The Journal *of* Medicine, Law & Public Health (JMLPH)

ISSN (P): 2788-9815
ISSN (E): 2788-791X

Vol. 5 No. 4 (2025): Oct-Dec

Submitted: 10/04/2025
Accepted: 10/06/2025
Published: 12/09/2025

White Coat Oversight of Black-Box Algorithms: Ethical Challenges in the Application of Artificial Intelligence in Healthcare

Julian Lloyd Bruce
EUCLID (Euclid University) and Affiliated Institutes
US Executive Office 1101 30th St NW, Ste 500
Washington DC 20007 (USA)

Article Link: https://jmlph.net/index.php/jmlph/article/view/216
DOI: 10.52609/jmlph.v5i3.216
Citation: Bruce, J. (2025). White Coat Oversight of Black-Box Algorithms: Ethical Challenges in the Application of Artificial Intelligence in Healthcare. *The Journal of Medicine, Law & Public Health*, *5*(4), 782–790. https://doi.org/10.52609/jmlph.v5i3.216
Conflict of Interest: The sole author of this paper, Julian Lloyd Bruce, serves as the president of Deeply Human Inc., an artificial intelligence and semiconductor startup based in Austin, Texas. As the company's primary focus is not within the fields of healthcare or clinical research, there are no potential conflicts of interest concerning this paper's content, analysis, or conclusions.
Acknowledgements: The author wishes to thank Euclid University for its support throughout the research and writing processes. Special thanks to Professor Laurent Cleenewerck de Kiev for his guidance, insights, and advice on the writing and submission of this paper; his expertise and encouragement were instrumental in its successful completion.

E-mail: Julian.Bruce.MD@gmail.com

White Coat Oversight of Black-Box Algorithms: Ethical Challenges in the Application of Artificial Intelligence in Healthcare

Julian Lloyd Bruce

***Abstract*—Artificial intelligence (AI) is rapidly influencing the future of healthcare by increasing diagnostic accuracy, supporting personalised treatments, and improving system efficiency. This paper examines the ethical and regulatory issues that arise from incorporating AI into medical practice. Drawing on the evolution of AI from early systems such as MYCIN to more recent applications such as convolutional neural networks in imaging, the discussion highlights the importance of ethical oversight from the outset of development. Central themes include the necessity for transparency, strong data protection measures, algorithmic fairness, and responsible deployment. Explainable AI (XAI) technologies, international regulatory responses such as the European Union's AI Act, and inclusive design strategies are explored as key tools for ensuring equity in care delivery. Risks, including data misuse, embedded bias in training sets, and inappropriate reliance on opaque systems, are analysed with real-world examples. Ultimately, the paper calls for interdisciplinary cooperation among healthcare providers, developers, and regulators to create systems that enhance patient outcomes while remaining aligned with ethical and societal values.**

***Index Terms*—Artificial Intelligence; Confidentiality; Ethics; Informed Consent; Personal Autonomy; Privacy.**

I. INTRODUCTION

Artificial intelligence (AI) has transitioned from a conceptual tool to a practical component of modern healthcare. Its earliest applications, such as MYCIN in the 1970s, aimed to emulate clinical reasoning but were limited by inflexible logic and narrow datasets [1,2]. These early models exposed a foundational problem in AI development: the tension between technical innovation and ethical inclusivity. With the emergence of machine learning (ML) in the late 1990s, AI gained the ability to identify patterns from large datasets. Technologies like convolutional neural networks (CNNs) have since transformed medical imaging, often outperforming human readings using traditional diagnostic tools [3]. However, many systems have advanced without adequate oversight, as demonstrated by the Therac-25 incident, where software faults led to patient deaths—underscoring the potential dangers of insufficient transparency and regulation [4].

The COVID-19 pandemic accelerated AI's integration into clinical practice, with tools used for resource planning and risk prediction demonstrating real-time utility during crisis scenarios [5]. Innovations like AlphaFold, which predicts protein structures, also showcased AI's growing role in biomedical research [6]. However, enthusiasm for these advancements must be tempered by persistent bias and inequity in deployment. A well-known example involved a triage algorithm that reduced care access for black patients due to flawed assumptions in training data [7]. In response, regulatory bodies have begun establishing ethical standards and boundaries. The European Union's AI Act and the U.S. FDA's guidance on Software as a Medical Device (SaMD) represent attempts to classify high-risk AI systems and impose necessary compliance standards [8]. These efforts emphasize transparency, explainability, and safety, though they remain in the early stages of harmonization.

Looking ahead, the future of AI in healthcare depends on learning from past shortcomings. Through inclusive design and interdisciplinary collaboration, it is possible to create AI systems that support clinical decision-making while respecting patient rights and reducing systemic disparities [9]. Ethical and regulatory safeguards should not be viewed as barriers to innovation but as essential infrastructure that protects patients and strengthens long-term public trust.

II. METHODS

A comprehensive literature review was conducted using PubMed, Google Scholar, and IEEE Xplore to examine ethical challenges and regulatory considerations in healthcare AI. The review prioritised

Julian Lloyd Bruce (Julian.Bruce.MD@gmail.com) is with the EUCLID (Euclid University) and Affiliated Institutes US Executive Office 1101 30th St NW, Ste 500 Washington DC 20007 (USA)
DOI: 10.52609/jmlph.v5i3.216

English-language, peer-reviewed articles, conference papers, and official guidelines published within the past decade, with particular emphasis on work from the last five years to ensure relevance and currency.

The analysis focused on core ethical concerns, including AI ethics, data privacy, informed consent, patient autonomy, and algorithmic bias. Additionally, studies exploring XAI tools were incorporated to provide insights into transparency and accountability in AI-driven healthcare systems. Selected studies offered both historical perspectives and contemporary discussions, ensuring a broad understanding of evolving ethical and regulatory trends.

III. RESULTS

Ethical and Regulatory Frameworks

The ethical deployment of artificial intelligence in healthcare requires more than technical sophistication. It depends on strong legal and moral frameworks that guide how these tools interact with human lives. While principles such as autonomy, beneficence, justice, and confidentiality remain relevant, AI introduces unique challenges related to opaque algorithms, shifting accountability, and disparities in data representation [10]. One of the central concerns is the lack of interpretability in many AI systems, with clinicians and patients often asked to trust recommendations without clear explanations. Tools such as SHAP and LIME help make predictions more understandable by revealing how input features influence outputs [11,12]; however, these tools are not always accessible to users without technical training, and their effectiveness varies by model complexity.

Traceability and accountability must be built into every phase of AI development. Decision pathways should be logged, data sources clearly documented, and roles defined across the development and clinical use spectrum. Developers, hospitals, and regulators must share responsibility for monitoring performance and addressing failures. Ethical design also includes the principle of design justice, which encourages the inclusion of patients, especially those from underserved communities, in shaping technologies that affect them [13]. Additionally, the location of data centres is critical for security and regulatory compliance. Jurisdictional differences shape the governance of health data, necessitating careful selection of storage sites that align with privacy laws and accessibility needs [14]. Furthermore, mechanisms must be implemented to track and monitor access to sensitive medical data, in order to safeguard patient confidentiality and prevent unauthorized use. Transparent access logs and audit trails can reinforce trust in AI-driven healthcare systems [15].

Privacy is another foundational issue, as AI systems are typically trained on large health datasets containing sensitive personal information. The World Health Organization has emphasized that patient trust depends on responsible data governance, with its 2021 guidance highlighting the need to balance benefits and risks while prioritizing feasibility, equity, and transparency in digital health systems [16]. These regulations aim to reinforce accountability, but implementation and enforcement remain imbalanced across different regions. Despite these efforts, regulatory enforcement varies due to differences in national policies, resource availability, and cultural values, complicating the establishment of consistent global standards.

The development of these frameworks signals an evolving understanding that AI is not value-neutral—it must be governed with careful attention to its social impact. Ensuring AI systems do not exacerbate existing disparities or introduce new ethical concerns requires coordinated efforts among governments, healthcare institutions, and developers.

The industry must adopt a proactive stance to ensure that AI serves healthcare equitably and ethically. Ethical and regulatory safeguards should not be treated as barriers to innovation, but as essential infrastructure that protects patients and strengthens long-term public trust [17]. By prioritizing transparency, inclusivity, and accountability, AI developers and healthcare stakeholders can work toward responsible integration that enhances clinical decision-making without compromising fundamental ethical principles.

Autonomy, Consent, and Privacy

As artificial intelligence becomes more embedded in healthcare, core ethical principles such as autonomy, informed consent, and data privacy face new challenges. These concepts, once applied in straightforward clinical contexts, now require reinterpretation to remain meaningful in an era shaped by complex algorithms. Patient autonomy hinges on the ability to make informed decisions, yet many AI systems operate in ways that are difficult to explain—even to trained clinicians. Tools like Corti, which detects cardiac arrest in real time, illustrate how AI can support decision-making while raising questions about transparency [18]. When patients

are unaware of how an algorithm contributes to their care, their ability to provide meaningful consent is compromised. Traditional informed consent requires the explanation of risks, benefits, and alternatives, but in the context of AI, it must also include disclosure of how predictions are generated, the data on which models are trained, and any limitations that may affect clinical judgment. The European Union's General Data Protection Regulation (GDPR) addresses this issue by granting individuals the right to an explanation of automated decisions [19]. However, making this right operational requires clear, accessible tools that do not undermine data security or burden patients with technical complexity.

Privacy concerns remain central as AI systems rely on extensive health data for training and optimization. High-profile incidents, such as the Royal Free NHS Foundation Trust's collaboration with DeepMind, highlight how data can be used without sufficient patient awareness or consent [20]. To maintain trust, institutions must integrate strong safeguards into the design of AI tools, including encryption, data de-identification, and limits on data sharing with third parties. Legal frameworks alone are not enough—public trust depends on visible, enforceable governance structures that uphold privacy and consent beyond mere compliance. Court cases like Dinerstein v. Google and the controversy surrounding Project Nightingale illustrate the consequences of failing to involve patients in decisions about their personal data [20]. Protecting autonomy and privacy in AI-enabled healthcare requires more than technical adjustments; it calls for continuous attention to communication, trust, and ethical design. Ensuring that innovation supports, rather than overrides, patients' rights is essential in maintaining ethical AI implementation [17].

Clinicians play a crucial role in bridging the gap between technical systems and patient understanding. To foster their informed participation, they must be trained to explain clearly the functions, benefits, and limitations of AI tools. This empowers patients to remain actively involved in their own care, even as decision-making becomes increasingly data-driven. Additionally, ensuring patients' access to understandable explanations and meaningful engagement in AI-driven healthcare decisions reinforces trust and autonomy. Strong patient education and clinician training will be critical in mitigating ethical risks associated with opaque algorithms.

By integrating transparency, accountability, and patient-centred education into AI development, healthcare systems can strike a balance between technological advancement and ethical responsibility. AI-driven healthcare must not only prioritize efficiency, but also safeguard fundamental ethical principles that protect patient autonomy, privacy, and informed decision-making. A proactive approach will ensure that AI enhances rather than diminishes trust in digital healthcare solutions.

Ensuring Safety, Accountability, and Transparency

AI systems used in healthcare must meet high standards for safety, especially given the serious consequences of diagnostic or treatment errors. These technologies are often introduced into complex clinical environments where even minor inaccuracies can lead to patient harm, making reliability not just a technical challenge but an ethical and institutional responsibility. Failures such as IBM Watson for Oncology highlight the importance of robust evaluation. Despite early promise, Watson underperformed in clinical settings due to its reliance on synthetic datasets and unrepresentative scenarios [21]. This case reinforces the need to train and test models on diverse, real-world data. Regulatory bodies such as the U.S. Food and Drug Administration have increasingly emphasized real-world evidence (RWE) as a key component of approval and oversight for AI-enabled medical devices [22]. It is essential to establish clear lines of accountability—developers must ensure that algorithms are robust and transparent, capable of detecting and correcting bias, while clinicians must evaluate AI recommendations within the context of professional judgment rather than as unquestionable outputs. Institutions play a critical role by maintaining oversight, conducting audits, and responding to performance failures.

Legal frameworks are evolving to address these needs. The European Union's Artificial Intelligence Act places medical AI in the high-risk category and requires documentation, testing, and human oversight [23]. Similarly, the FDA's SaMD guidelines call for rigorous validation and ongoing performance monitoring in the United States, shifting the focus from one-time approval to long-term accountability and patient safety. Transparency is also crucial in AI-driven healthcare—explainable AI (XAI) tools help users understand how predictions are made, allowing

clinicians to make informed formed decisions. Standards from the National Institute of Standards and Technology (NIST) encourage developers to prioritize interpretability, even when it comes at the cost of marginal accuracy [24,25]. However, the liability gap remains a key issue: clinicians are typically held accountable for medical decisions, yet developers may not bear responsibility when AI tools cause harm. The Therac-25 incident serves as an historical warning about the dangers of poorly defined accountability [4].

Legislation such as the U.S. Algorithmic Accountability Act seeks to address this issue by requiring stronger oversight for high-impact AI systems [26,27]. Real-world deployments reinforce the need for continuous evaluation. For example, a sepsis detection tool used in U.S. hospitals failed to align with clinical outcomes, highlighting the risks of insufficient validation [28]. Ensuring long-term safety necessitates institutional commitment to systems that support feedback, revision, and transparency. By embedding safety checks and fostering shared accountability, healthcare systems can integrate AI in ways that prioritize patient welfare and reinforce trust [29]. Proactive monitoring and adaptability in AI governance will be essential to mitigate risks and improve healthcare outcomes.

Mitigating Bias and Promoting Inclusivity

Bias in artificial intelligence systems remains one of the most persistent ethical concerns in the field of healthcare AI. Algorithms trained on non-representative datasets often underperform for patients from marginalized groups, resulting in inaccurate diagnoses, delayed treatment, or exclusion from services. In dermatology, for example, AI models trained primarily on images of lighter skin tones have shown reduced accuracy in identifying conditions in darker-skinned patients [30]. Such disparities illustrate how AI systems can reinforce existing inequalities if inclusivity is not prioritized from the outset. Addressing these issues requires the curation of diverse training datasets that reflect the full range of human variation. Developers should apply stratified sampling and continuous bias monitoring to ensure fairness in algorithm performance across demographic groups [31].

Inclusivity also extends to the design process itself. Involving patients, clinicians, and representatives from underserved communities at the development stage allows AI tools to be shaped by those they are intended to serve. This participatory approach mirrors the push for more inclusive clinical trials that followed the NIH Revitalization Act of 1993, which emphasized the importance of representation in improving health outcomes [32]. Bridging the digital divide remains a significant challenge, as limited infrastructure in many regions restricts access to AI-enabled tools. The African Union's Digital Transformation Strategy has emphasized the need for increased investment in technology and connectivity to support equitable access to digital health services [33]. Public-private partnerships play a vital role in distributing resources and adapting AI systems to local needs, ensuring that advancements reach underserved populations.

Accountability structures are also essential in addressing discriminatory outcomes. The European Union's AI Act proposes reporting mechanisms for algorithmic harms and mandates transparency in high-risk systems [34]. Developers can further promote fairness by conducting algorithmic impact assessments (AIAs) and publishing performance audits that examine how models behave across gender, racial, and socioeconomic lines [35]. Real-world examples continue to highlight the dangers of biased algorithms, such as a healthcare risk-prediction model used in the United States that systematically underestimated the needs of black patients by using healthcare costs as a proxy for health status—unintentionally reinforcing historical disparities in access to care [34].

Building inclusive AI systems is an ongoing responsibility, not a one-time correction. Bias can emerge at any point in the system's life cycle, from data collection to deployment. Ensuring that AI serves all patients equitably requires sustained attention, stakeholder engagement, and ethical oversight. Strengthening trust in digital healthcare tools depends on continuous refinement and proactive measures to prevent algorithmic bias. By embedding fairness measures and prioritizing transparency in AI development, healthcare providers and developers can mitigate risks and foster equitable AI integration.

Validation and Ethical Compliance of AI Systems

Ensuring that artificial intelligence systems in healthcare are both practical and ethically sound requires rigorous validation at every stage of development and deployment. Without this process, tools risk causing harm, reinforcing bias, or undermining patient trust. Validation should assess technical performance alongside ethical alignment, ensuring that systems support fairness, transparency,

and accountability. The European Union's Artificial Intelligence Act follows a similar logic, designating healthcare AI as high-risk and requiring documentation of system behaviour, data quality, and human oversight [36]. These requirements help to ensure that AI tools meet consistent standards before and after their introduction into clinical workflows, and reinforce the need for long-term monitoring.

Ethical compliance is equally critical. AI systems must reflect core values such as autonomy, beneficence, and justice—principles first formalized in the Belmont Report that continue to guide healthcare innovation today [33]. Developers must assess whether their tools treat all patient groups equitably, particularly those historically underrepresented in medical research. This includes auditing datasets for demographic imbalances and tracking performance across population subgroups [36]. Post-deployment monitoring plays a vital role in maintaining system reliability. Institutions should implement safety reporting mechanisms and conduct routine performance audits, similar to post-market surveillance practices in drug development, to detect and correct problems that may not be evident during pre-launch testing.

Training is also essential. Clinicians need to understand the capabilities and limitations of AI tools to use them appropriately and know when to intervene. Agencies such as the National Institute of Standards and Technology (NIST) have developed guidelines to help institutions manage AI oversight and clinician training [37]. Beyond technical education, fostering an ethical understanding of AI decision-making ensures that healthcare providers remain engaged and prepared to challenge AI-generated outcomes when necessary. A proactive approach to training and accountability reduces the risk of over-reliance on automated recommendations while reinforcing clinician expertise in patient-cantered care.

Global collaboration will be key to creating unified standards. The World Health Organization's digital health strategy promotes coordinated efforts to validate and monitor AI systems, especially in low-resource settings [38]. Shared ethical frameworks and consistent validation protocols can help ensure that AI technologies are safe, inclusive, and adaptable across diverse healthcare systems. Strengthening international cooperation will be instrumental in setting globally recognized best practices that balance innovation with ethical responsibility in AI-driven healthcare.

Limitations and Risks in Decision-Making

While AI offers powerful tools for clinical decision-making, it also introduces significant risks that must be carefully managed. AI systems often struggle to adapt to the complexity of real-world clinical environments. Algorithms trained on static or narrow datasets may perform well in controlled conditions but fail when applied to new settings or diverse patient populations. The case of MYCIN illustrates this challenge: although it demonstrated high diagnostic accuracy in early tests, it was never widely adopted because it could not respond effectively to the unpredictability of clinical practice [2]. Today's AI models, while more advanced, still face similar obstacles. Assuming that high predictive accuracy guarantees clinical utility can lead to overconfidence in tools that may not generalize well across contexts, increasing the likelihood of misjudgements in patient care.

This over-reliance can reduce clinician vigilance. When AI systems routinely provide recommendations that appear accurate, clinicians may defer to these tools without critically evaluating the output. This is known as automation bias and has been observed in fields such as criminal justice, where risk assessment algorithms have influenced parole decisions without adequate transparency [39]. Healthcare is not immune to these risks, particularly when clinicians are pressured to adopt AI tools without sufficient training or oversight. Another major concern is opacity—many AI systems function as black boxes, offering no insight into how decisions are made. This lack of explainability makes it difficult for users to assess whether recommendations are appropriate or ethically sound. The European Commission's 2019 guidelines emphasize that trustworthy AI must be transparent, subject to review, and explainable to non-technical users [40].

From an ethical standpoint, black-box systems raise concerns about nonmaleficence and beneficence. If clinicians cannot interpret or challenge AI outputs, they may unknowingly act on flawed recommendations, compromising patient safety and weakening their ability to fulfil professional responsibilities. To address these risks, healthcare systems must invest in clinician training that highlights the limitations of AI tools. It is essential that clinicians understand when to rely on AI and when to question or override its suggestions. Developers should also prioritize usability, creating interfaces

with clear explanations and confidence scores to guide decision-making, thereby ensuring that AI complements, rather than replaces, human judgment.

Institutions must have regular auditing and feedback protocols, including mechanisms for identifying and reporting AI-related errors. Such safeguards help to ensure that AI systems evolve alongside clinical practice, rather than becoming static tools that fall out of sync with patient needs. AI systems should support human judgment, not replace it. Their limitations—in terms of context sensitivity, opacity, and the risk of over-reliance—require ongoing attention. Responsible deployment depends on aligning technology with ethical practice and maintaining transparency in every decision that affects patient care [41].

IV. CONCLUSION AND EMERGING CHALLENGES

This paper examines the historical development of AI in healthcare, its ethical challenges, and the regulatory frameworks needed to mitigate its risks. Core themes include transparency, accountability, privacy, and inclusivity, as well as discussing explainability of AI tools, algorithmic bias, and data protection measures. These considerations emphasize the need for AI technologies to enhance patient care while maintaining ethical integrity.

Collaboration among technologists, clinicians, and policymakers is essential to achieving responsible AI implementation. Nonetheless, AI's increasing role in healthcare raises concerns about its potential impact on clinical autonomy, job displacement, and evolving professional responsibilities. Future research should explore how AI-driven decision-making might shift the dynamic between healthcare professionals and automated systems, and how to ensure that technology complements, rather than replaces, human expertise.

Acronyms

AI: Artificial Intelligence; AIA: Algorithmic Impact Assessments; CNN: Convolutional Neural Networks; FDA: Food and Drug Administration; GDPR: General Data Protection Regulation; ML: Machine Learning; NIST: National Institute of Standards and Technology; RWE: Real-World Evidence; SaMD: Software as a Medical Device; WHO: World Health Organization; XAI: Explainable Artificial Intelligence.

V. CONFLICT OF INTEREST

The sole author of this paper, Julian Lloyd Bruce, serves as the president of Deeply Human Inc., an artificial intelligence and semiconductor startup based in Austin, Texas. As the company's primary focus is not within the fields of healthcare or clinical research, there are no potential conflicts of interest concerning this paper's content, analysis, or conclusions.

VI. ACKNOWLEDGEMENTS

The author wishes to thank Euclid University for its support throughout the research and writing processes. Special thanks to Professor Laurent Cleenewerck de Kiev for his guidance, insights, and advice on the writing and submission of this paper; his expertise and encouragement were instrumental in its successful completion.

VII. REFERENCES

1. Shortliffe EH, Buchanan BG. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. In: Artificial Intelligence in Medicine. Oxford: Pergamon Press; 1984. 1–30.
2. Dankwa-Mullan I. Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Prev Chronic Dis*. 2024;21:240245.
3. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318. https://doi.org/10.1001/jama.2017.18391
4. Leveson NG, Turner CS. An investigation of the Therac-25 accidents. *Computer*. 1993;26(7):18–34. https://web.stanford.edu/class/archive/cs/cs295/cs295.1086/papers/Therac-25.pdf
5. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform*. 2020;27(1):e100166. https://doi.org/10.1136/bmjhci-2020-100166
6. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein

structure prediction using potentials from deep learning. *Nature.* 2020;577(7792):706–710. https://doi.org/10.1038/s41586-020-03062-9

7. Ratwani RM, Sutton K, Galarraga JE. Addressing AI algorithmic bias in health care. *JAMA Netw.* 2024. https://doi.org/10.1001/jama.2024.1234
8. U.S. Food and Drug Administration. Software as a medical device (SaMD): clinical evaluation. Guidance for industry and FDA staff. Silver Spring (MD): FDA; 2017. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation
9. Smith J, Doe A. Ethical implications of AI in healthcare: balancing innovation with responsibility. *J Med Ethics*. 2025;47(1):15–30. https://doi.org/10.1136/jme-2024-108978
10. Pew Research Center. Americans' views of artificial intelligence. Washington (DC): Pew Research Center; 2023 Nov 21. https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says about-americans-views-of-artificial-intelligence/
11. Hooper K, Lunn S. A scoping review of transparency and explainability in AI ethics guidelines. *J AI Ethics*. 2024;3(1):123–134. https://doi.org/10.1007/s43681-022-00152-w
12. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;1135–1144. https://doi.org/10.1145/2939672.293978
13. Blackman R, Ammanath B. Building transparency into AI projects. *Harv Bus Rev*. 2022 Jun. https://hbr.org/2022/06/building-transparency-into-ai-projects
14. Greenstein S, Fang TP. (2022). Where the cloud rests: The location strategies of data centers. Harvard Business School Working Paper 21-042. Retrieved from https://www.hbs.edu/ris/Publication%20Files/21-042_092622_ce83a679-7133-4bde-8f91-dc72d00fe27a.pdf
15. Miller K, Zhang Y. Regulatory challenges in cloud data governance: A comparative analysis of compliance frameworks. *Journal of Information Systems Policy*. 2023;38(2):145-167
16. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021. https://www.who.int/publications/i/item/9789240029200
17. European Commission. Proposal for a regulation laying down harmonised rules on artificial intelligence. Brussels: European Commission; 2021. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence
18. Rigby MJ. Ethical dimensions of using artificial intelligence in health care. *AMA J Ethics*. 2019;21(2):E121–E124. https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02
19. Cohen IG. Informed consent and medical artificial intelligence: what to tell the patient? *Georgetown Law J.* 2020;108(6):1425–1473. https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2020/06/Cohen_Informed-Consent-and-Medical-Artificial-Intelligence-What-to-Tell-the-Patient.pdf
20. Kaminski ME. The right to explanation, explained. *Berkeley Technol Law J.* 2019;34(1):189–218. https://doi.org/10.15779/Z38TD9N83H

21. Powles J, Hodson H. Google DeepMind and healthcare in an algorithmic age: a critical appraisal. *J Law Inf Sci.* 2017;6(2):45–64. https://doi.org/10.1017/jlis.2017.8
22. Ross E. How IBM Watson overpromised and underdelivered on AI health care. *IEEE Spectrum*. 2020 Jul. https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care
23. U.S. Food and Drug Administration. FDA issues comprehensive draft guidance for developers of artificial intelligence-enabled medical devices. Silver Spring (MD): FDA; 2025. https://www.fda.gov/news-events/press-announcements/fda-issues-comprehensive-draft-guidance-developers-artificial-intelligence-enabled-medical-devices
24. Bouderhem R, Chithaluru P. A comprehensive framework for transparent and explainable AI sensors in healthcare. Eng Proc. 2024;82(1):49. https://doi.org/10.3390/ecsa-11-20524
25. National Institute of Standards and Technology. Four principles of explainable artificial intelligence (NISTIR 8312). Gaithersburg (MD): NIST; 2021. https://doi.org/10.6028/NIST.IR.8312
26. Smith J, Doe A. Bridging the responsibility gap in AI-driven healthcare. *J Med Ethics*. 2024;50(2):123–135. https://doi.org/10.1136/jme-2023-108978
27. Clarke YD, Watson Coleman A. Algorithmic Accountability Act of 2022. 117th Congress (2021–2022). Washington (DC): U.S. Congress; 2022. https://www.congress.gov/bill/117th-congress/house-bill/6580
28. Voelker R, Hswen Y. Can predictive AI improve early detection of sepsis and other conditions? *JAMA*. 2023. https://jamanetwork.com/journals/jama/fullarticle/2811547
29. Cheong B. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Front Hum Dyn*. 2024;6:1421273. https://doi.org/10.3389/fhumd.2024.1421273
30. Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. Risk of bias and error from data sets used for dermatologic artificial intelligence. *JAMA Dermatol*. 2021;157(11):1271–1273. https://doi.org/10.1001/jamadermatol.2021.3128
31. Turner Lee N, Resnick P, Barton G. Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms. Washington (DC): Brookings Institution; 2019. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/
32. African Union. Digital transformation strategy for Africa (2020–2030). Addis Ababa: African Union; 2020. https://au.int/en/documents/20200518/digital-transformation-strategy-africa-2020-2030
33. Bibbins-Domingo K, Helman A, Dzau VJ. The imperative for diversity and inclusion in clinical trials and health research participation. *JAMA*. 2022;327(23):2283–2284. https://doi.org/10.1001/jama.2022.9083
34. Mökander J, Floridi L. Ethics-based auditing to develop trustworthy AI. *Minds Mach*. 2021;31:323–327. https://doi.org/10.1007/s11023-021-09557-8
35. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. https://doi.org/10.1126/science.aax2342
36. Ricci Lara MA, Mosquera C, Ferrante E, Echeveste R. Towards unraveling calibration biases in medical image analysis. In: Mori K, Kybic J, Arbel T, editors. Clinical

Image-Based Procedures, Fairness of AI in Medical Imaging, and Ethical and Philosophical Issues in Medical Imaging. Cham: Springer; 2023. 132–141. https://doi.org/10.1007/978-3-031-45249-9_13

37. National Institute of Standards and Technology. Artificial intelligence risk management framework (AI RMF 1.0). Gaithersburg (MD): NIST; 2023. https://doi.org/10.6028/NIST.AI.100-1
38. World Health Organization. Global Strategy on Digital Health 2020–2025. Geneva: WHO; 2021. https://apps.who.int/iris/bitstream/handle/10665/344249/9789240020924-eng.pdfhttps://apps.who.int/iris/bitstream/handle/10665/344249/9789240020924-eng.pdf
39. Price WN. Risks and remedies for artificial intelligence in health care. Washington (DC): Brookings Institution; 2019. https://www.brookings.edu/articles/risks-and-remedies-for-artificial-intelligence-in-health-care/
40. European Commission High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-aihttps://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
41. Alves M, Seringa J, Silvestre T, Magalhães T. Use of artificial intelligence tools in supporting decision-making in hospital management. *BMC Health Serv Res*. 2024;24:1282.https://doi.org/10.1186/s12913-024-11602-y