Life Studies, 2025, 1(3), 80-90
https://doi.org/10.71204/3cjmek14

Humanizing Medicine in the Age of Artificial Intelligence: Challenges, Transformations, and Prospects for Medical Humanities

Li Jia [1, *], Xianghan Du [2]

[1] Guangdong Medical University, Dongguan 523808, China

[2] Heilongjiang University, Harbin 150080, China

*** Correspondence:**

Li Jia

jiali@gdmu.edu.cn

Received: 31 October 2025 /Accepted: 9 November 2025 /Published online: 11 November 2025

Abstract

Artificial intelligence (AI) is rapidly reshaping clinical knowledge, workflow, and relationships, and it is doing so at a pace that presses the medical humanities to reinterpret their aims and methods. This article argues that, far from being peripheral to the algorithmic turn, the medical humanities are central to judging when, how, and under what conditions AI supports humane care. Drawing on scholarship from bioethics, science and technology studies, narrative medicine, and health services research, I first situate AI′s rise within long-standing debates about evidence, expertise, and the moral foundations of medicine. I then develop a critical analysis of the principal challenges AI poses for the human dimensions of care, including opacity and accountability, bias and justice, privacy and consent, erosion of clinical judgment and identity, and the risk of substituting datafication for meaning. In a parallel analysis, I identify opportunities where medical humanities can shape AI toward more trustworthy, equitable, and relationally sensitive practices: augmenting empathy and narrative attention with computational tools, reframing explainability as a communicative achievement rather than a technical property alone, embedding participatory design with patients and communities, renovating curricula to integrate critical data literacy with humanistic formation, and aligning governance with values such as dignity and solidarity. The article concludes by proposing a practical research and policy agenda in which humanities scholars collaborate with clinicians, patients, and engineers to evaluate AI not only by its predictive or operational performance but also by its contributions to understanding, moral repair, and shared decision-making in the everyday clinic.

Keywords: Medical Humanities; Artificial Intelligence; Algorithmic Bias; Narrative Medicine; Clinical Judgment

1. Introduction

The contemporary clinic is an environment increasingly saturated with algorithmic systems. Diagnostic classifiers interpret radiographs with uncanny speed, predictive models triage patients for scarce resources, conversational agents summarize visits and draft chart notes, and decision-support tools forecast deterioration hours in advance. Against this backdrop, the medical humanities—encompassing ethics, history, philosophy, literature, anthropology, and arts-based inquiry—face a two-sided question: How should they respond to the transformations AI is enacting in medicine, and what distinctive contributions can the humanities offer to shape these transformations toward the good? It is tempting to answer by appealing to an abstract notion of "human-centeredness" that counters technological exuberance with calls for empathy. Yet the stakes are more precise. AI intervenes in how illness is known, how responsibility is distributed, and how meaning is made from the irreducibly personal experience of being a patient or a clinician. These are the domains in which the medical humanities have long cultivated sensitivity and judgment.

The rise of AI does not begin from a vacuum. The evidence-based medicine movement habituated clinicians to regard statistical regularities as the gold standard for action, even as critics warned against reduction to decontextualized metrics. Machine learning can be seen as a continuation and intensification of this logic: it uses vast quantities of data to detect patterns that exceed unaided human perception, while often rendering the link between input and recommendation opaque to the user (London, 2019; Rajkomar et al., 2019). The biomedical promise is clear—earlier diagnosis, fewer missed signals, and greater consistency. Yet, as multiple observers have cautioned, predictive success at the population level is not synonymous with ethical acceptability at the bedside, especially where models encode historical inequities or displace forms of attention that matter for trust and understanding (Char et al., 2018; Obermeyer et al., 2019; Vyas et al., 2020).

Within popular discourse, AI is frequently framed as a tool for "freeing" clinicians to be more human by offloading repetitive tasks. This vision, while not false, insufficiently acknowledges that tools do not merely liberate; they reorganize the work and identities of those who use them. Clinical judgment may be reshaped as the "right" answer becomes what an algorithm predicts; empathic labor may be squeezed between documentation demands and the interpretive work that AI now performs. The humanities, attuned to the ways technologies have moral affordances and social lives, can illuminate how AI changes the grammar of care. They can show how algorithms invite certain ways of seeing, knowing, and valuing while occluding others; how they redistribute moral responsibility between humans and machines; and how they might be governed to sustain dignity, solidarity, and meaning (Floridi et al., 2018; Dignum, 2019).

This article proceeds on the premise that AI's clinical future will be decided not only by technical performance but by institutional choices about what kinds of evidence count, what harms we are willing to risk, and what aspects of care we wish to protect. The next chapter analyzes challenges that AI poses for medical humanities' core commitments: making sense of suffering, preserving moral agency, and fostering justice. The following chapter turns to opportunities—how humanities frameworks can partner with technical innovation to amplify the humane goods of medicine. The argument is not that AI will either save or ruin medicine, but that it is a site where the human and the technical are co-produced, demanding interpretive and ethical work commensurate with its power.

2. Challenges: Opacity, Justice, Privacy, Judgment, and Meaning in the Algorithmic Clinic

The most frequently cited challenge of clinical AI is opacity. Deep learning systems that achieve state-of-the-art performance often do so through representational strategies that resist straightforward explanation, yielding highly accurate but largely inscrutable outputs (London, 2019). From a humanistic point of view, the problem is not merely cognitive but moral. When an algorithm recommends discharging a patient or downgrades their priority for follow-up, who bears responsibility if harm ensues? Without intelligible rationales, clinicians cannot meaningfully contest or contextualize recommendations, and patients cannot evaluate whether reasons accord with their values. The right to an explanation, even if not legally guaranteed in all jurisdictions, signals a broader ethical need for justificatory narratives that connect decisions to reasons people can understand and challenge (Wachter et al., 2017). Attempts to retrofit "explainability" through post-hoc saliency maps or feature importance rankings address part of this need but risk giving an illusion of understanding where the underlying reasoning remains unexamined (Ghassemi et al., 2021). From the vantage of medical humanities, explanation is a communicative practice achieved in dialogue, not merely a technical property of a model; opacity thus implicates the conditions for moral relationship, not only the architecture of algorithms.

Closely allied to opacity is the risk of bias and injustice. Clinical AI inherits and can amplify the inequities inscribed in the data from which it learns. Obermeyer and colleagues (2019) demonstrated that an algorithm widely used to allocate extra care management systematically underestimated the needs of Black patients because it used health care spending as a proxy for illness—a measure already depressed by unequal access and structural racism. The result was a technocratic reinforcement of injustice under a veneer of objectivity. Vyas et al. (2020) catalogued race-adjusted algorithms in clinical practice that risk embedding racial essentialism in the guise of precision. Such cases highlight that fairness is not a post-hoc tweak but a design choice grounded in social theory and historical consciousness. The medical humanities, with their sustained engagement with structural determinants of health, are uniquely placed to articulate how data encodes histories of exclusion and how algorithmic optimization can be misaligned with justice. Yet the challenge deepens: even fairness metrics that aim to equalize error rates can collide with one another, and trade-offs must be made. Determining whose harms count and how to balance them is an ethical and political question that cannot be settled within code alone (Mittelstadt et al., 2016; Floridi et al., 2018).

A third cluster of challenges concerns privacy, consent, and the reconfiguration of intimacy. AI thrives on data density: continuous streams from wearables, sensor-rich hospital rooms, digitized narratives, and images harvested at scale. The boundary between clinical data and the textures of daily life becomes porous, inviting new forms of surveillance and secondary use. Traditional consent models—event-based, document-signed—strain under the continuous and inferential character of algorithmic analytics. Patients may consent to one use and find themselves subject to predictive inferences they did not anticipate or desire. Scholars have argued for governance that treats privacy not only as control over information but as a social value, connected to respect for persons and community norms (Price & Cohen, 2019). For the humanities, this means attending to how datafication can alter what it feels like to be a patient: to be watched by unseen systems, to have one's story parsed into features, to experience care through the lens of risk scores. The challenge is not simply compliance but the preservation of relationships marked by trust, appropriate vulnerability, and mutual recognition (Kerasidou, 2021).

The clinician's role and identity also come under pressure. Enthusiasts suggest that AI will relieve cognitive load, eliminate clerical drudgery, and standardize best practices. Yet the history of clinical technologies cautions that new tools often create new forms of work, particularly monitoring and exception-handling labor that remains invisible in planning documents but heavy in practice. If AI becomes a default recommender, clinicians may feel compelled to justify departures, a dynamic that tacitly shifts accountability from human to algorithm while also diminishing the perceived legitimacy of experiential knowledge. In training, students might learn to treat patient narratives as noise to be filtered rather than signals to be interpreted. The medical humanities have long emphasized the cultivation of practical wisdom—phronesis—that integrates technical knowledge with moral judgment attentive to particulars. A risk of AI-mediated care is the atrophy of this sensibility through over-reliance on statistically derived generalities (Beauchamp & Childress, 2019; Kleinman, 1988). The very shape of clinical reasoning could narrow, with model outputs crowding out slower, narrative modes of understanding. When a diagnostic suggestion appears with an aura of mathematical authority, it can displace the curiosity that might otherwise pursue discrepant clues and patient meanings.

Another set of challenges arises from the institutional embedding of AI. Health systems seeking efficiency may adopt models whose benefits are measured in throughput or documentation time saved rather than patient-centered outcomes. The humanities remind us that metrics shape action. When what is counted becomes what counts, phenomena resistant to measurement—comfort, dignity, the relational tissue of care—risk marginalization. Topol (2019) argues that AI could "restore" the human connection by freeing time; but without deliberate design and governance, freed minutes may be captured by productivity targets rather than conversation, leaving the relational promise unfulfilled. Moreover, safety science cautions that introducing new automation can generate unanticipated failure modes, including automation bias, deskilling, and brittle responses in rare cases—the very situations in which clinicians most need the capacity to improvise (Challen et al., 2019). Humanities-informed inquiry asks how institutions will cultivate cultures where clinicians can question model advice, where dissent is possible without sanction, and where responsibility is shared rather than offloaded.

Global and cross-cultural considerations compound these issues. AI models trained in data-rich settings may travel to low- and middle-income contexts with different disease burdens, infrastructures, and cultural meanings of illness. Importing tools without attention to local values and care practices risks epistemic injustice—silencing local knowledge in favor of algorithmic authority. The humanities, especially medical anthropology and cross-cultural ethics, can illuminate how categories embedded in datasets—diagnoses, symptoms, social risk—are not neutral but inherit specific histories. They can also challenge a universalizing narrative of AI "diffusion" by foregrounding co-design with communities and the notion that good care is plural, embedded, and negotiated. Without such attention, AI could widen global inequities by concentrating control over models and data in the Global North while exporting systems poorly matched to the sociocultural realities of receiving contexts (WHO, 2021).

Even where AI demonstrably improves certain outcomes, there is a danger of moral overclaiming: treating prediction as equivalent to understanding. Predictive accuracy does not resolve the interpretive task of making sense of illness in a life, nor does it determine what ought to be done. The humanities insist on the irreducibility of meaning, and they warn against mistaking algorithmic correlation for clinical explanation. This is not to reject prediction but to resist its colonization of other forms of knowing that patients and clinicians require for wise action. Greenhalgh and Papoutsi (2018) argue for embracing complexity—recognizing health care as a socio-technical ecosystem where interventions have nonlinear effects. An ethics of AI in medicine thus demands humility about what models can reliably deliver, attentiveness to context, and a posture of inquiry that remains responsive to surprise.

Finally, there is the challenge of governance: How should institutions evaluate AI beyond accuracy and calibration? Frameworks abound—principles of beneficence, nonmaleficence, autonomy, and justice; calls for transparency, fairness, accountability, and contestability (Beauchamp & Childress, 2019; Floridi et al., 2018; Jobin et al., 2019). Yet principles are inert without practices. The medical humanities can help articulate practices that render values operational: narrative review boards that examine patient stories alongside model validation; ethics rounds that include engineers and community representatives; consent processes that are dialogic and ongoing rather than transactional; documentation that records reasons for overriding algorithmic advice; and evaluation protocols that include ethnographic observation of workflow changes and relational effects. The hard challenge is sustaining these practices under the pressures of cost containment and administrative efficiency. Nevertheless, if the humanities do not insist that AI be answerable to the purposes of care, algorithms may come to define those purposes by default.

3. Opportunities: Re-centering Relation, Understanding, and Justice with Humanities-Informed AI

If the challenges are substantial, so too are the opportunities for medical humanities to guide AI toward more humane care. The first lies in reframing explainability as a relational achievement. Clinicians do not require every internal weight of a deep network to be legible; they need reasons they can responsibly present to patients, reasons that connect the recommendation to features of the case that matter to the person's goals and concerns. Humanities scholarship on communication, narrative, and practical reasoning can inform the design of interfaces and workflows that scaffold such sense-making. Instead of static "explanations," we might develop dialogic tools that allow clinicians to ask "what-if" questions, compare counterfactuals, and surface model limitations in language calibrated to the clinical conversation (Tonekaboni et al., 2019). This is an opportunity to integrate rhetorical and hermeneutic expertise into technical development so that explanation supports shared decision-making rather than merely satisfying compliance checklists.

Second, AI can be recruited to amplify, rather than erode, narrative attention. Natural language processing already extracts patterns from clinical notes and patient messages; with humane design, it can help clinicians notice overlooked meanings and social risks. Narrative medicine teaches clinicians to listen for metaphors of illness, for the moral stakes of a person’s story, and for the cultural scripts that shape expectations. Computational tools might, for example, flag shifts in a patient’s language that indicate despair or growing distrust, or summarize long timelines in ways that preserve the arc of a person’s experience rather than merely enumerating events. Such tools would require careful governance to avoid surveillance harms, but they point to a possibility: AI as a prosthesis for noticing, an ally in the interpretive labor that busy clinics can sideline. Kleinman’s (1988) insistence that illness is an interpretive object, not a purely biological state, offers a conceptual foundation for designing “narrative-sensitive” AI that respects the primacy of meaning.

Third, the medical humanities can help articulate standards for justice-aware design. The literature on algorithmic bias provides diagnostics; humanities-informed praxis can supply orientation. This includes embedding historical analysis at the outset of model development to anticipate how proxies might encode inequities; establishing community advisory boards with real authority over dataset construction and labeling; and defining success metrics that weight reduction in disparities as highly as overall accuracy (Obermeyer et al., 2019; Floridi et al., 2018; WHO, 2021). Participatory design methods, honed in anthropology and design studies, can be adapted so that those most at risk of harm, including minoritized patients, co-determine the goals of AI projects. Justice here is not an add-on but a through-line that shapes what problems are deemed worth solving. For example, rather than optimizing no-show predictions to penalize patients, systems could allocate resources for transportation or flexible scheduling, reflecting an ethics of solidarity. Medical humanities practitioners can help reframe datasets from mere technical artifacts into sites of moral choice.

Fourth, professional formation can be renovated to integrate data literacy with humanistic cultivation. Curricular experiments might pair courses in machine learning with seminars in philosophy of science and narrative ethics; clinical rotations could include “algorithm rounds” in which students practice articulating reasons for adopting or resisting a model’s advice, supported by the principles of beneficence, autonomy, and justice (Beauchamp & Childress, 2019; Dignum, 2019). Students could learn to read model cards as moral documents, to ask about training data provenance, to identify where values enter into labeling decisions, and to perform mini-ethnographies of workflow changes introduced by AI. The opportunity is to graduate clinicians who neither fetishize nor fear algorithms but situate them within a broader ecology of judgment. Kerasidou’s (2021) work on the enduring centrality of empathy and trust underscores that technical skill without relational competence undermines the very ends of medicine; training must therefore cultivate both together.

Fifth, the turn to AI invites reconsideration of evaluation. Medicine excels at measuring sensitivity, specificity, and net benefit; it is less adept at capturing dignity, comfort, and relational repair. Humanities-informed research can bring mixed methods to bear on AI deployments: ethnography to trace how tools reshape interactions; narrative analysis to understand how patients incorporate algorithmic advice into their identities and choices; and deliberative forums to surface contested values. Greenhalgh and Papoutsi (2018) call for complexity-sensitive evaluation; the medical humanities can help operationalize this by designing studies that follow AI over time, attending to adaptation, workarounds, and unintended consequences. Health systems could adopt dashboards that track not only model performance but also indices of trust, moral distress among clinicians, and equity impacts across subpopulations. The opportunity is to redefine "success" to include relational and justice outcomes that matter.

Sixth, AI can help return time to conversation if—and only if—institutions protect that time for care. Topol (2019) famously suggests that automation can restore the human connection; humanities scholarship can translate that aspiration into policy levers. For instance, documentation-drafting tools could be coupled with scheduling reforms that guarantee clinicians a portion of the reclaimed time for extended dialogue, with reimbursement codes recognizing communicative labor as clinical work. Narrative prompts integrated into electronic records could encourage clinicians to record the patient's "explanatory model" alongside the diagnosis, preserving meaning in the chart itself. Arts-based interventions—poetry, visual storytelling—might be supported by AI tools that curate reflective materials keyed to a patient's concerns, inviting clinicians to attend to the lived texture of illness. These proposals treat AI not as a time-saver plowed back into throughput but as a catalyst for realigning incentives with care's human ends.

Seventh, medical humanities can partner with technical communities to rethink explainability as design for contestability. The goal is not only to make models clear but to make them criticizable. Interfaces can surface data lineage and uncertainty, show where training data is sparse for a patient like this one, and offer pathways for clinicians to register disagreement that flows back into model governance. Such features enact accountability: they make it possible to say no, to ask for reasons, and to revise tools in light of practice. Social scientists can help design feedback loops where clinician and patient experiences become structured evidence in post-deployment monitoring, complementing quantitative surveillance of drift and performance (Mittelstadt et al., 2016; Jobin et al., 2019). The opportunity is a living ethics where models are not frozen artifacts but participants in a learning health system governed by human values.

Eighth, AI affords novel forms of compassionate anticipation. Predictive risk models, if governed justly, can identify patients likely to fall through the cracks, enabling proactive outreach that honors vulnerability. The humanities can shape the scripts by which such outreach is conducted, avoiding stigmatizing framings ("high-risk noncompliant") in favor of relational invitations ("we noticed you might need extra support—can we help?"). Privacy scholarship guides the design of consent processes that are ongoing and meaningful: granular permissions, opt-out by default for secondary use not directly related to care, and community governance for population-level analytics (Price & Cohen, 2019). Rather than letting risk scores trigger punitive measures, systems can embed social support referrals, aligning algorithmic foresight with compassion.

Ninth, the research ecosystem can incorporate humanities expertise upstream. Grant programs and institutional review processes might require interdisciplinary teams that include humanists with decision-making authority, not merely advisory roles. Model documentation could include "value statements" authored jointly by engineers and humanists, spelling out the moral aims and acceptable trade-offs of the tool. Ethics impact assessments could be published alongside performance papers, detailing steps taken to address bias, consent, and communicative transparency. The AI4People framework (Floridi et al., 2018) and the WHO guidance (2021) provide starting principles; medical humanities can translate them into practices and instantiate them in organizational routines. Over time, such integration could normalize a culture in which building a clinical model also means building a story about how it will support humane care.

Tenth, there is an opportunity to harness AI for reflexivity about medicine itself. By analyzing large corpora of clinical narratives, patient forums, and policy documents, AI can surface patterns that prompt critical reflection: how diagnostic categories evolve, which patient voices are marginalized, how institutional priorities shift the language of care. The humanities can interpret these findings, relating them to histories of professionalization, marketization, and sociopolitical change. This "AI-assisted critique" would not be a detached commentary but a practical tool for reform, identifying where systems drift from their professed values and where design could repair trust. In this sense, AI becomes both an object and an instrument of the humanities: a technology to be governed and a lens through which to see the moral life of medicine more clearly.

Realizing these opportunities requires pragmatic steps. Health systems can create interdisciplinary AI clinics where patients consent to enhanced observation of relational effects, with ethics oversight and community governance. Regulators can require evidence of equity impact and communicative adequacy as conditions of approval. Journals can solicit humanities-informed evaluations of deployments, not only technical metrics. Educators can develop competency frameworks that treat relational and ethical fluency as core to AI-era professionalism. Funders can prioritize projects that build tools for meaning-making—explainability for conversation, narrative-aware summarization—alongside diagnostic prowess. These steps do not romanticize the humanities as a panacea; they acknowledge that building humane AI is labor, and that such labor must be resourced and institutionalized.

Underlying all these opportunities is a simple claim: medicine is a moral practice oriented to the good of particular persons in particular communities. AI can support that practice when it is answerable to those goods. The medical humanities equip us with concepts—dignity, solidarity, vulnerability, narrative, justice—and with methods—close reading, ethnography, conceptual analysis, historical inquiry—that can keep algorithms aligned with what matters in care. In the best case, AI does not replace these humanistic competencies but becomes a partner that extends them, helping clinicians listen better, decide more wisely, and act more justly.

4. Conclusion

AI in medicine is neither destiny nor cure-all. It is a set of tools, institutions, and imaginaries that reorganize how we know, decide, and relate. The medical humanities have a vital role in steering these reorganizations toward humane ends. The challenges are real: opacity can corrode accountability; bias can conceal injustice beneath mathematical form; data hunger can erode privacy and intimacy; automation can deskill judgment and crowd out meaning. Yet the opportunities are equally tangible: to design explanations as conversations, to amplify narrative attention, to embed justice in the fabric of dataset construction and evaluation, to reform training so that data literacy and moral imagination grow together, and to align governance with values that medicine professes but does not always practice.

The work ahead is collaborative. Humanists, clinicians, engineers, and patients must build a shared language and shared practices for evaluating AI's contributions to care. Success will be measured not only in AUCs and workflow gains but in whether patients feel heard, whether disparities shrink, whether clinicians recover time and courage to engage the moral complexity of illness, and whether institutions cultivate a culture of contestability and repair. This article has sought to show that the medical humanities are not adjuncts to an already technical future; they are co-authors of the algorithmic clinic, capable of shaping AI to serve the goods of medicine. The task is urgent, but it is also hopeful: with deliberate design and governance, AI can become less a threat to humanistic care than a catalyst for its renewal.

Author Contributions:

All authors have read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of Interest:

The authors declare no conflict of interest.

References:

Beauchamp, T. L., & Childress, J. F. (2019). Principles of biomedical ethics (8th ed.). Oxford University Press.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. BMJ Quality & Safety, 28(3), 231–237.

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care — Addressing ethical challenges. The New England Journal of Medicine, 378(11), 981–983.

Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer. https://doi.org/10.1007/978-3-030-30371-6

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable AI in health care. The Lancet Digital Health, 3(11), e745–e750.

Greenhalgh, T., & Papoutsi, C. (2018). Studying complexity in health services research: Desperately seeking an overdue paradigm shift. BMC Medicine, 16, 95.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399.

Kerasidou, A. (2021). Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. Bulletin of the World Health Organization, 99(9), 605–606.

Kleinman, A. (1988). The illness narratives: Suffering, healing, and the human condition. Basic Books.

London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. Hastings Center Report, 49(1), 15–21.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. Social Science & Medicine, 260, 113172.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — Big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216–1219.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Price, W. N., II, & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25, 37–43.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. The New England Journal of Medicine, 380(14), 1347–1358.

Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. Journal of the Royal Society of Medicine, 112(1), 22–28.

Saria, S., Butte, A., & Sheikh, A. (2018). Better medicine through machine learning: What's real, and what's artificial? PLoS Medicine, 15(11), e1002699.

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want from explainable AI. npj Digital Medicine, 2, 90.

Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.

Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight — Reconsidering the use of race correction in clinical algorithms. The New England Journal of Medicine, 383(9), 874–882.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the GDPR. International Data Privacy Law, 7(2), 76–99.

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization.