A Practical Consideration of the Ethical Challenges and Solutions of AI Implementation in Healthcare: A Systematic Review

Angelica Akrami[1], Martin Ga Zen Tam[1], Dr. Fatema Aftab[2]
1. School of Clinical Medical, University of Cambridge 2. Cambridge University Hospitals

Purpose

Although many AI (artificial intelligence) models are still in their nascent stages, they have great potential to transform many components of healthcare practice, such as disease diagnosis (1), management and the optimisation of clinical workflows (2). However, the transition from *in silico* development to clinical implementation requires the navigation of a multitude of ethical challenges, such as, but not limited to, data security, bias (3), explainability (4), validation, non-maleficence, and autonomy (3). This systematic review aims to understand where AI has been deployed into clinical settings, the ethical challenges that might be considered as a result of this implementation, and the potential solutions devised or utilised. Additionally, we aim to provide insights and recommendations to facilitate the ethical and effective integration of AI into clinical workflows.

Methods

A systematic search of PubMed, EMBASE, Web of Science, and Scopus identified literature on AI deployment in healthcare (inception to October 2024). Research articles, reviews, empirical studies, and case studies in all languages were considered. Following removal of duplicates, 478 articles went through title and abstract screening according to a pre-defined inclusion and exclusion criteria for details on real-world AI implementation in patient-facing applications within clinical environments, decision-making systems, and medical technologies. We defined real-world AI implementation as the deployment of an AI model that has completed internal (e.g. publicly available datasets) and/or external validation (e.g. clinical data) and is integrated into clinical workflows, actively influencing clinical decision-making. Patient-facing AI refers to models that directly interact with or impact patients in a healthcare context. Full-text screening resulted in 9 articles being selected for analysis of ethical challenges associated with AI deployment in healthcare settings.

Results

Ethical Challenges

We defined an ethical challenge as an issue that directly involved one of the four pillars of biomedical ethics, as defined by Beauchamp and Childress (5).

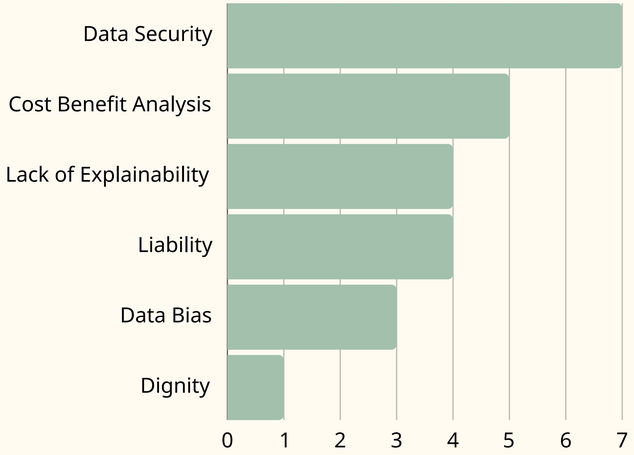


Figure 1 *- A graphical representation showcasing the frequency of ethical challenges identified across the 9 included papers.*

The ethical challenges discussed in the papers can be grouped into the categories of data security, cost benefit analysis, lack of explainability, liability, data bias, and dignity. We measured the frequency in which these challenges were included in the selected papers (Figure 1).

Solutions

The solutions were sorted based on the ethical challenges they addressed and classified as either theoretical proposals or practical implementations developed in response to these challenges. The implemented solutions have been showcased as seen below (Table 1).

Ethical Challenge	Implemented Solutions
Data Security	Encryption, 2-factor authentication, continual staff education, secure data storage including reduction of data transmission, appropriate physical security measures, survey to understand patient perspective of sharing data and engagement initiatives including public events to discuss concerns
Data Bias	Inclusion of clinicians in the decision of what data to use in AI training
Lack of Explainability	None reported
Liability	Public and external accountability
Cost Benefit Analysis	Integration into legacy systems / EHRs, inclusion of clinicians to understand what is effective
Dignity	MDT approach to involve all stakeholders, establish an ethical research review process

Table 1 *- Implemented solutions sorted according to the ethical challenge that they aim to address.*

Conclusion

AI, as a tool in the healthcare system, represents an unprecedented paradigm. It has the capacity to independently influence clinical decision-making but lacks the consciousness to reflect on its actions when outcomes are unfavourable. This places AI in a distinct category—neither human nor a conventional medical device—necessitating the creation of novel regulations and frameworks. These would guide stakeholders through the challenges of implementation and foster the development of innovative solutions.

Transparency must also be incentivized, enabling developers and deployers to document insights and challenges without fear of intellectual property infringement. This collaborative approach would accelerate the advancement of AI solutions and benefit the wider healthcare community.

Many deployed AI systems lack clearly defined key performance indicators (KPIs), making it difficult for stakeholders to assess their value and impact in practical terms, such as cost savings or time efficiency. Establishing robust KPIs would not only highlight challenges but also enable benchmarking of solutions and measurable progress.

Finally, evaluating ethical challenges remains complex due to the absence of a centralized consensus on what constitutes an “ethical challenge” in the context of AI implementation. Defining these parameters is essential for addressing ethical concerns systematically and ensuring responsible integration of AI into healthcare.

Additionally, the practical usefulness of stratifying ethical challenges into often abstract terms must also be considered. Sendak et al (5) mentioned the use of the term explainability, and how forcing clinicians to understand every aspect of AI, might be counter-intuitive and lead to overburden and clinician fatigue, and that the focus might be better placed on trying to increase trust and accountability through other means. However, what those other means may be remains unclear.

References

1. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018 Feb 22;172(5):1122-1131.e9.
2. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019 Oct 29;17(1):195.
3. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare. 2023 Jan;11(6):887.
4. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy. 2021 Jan;23(1):18.
5. Beauchamp TL, Childress JF. Principles of biomedical ethics. 5th. New York: Oxford University Press; 2001.
6. Sendak M, Elish MC, Gao M, Futoma J, Ratliff W, Nichols M, et al. 'The human body is a black box': supporting clinical decision-making with deep learning. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2024 Nov 25]. p. 99–109. (FAT* '20). Available from: https://dl.acm.org/doi/10.1145/3351095.3372827