OPEN ACCESS

Global perspectives on governing healthcare AI: prioritising safety, equity and collaboration

Ghasem Dolatkhah Laein

Correspondence to
Dr Ghasem Dolatkhah Laein, Mashhad University of Medical Sciences, Mashhad, Razavi Khorasan, Iran (the Islamic Republic of); dr.ghasemdolatkhah@gmail.com

Received 7 September 2023
Accepted 14 May 2024

INTRODUCTION

The adoption of artificial intelligence (AI) tools in healthcare administration and delivery continues to accelerate rapidly worldwide. Machine learning (ML) holds immense potential to enable transformative improvements in patient outcomes and care quality.[1] However, emerging evidence reveals significant risks if deployment proceeds without thoughtful oversight.[2] Studies demonstrate medical AI systems can propagate disparities based on race, gender, age and other factors through subtle pathways beyond just training data biases.[3] Additionally, lack of transparency into proprietary 'black box' systems poses challenges for evaluating real-world safety, effectiveness and equitable performance post-deployment.[4] These examples illuminate the growing need for judicious, cooperative governance to responsibly steer AI innovation in medicine.[5 6] While the European Union (EU) pioneered proactive oversight initiatives, regulatory regimes remain fragmented across most nations.[7] Realising AI's immense potential to improve care equitably worldwide requires gradually harmonising evidence-based approaches through collaboration among policymakers, companies, researchers and civil society.[6] Nuanced regulation can nurture development and adoption of AI systems that live up to their promise ethically across diverse patient populations.[8] This commentary analyses critical considerations for advancing responsible healthcare AI globally. It examines key challenges and limitations of existing frameworks to derive recommendations for pragmatic governance promoting safety, transparency and equity through cross-border cooperation. With inclusive innovation and oversight, medical AI can progress as a force for enhanced access and quality universally.

KEY RISKS AND CHALLENGES

Medical AI systems can propagate biases in multifaceted ways beyond just race or gender, with subtle root causes but serious impacts on equitable care.[3 9–11] Data representation biases arise when algorithmic models are trained on datasets that under-represent certain populations, systematically disadvantaging them.[10 12] For instance, an algorithm to predict 5-year mortality showed nearly 40% lower accuracy for black patients versus white patients due to under-representation of minorities in the underlying electronic health record data.[10] Even well-intentioned efforts to mitigate representation bias can paradoxically penalise underserved groups further.[9 13] Additionally, algorithm design choices can introduce bias even when training data are balanced across groups.[3 9] Failing to account for differences in comorbidities, social determinants and structural disadvantages between populations can skew predictions and recommendations in ways that propagate disparities.[9] This illustrates the subtle but impactful pathways through which biases emerge beyond just data representation issues.[3 11] These interconnected sources of algorithmic bias demand nuanced governance approaches.[9 11] While technical safeguards like high-quality datasets are crucial, oversight must also address root societal causes that infiltrate data and healthcare structures.[11] As leading AI ethicists argue, 'It takes an infrastructure to prevent an infrastructure from discriminating'.[11] Beyond fairness, lack of transparency in commercial algorithms severely constrains accountability.[6 14] Researchers have found that only 9 of 53 widely used medical AI systems allowed external validation of their performance claims.[14] Absent mandates, developers lack incentives to enable auditing that could reveal deficiencies.[6] The Food and Drug Administration's (FDA) attempt to evaluate one 'black box' AI diagnostic device was unfruitful given its inscrutable design.[6] This opaqueness leaves clinicians and regulators unable to properly assess hidden biases or monitor models over time.[6] Evidence-based governance balancing transparency and innovation is essential but complex given competing interests.[6 15] Approaches like holding proprietary algorithms' methods as trade secrets while mandating disclosure of validation studies offer potential paths forward.[6] Global harmonisation could also accelerate progress by preventing 'forum shopping' for lenient jurisdictions.[15] Ultimately, realising AI's promise requires reconciling public demands for accountability with developers' incentives to protect competitive advantages.[6 15]

GLOBAL LANDSCAPE ANALYSIS

While the EU mandates extensive transparency provisions for healthcare AI systems, the US FDA's flexible approach risks deficiencies going undetected according to experts.[6] The FDA's recent AI/ML software as a Medical Device Action Plan acknowledges the need for greater international cooperation and clarity to avoid fragmented policy landscapes.[16] Absent rigorous validation and monitoring, defective or biased algorithms could propagate inequities and cause preventable patient harm.[2 3] One study found a widely used algorithm that underestimated health risks for black patients was likely not adequately vetted for racial disparities before deployment.[3] Lack of transparency also

To cite: Dolatkhah Laein G. *BMJ Leader* Published Online First: [*please include* Day Month Year]. doi:10.1136/leader-2023-000904

erodes public trust in AI reliability.[17] Thought leaders argue opaque development and deployment processes necessitate strong oversight to ensure safety and accountability.[17 18] While the EU's proactive transparency and oversight mechanisms are a step forward, critics argue the regulations lack sufficient binding enforcement mechanisms.[18] The focus on premarket review also risks inflexibility in responding to issues arising post-deployment.[19] Though pioneering, the EU framework has gaps to address. Comparatively, the FDA's voluntary AI/ML software as a Medical Device Action Plan taps industry initiative but lacks binding enforcement mechanisms to compel accountability.[6] The plan's stated goals around Good Machine Learning Practice (GMLP) development, bias evaluation methods and real-world performance monitoring offer a framework for nuanced oversight.[16] However, absent mandates, developers may still lack sufficient incentives to fully enable reviews revealing potential deficiencies.[17] While the flexible approach avoids burdensome bureaucracy, experts note it relies heavily on corporate goodwill despite competitive disincentives.[6] In contrast, the EU's AI Act would implement binding requirements scaled by risk level.[18] For 'high-risk' AI like healthcare systems, mandatory conformity assessments and transparency obligations aim to address accountability gaps, though some question if these regulations will be responsive to issues arising post-deployment.[19] Analysis indicates the EU and FDA frameworks offer examples of nuanced attempts to balance innovation and oversight through tailored, cooperative approaches. However, experts underscore need to monitor their real-world efficacy at achieving safety and accountability goals.[17–19] However, international harmonisation faces political and practical barriers.[20] Competing national interests, intellectual property concerns and costs of aligning disparate regimes all pose challenges.[20] But multilateral collaborations like the WHO's new global AI health partnership may provide productive forums for finding common ground on core principles.[18] International cooperation is essential for translating principles into real-world impact.[17] Smaller-scale progress through transnational networks of experts and policymakers could also gain momentum towards responsible healthcare AI adoption worldwide.[17] In summary, preventing unsafe or unethical healthcare AI depends on transparency coupled with gradual multilateral alignment on core principles of accountability and equitable access.[17 18 21] Thoughtful regulation and global cooperation can enable AI to enhance care universally through gains in quality, safety and efficiency.

EXAMINING GOVERNANCE FOR ALGORITHMIC EQUITY

Effective governance is imperative to address algorithmic bias and ensure healthcare AI promotes equity. A critical perspective reveals important considerations across potential mechanisms including transparency, testing, auditing, collaboration and oversight. Transparency enables external scrutiny to uncover biases but appropriately scoping requirements remains challenging. Excessive transparency risks negative innovation incentives or privacy violations while limited visibility hampers accountability. Policymakers face difficult tradeoffs crafting rules balancing stakeholder interests. Pre-deployment testing methodologies are valuable but may also introduce new biases if datasets and benchmarks contain flaws. While testing surfaces some issues, real-world performance often diverges, underscoring the importance of ongoing monitoring after deployment. Independent auditing provides fresh perspective compared with internal testing. But legal and intellectual property (IP) barriers limit third-party access currently. Incentives and safe harbour protections may facilitate greater voluntary participation, but mandates may still prove necessary if progress lags. Collaboration on best practices offers benefits but views on acceptable bias levels still vary. Customised standards tailored to different use cases may suit the diversity better than one-size-fits-all edicts. Inclusive multistakeholder processes strengthen legitimacy and buy-in. Incentivising voluntary adherence to processes promoting fairness provides a flexible approach aligned with innovation culture. However, good faith alone risks 'ethics washing' without accountability mechanisms verifying outcomes. As Floridi argues, such washing concentrates on superficial ethical branding rather than enacting meaningful solutions.[22] Hybrid governance blending top-down rules and bottom-up collaboration may prove most effective.

STAKEHOLDER ROLES AND INCENTIVES

Aligning healthcare AI with ethical obligations requires reconciling differing stakeholder motivations shaped by political and commercial forces.[5 6] Companies prioritise rapid innovation, profits and IP protections.[23] But opacity around algorithms and data to protect IP can enable deficiencies and biases to go undetected without transparency mandates.[5 23–25] Proactive governance requiring explainability and validation, like the EU's documentation requirements, is vital though may be initially opposed.[5 23–25] However, heavy-handed restrictions may also limit progress without developer buy-in.[24]

Potential incentives include research and development tax credits, expedited reviews for voluntary accountability programmes and data sharing policies that protect IP.[2] The UK's collaborative development of AI best practices exemplifies multistakeholder cooperation.[24] Patients and providers can also champion reforms through advocacy and expert panels.[5 25] Groups like the American Medical Association (AMA) promote transparency, bias evaluations and human oversight.[25] But enabling inclusive public input remains challenging.[5 24]

The ethical implementation of healthcare AI fundamentally requires reconciling the inherent tensions between privatised technological development and public health responsibilities. While incentives and collaborations can partially align stakeholder interests, the profit-driven model underlying healthcare AI likely necessitates some degree of external regulation and oversight to prevent capitalistic motives from wholly capturing development. However, finding the right balance is challenging. Overly permissive approaches risk allowing potentially unethical products to proliferate until issues emerge post-market and public trust erodes. Yet, excessive restrictions may limit innovation and availability, depriving many of healthcare AI's benefits. In my view as a governance researcher, an evidence-based, iterative approach can help strike this balance. For instance, policy pilots allowing constrained rollout of new AI tools under heightened monitoring may help gauge real-world impacts. Data could inform continuously optimised regulation balancing innovation, accountability and access. Additionally, institutions like the WHO have published extensive guidance on the ethics of AI in healthcare, including recently on generative AI models. National health bodies also play vital roles synthesising cross-disciplinary academic research to derive best practices and core principles for healthcare AI. These can anchor binding regulations and voluntary programmes alike. Overall, the path forward likely requires policy humility—pragmatic trial-and-error governance evolution guided by multidisciplinary research illuminating healthcare AI's complex sociotechnical dynamics. With careful stewardship and research, the profound risks posed by healthcare AI's profit-based model can be mitigated to realise

its full equitable potential. But this will demand nuanced, adaptable and data-driven cooperation between stakeholders. Innovative deliberative formats like online platforms and weighted lotteries for representation may help.[23] In summary, cooperation across sectors coupled with incentives aligning social value and business value is key to optimising healthcare AI equitably.[5 23–25] Further research on optimal collaborative governance models is warranted.

REALISING HEALTHCARE AI'S EQUITABLE POTENTIAL: A STRATEGIC GOVERNANCE FRAMEWORK

Advancing healthcare AI responsibly poses complex challenges, but incremental progress is achievable through targeted governance efforts centred on collaboration, participatory processes and evidence-based oversight.[5 26 27] I propose policymakers pursue five strategic imperatives: first, implement premarket reviews calibrated to risks but expanded to require bias evaluations including analysis of training data representativeness and testing on diverse populations.[5 26 27] This prevents overburdening innovation while mandating assessments of system biases and disparate performance. Second, use public engagement mechanisms like citizens' juries to incorporate diverse perspectives, as research shows they enable inclusive deliberation on algorithm design and oversight.[5 26 27] Directly engaging affected groups through such structured formats will productively surface biases and needs that should shape governance. Third, cultivate global capacity through collaborative networks connecting health agencies, particularly in the developing world, with regional centres of excellence in AI safety and ethics.[27] Such partnerships can share expertise and co-create tailored frameworks addressing local priorities and constraints.[27] This complements essential top-down efforts by supranational organisations.[26] However, reticence around sharing sensitive technologies merits consideration. Fourth, augment premarket assessments with post-deployment surveillance infrastructure encompassing outcome monitoring, algorithm registration and adverse event reporting.[2 14 26 28] This creates responsive feedback channels enabling agile interventions when issues emerge in complex real-world contexts.[2 26 28] But lacking transparency into proprietary systems poses obstacles.[14 26] Policy creativity is required to enable oversight absent full disclosure.[14 26] Finally, support collaborative networks to develop best practices for bias detection techniques. Studies demonstrate approaches like clinical trial analyses stratified by social groups and independent algorithm reviews can improve bias evaluation.[3]

Developing and aligning on standardised bias detection methods will require tapping multidisciplinary expertise across stakeholders. Incentives tying preferential policies to voluntary best practices adoption can also help bridge gaps between business and social obligations.[2 26 29] However, scepticism persists regarding the efficacy of self-regulation.[26] Oversight mechanisms must therefore verify adherence and accountability. Realising equitable healthcare AI poses complex scientific challenges. Merely asserting aspirations would not progress solutions. Instead, we must illuminate core obstacles transparently. For instance, algorithmic biases emerge from myriad sources—insufficiently diverse training data, mismatching population data distributions, poorly constructed feature engineering and judgement criteria, among others. And biases propagate in equally multifaceted ways. Solely solving discrete technical issues often proves inadequate. Truly overcoming algorithmic biases demands a systems perspective encompassing the sociotechnical complexities of how data are generated, algorithms are constructed and outputs are used in real-world contexts. This necessitates cross-disciplinary expertise spanning computer science, healthcare, ethics, social science, regulatory policy and beyond. Collaborative, participatory processes engaging diverse stakeholders can surface insights and constraints that technologists alone may overlook. Community partnerships, advocacy boards and participatory design are invaluable but underemployed resources. Healthcare AI equitably enhancing access and outcomes is possible but will require scientific humility—acknowledging the inherent uncertainties and multidimensional challenges. With transparent analysis and collective determination, the research community can elucidate solutions. But this begins with clearly explaining the complex problems themselves. In closing, advancing healthcare AI equitably is surmountable through evidence-based, cooperative governance.[26 27] We can fulfil these technologies' promise through pragmatic progress grounded in expertise and shared values.

CONCLUSION

Realising the immense potential of healthcare AI to enhance access and quality universally demands judicious governance upholding ethical principles despite political and economic impediments. Pragmatic policies grounded in multidisciplinary research offer paths to responsible adoption. Premarket reviews, participatory design, post-market monitoring, regulatory harmonisation and incentives for voluntary ethics programmes could help maximise benefits while addressing harms proactively and responsively across contexts. However, meaningful progress requires transparency coupled with binding accountability mechanisms that compel demonstrable equity—well-intentioned guidelines alone are insufficient without vigilant enforcement. Through nuanced calibration of rules and guidance, open knowledge exchange and participatory policymaking, incremental gains are achievable. But good faith cooperation must be complemented by upholding accountability when commercial incentives conflict with social welfare. With sustained constructive analysis and willingness to openly confront hard truths, healthcare AI can equitably enhance quality, safety and access for diverse populations worldwide. This will demand enlightened leadership constantly oriented by human rights values and scientific humility. But if we collectively steer these technologies judiciously, healthcare AI can transform modern medicine in service of a more just future for all. The complex way forward begins with choosing progress through principles over profits or efficiency alone. Our shared humanity deserves our absolute best efforts to equitably harness the remarkable potential of AI to heal and unite across borders.

Contributors GD solely conceived and designed the study, conducted the literature review, analysed the data and wrote the manuscript. GD takes responsibility for the integrity of the data analysis and manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD
Ghasem Dolatkhah Laein http://orcid.org/0009-0003-7325-8029

REFERENCES

1 Jiang F, Jiang Y, Zhi H, *et al*. Artificial intelligence in Healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43.
2 Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. *N Engl J Med* 2018;378:981–3.
3 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
4 Kelly CJ, Karthikesalingam A, Suleyman M, *et al*. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
5 Morley J, Floridi L, Kinsey L, *et al*. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics* 2020;26:2141–68.
6 Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019;322:1765–6.
7 Heikkilä M. Five things you need to know about the EU's new AI act. MIT Technology Review; 2023. Available: https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/
8 Mitchell M. Why AI is harder than we think. GECCO '21; Lille France, 2021:3. 10.1145/3449639.3465421 Available: https://dl.acm.org/doi/proceedings/10.1145/3449639
9 Vyas DA, Eisenstein LG, Jones DS. Reconsidering the use of race correction in clinical Algorithms. *N Engl J Med* 2020;383:874–82.
10 Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care. *AMA J Ethics* 2019;21:E167–179.
11 Benjamin R. Assessing risk, automating racism. *Science* 2019;366:421–2.
12 Gianfrancesco MA, Tamang S, Yazdany J, *et al*. Potential biases in machine learning Algorithms using electronic health record data. *JAMA Intern Med* 2018;178:1544–7.
13 Eneanya ND, Yang W, Reese PP. Reconsidering the consequences of using race to estimate kidney function. *JAMA* 2019;322:113–4.
14 Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and Algorithms: an online database. *NPJ Digit Med* 2020;3:118:118:.
15 Yeung K. Regulation by Blockchain: the emerging battle for supremacy between the code of law and code as law. *Mod Law Rev* 2019;82:207–39.
16 U.S. Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-Based software as a medical device (Samd). *Action Plan* 2021. Available: https://www.fda.gov/media/145022/download
17 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
18 Yu K-H, Kohane IS. Framing the challenges of artificial intelligence in medicine. *BMJ Qual Saf* 2019;28:238–41.
19 Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018;15:e1002689.
20 Panch T, Mattie H, Celi LA. The inconvenient truth about AI in Healthcare. *NPJ Digit Med* 2019;2:77.
21 Floridi L. Translating principles into practices of Digital ethics: five risks of being unethical. *Philos Technol* 2019;32:185–93.
22 Davies J, Procter R. Online platforms of public participation -- a Deliberative democracy or a delusion? 2020. Available: https://arxiv.org/abs/2009.14074
23 UK Government. *Establishing a Pro-Innovation Approach to Regulating AI*. London: GOV.UK, 2022. Available: https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai
24 Gasser U, Almeida VAF. A layered model for AI governance. *IEEE Internet Comput* 2017;21:58–62.
25 Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318:517–8.
26 Wiens J, Saria S, Sendak M, *et al*. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019;25:1337–40.
27 Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health* 2019;1:e271–97.
28 Whittaker M, Crawford K, Dobbe R, *et al*. AI now report 2018. AI Now Institute at New York University; 2018.
29 Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 2019;1:501–7.