC11276149.

Joshi, S Wang Y, Lee J, Patel R, Kim H, Brown T, et al. AI as intervention: improving clinical outcomes through causal development and validation frameworks. *J Am Med Inform Assoc.* 2025;32(7):1245-1256. doi:10.1093/jamia/ocae301.

Kernbach JM, Staartjes VE. Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. *Acta Neurochir Suppl*. 2022;134:15-21. doi: 10.1007/978-3-030-85292-4_3. PMID: 34862523.

Kostick-Quenet KM, Martel ML, Bates DW, McCoy LG. AI in the hands of imperfect users: why medical education must include AI competencies. *NPJ Digit Med.* 2022;5:66. doi:10.1038/s41746-022-00579-y. PMID: 35611383; PMCID: PMC9126174.

Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. *Bioeng (Basel).* 2024;11(4):337. doi: 10.3390/bioengineering11040337. PMID: 38671759; PMCID: PMC11047988.

Royal College of Physicians (2018). *Artificial Intelligence in Health: Policy Position*. https://www.rcp.ac.uk/policy-and-campaigns/policy-documents/artificial-intelligence-ai-in-health

Sivaraman V, Bukowski LA, Levin J, Kahn JM, Perer A. Ignore, trust, or negotiate: understanding clinician acceptance of AI-based treatment recommendations. *NPJ Digit Med*. 2023;6:17. doi:10.1038/s41746-023-00792-2. PMID: 36646068; PMCID: PMC9843915.

Smiley A. Reátegui-Rivera CM, Villarreal-Zegarra D, Escobar-Agreda S, Finkelstein J. Exploring biases in artificial intelligence predictive models for cancer diagnosis. *Cancers (Basel).* 2025;17(3):407. doi:10.3390/cancers 17030407. PMID: 40103216; PMCID: PMC11782954.

Wang X, Yang CC. Enhancing multi-attribute fairness in healthcare predictive modeling. *arXiv* [preprint]. 2025 Jan 22. doi:10.48550/arXiv.2501.13219. arXiv:2501.13219.

World Medical Association (2025). *Statement on Augmented Intelligence in Medical Care*. https://www.wma.net/policy-tags/artificial-intelligence

Wu L Zhang H, Li Y, Chen Q, Zhao X, Wang J, et al. Artificial intelligence in ovarian cancer ultrasound diagnostics: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol.* 2025;65(2): 145-158. doi:10.1002/uog.27580. PMID: 39811245.