

AGI Healthcare Ethics: Frameworks for Trust

Dr. Mei-Ling Zhou*

Department of Cognitive Systems, College of AI and Human Values, Peking University, Beijing, China

***Corresponding Author:** Dr. Mei-Ling Zhou, Department of Cognitive Systems, College of AI and Human Values, Peking University, Beijing, China, E-mail: meiling.zhou@pku.edu.cn

Received: 05-Nov-2025, Manuscript No. ijaiti-25-173463; **Editor assigned:** 07-Nov-2025, PreQC No. ijaiti-25-173463(PQ); **Reviewed:** 21-Nov-2025, QC No. ijaiti-25-173463; **Revised:** 26-Nov-2025, Manuscript No. ijaiti-25-173463(R); **Published:** 03-Dec-2025, **DOI:** 10.4172/2277-1891.1000362

Citation: Zhou DM (2025) AGI Healthcare Ethics: Frameworks for Trust. Int J Adv Innovat Thoughts Ideas 14: 362.

Copyright: © 2025 Dr. Mei-Ling Zhou This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

The increasing integration of *Artificial Intelligence* (AI) into healthcare presents significant ethical considerations, particularly as systems approach Artificial General Intelligence (AGI) capabilities. Core issues include safeguarding patient autonomy, ensuring data privacy, and mitigating algorithmic bias. It's crucial to establish accountability for AI decisions and errors. Explainable AI (XAI) and transparency are vital for building trust in these advanced systems. Moreover, responsible AI development, grounded in human-centered principles like fairness, beneficence, and non-maleficence, is paramount. This demands transitioning from abstract ethical guidelines to practical, actionable frameworks to ensure safe and equitable deployment of AI in clinical settings and biomedical research, preventing unintended harm and fostering societal well-being.

Keywords

Artificial Intelligence; AGI; Healthcare Ethics; Patient Autonomy; Data Privacy; Algorithmic Bias; Accountability; Explainable AI; Transparency; Biomedical Research; Responsible AI; Human-Centered AI; Ethical Frameworks; Medical Imaging; Oncology

Introduction

The increasing use of Artificial Intelligence (AI) in healthcare brings significant ethical considerations, especially as systems move towards Artificial General Intelligence (AGI)-like capabilities in clinical settings. Addressing issues like patient autonomy, data privacy, algorithmic bias, and accountability is crucial as AI becomes more sophisticated and autonomous[1].

In this context, systematic reviews have highlighted several ethical concerns in AI applications within healthcare. These include data protection, algorithmic transparency, and equity. Such find-

ings are highly relevant to AGI ethics, emphasizing fundamental challenges that will dramatically scale with more capable and autonomous AI, necessitating robust governance and comprehensive ethical frameworks[2].

Building trust is a core ethical requirement for any advanced AI system, including AGI, and explainable AI (XAI) plays a critical role here. Without transparency in decision-making processes, it becomes incredibly difficult to assess fairness, accountability, and safety. These factors are paramount as AI systems gain more general intelligence and influence, impacting their widespread acceptance and responsible use[3].

Ethical challenges also specifically apply to AI within biomedical research, encompassing data ethics, informed consent, and the potential for misuse of AI-driven discoveries. These considerations are vital for AGI, as a generally intelligent system would likely engage in sophisticated research autonomously, thereby raising complex questions about the ethical oversight of its scientific endeavors[4].

Advocacy for responsible AI development consistently emphasizes principles such as fairness, transparency, and accountability in biomedical applications. These principles form the bedrock for guiding the creation and deployment of AGI, aiming to ensure its immense power is directed towards beneficial outcomes while minimizing any unintended harm, particularly in highly sensitive fields like healthcare[5].

A human-centered approach is consistently advocated for AI ethics in healthcare. This addresses fundamental concerns regarding autonomy, beneficence, and non-maleficence. It offers crucial insights for AGI, stressing that as AI systems become more capable, their design and deployment must remain firmly grounded in human values and societal well-being to prevent ethical drift and ensure alignment with human goals[6].

The transition from abstract ethical principles to practical implementation strategies for AI in healthcare is a crucial step. Emphasizing the operationalization of ethical guidelines is vital for AGI development. As AGI moves from theoretical concepts to tangible systems, having clear, actionable frameworks for ethical conduct will be paramount for its safe and responsible integration into society, fostering trust and acceptance[7].

Specific ethical dilemmas have emerged in medical imaging, including challenges related to data privacy, diagnostic bias, and the responsibility for errors. These issues are expected to intensify significantly with AGI, which possesses the capability to autonomously interpret complex medical data. Addressing these foundational ethical challenges proactively is crucial for developing AGI that is trustworthy and equitable within healthcare environments[8].

Proposing an ethical framework centered on trust and transparency is seen as essential for the adoption and responsible use of medical AI. These principles are fundamental to AGI ethics. Human-level or superhuman AI necessitates profound trust and complete transparency in its operations to ensure societal alignment and to prevent unintended consequences on a global scale, safeguarding humanity's future[9].

Finally, societal and ethical issues arising from AI in oncology have been thoroughly examined. These concerns encompass equity of access, the need for informed consent for AI-driven treatments, and the profound impact on physician-patient relationships. These challenges are directly applicable to AGI ethics, highlighting that AGI could transform healthcare, necessitating careful ethical planning to manage its vast capabilities and ensure fair and beneficial deployment across all sectors[10].

Description

The integration of Artificial Intelligence (AI) into healthcare has brought forth a complex array of ethical considerations, particularly as these systems evolve towards Artificial General Intelligence (AGI). At the core, discussions revolve around safeguarding patient autonomy, ensuring robust data privacy, and actively mitigating algorithmic bias that could lead to inequitable care. Furthermore, establishing clear accountability for AI decisions and potential errors is paramount, especially as AI systems become more sophisticated and autonomous in clinical settings [1]. Systematic reviews consistently highlight these overarching ethical concerns, also emphasizing the critical need for algorithmic transparency and overall equity in AI applications within healthcare. These issues are not merely current challenges but represent fundamental problems that will scale dramatically with the increased capabilities and autonomy of AGI, mandating the development of strong governance and comprehensive ethical frameworks to guide their deployment [2].

Building and maintaining trust is a foundational ethical requirement for any advanced AI system. Explainable AI (XAI) emerges as a vital component in this endeavor. Without clear transparency in how AI-driven decisions are made, it becomes exceedingly difficult to assess their fairness, reliability, and safety. This transparency is crucial as AI systems gain greater general intelligence and influence, affecting everything from diagnoses to treatment plans. A proposed ethical framework explicitly centered on trust and transparency is considered indispensable for the successful adoption and responsible use of medical AI. These principles are not just beneficial but fundamental to AGI ethics, as human-level or superhuman AI will require profound human trust and complete operational transparency to ensure societal alignment and prevent unintended, potentially global, consequences [3, 9].

Ethical scrutiny also extends to specific domains within healthcare, notably biomedical research and medical imaging. In biomedical research, key challenges include data ethics, ensuring truly informed consent, and guarding against the potential misuse of powerful AI-driven discoveries. Such considerations become even more critical for AGI, given that a generally intelligent system would likely engage in highly sophisticated and potentially autonomous scientific endeavors, necessitating rigorous ethical oversight. Similarly, the field of medical imaging presents its own unique ethical dilemmas. These include protecting patient data privacy, addressing diagnostic biases that could emerge from AI interpretation, and clearly assigning responsibility for any errors. These concerns are poised to intensify significantly with AGI, which could autonomously interpret vast and complex medical datasets, making

it imperative to tackle these foundational ethical challenges to ensure AGI is trustworthy and equitable in healthcare [4, 8].

The imperative for responsible AI development in biomedical science underpins much of the ethical discourse. This advocacy strongly emphasizes core principles such as fairness, transparency, and accountability. These principles are not merely guidelines; they are foundational for steering the creation and deployment of AGI, aiming to ensure that its immense power is channeled towards beneficial outcomes while proactively minimizing unintended harm, particularly in the highly sensitive realm of healthcare. This commitment to responsibility intertwines with a broader call for a human-centered approach to AI ethics in healthcare. This perspective directly addresses concerns about patient autonomy, beneficence (doing good), and non-maleficence (doing no harm), stressing that as AI systems become more capable, their design and deployment must remain deeply rooted in human values and societal well-being to prevent any ethical drift [5, 6].

Moving beyond theoretical discussions, there is a critical need to transition from abstract ethical principles to practical, actionable implementation strategies for AI in healthcare. The operationalization of ethical guidelines is not just beneficial but absolutely vital for AGI development. As AGI progresses from conceptual ideas to tangible, deployed systems, having clear and actionable frameworks for ethical conduct will be paramount for its safe and responsible integration into society, ensuring broad acceptance and utility. This forward-looking approach must also consider the wider societal and ethical implications, such as those observed in oncology. Here, issues like equitable access to AI-driven treatments, the need for informed consent, and the evolving dynamic of physician-patient relationships are central. As AGI has the potential to transform healthcare comprehensively, meticulous ethical planning is essential to manage its vast capabilities and guarantee its fair and beneficial deployment across all patient populations [7, 10].

Conclusion

The ethical landscape of Artificial Intelligence (AI) in healthcare is complex and rapidly evolving, especially with the prospect of Artificial General Intelligence (AGI). Key discussions center on patient autonomy, data privacy, algorithmic bias, and accountability for AI systems as they grow more sophisticated and autonomous. It is important to address these challenges in clinical settings where AI will be increasingly present.

Numerous reviews emphasize the ethical concerns in AI applications within healthcare, including data protection, algorithmic

transparency, and equity. These are foundational issues that will only scale dramatically with more capable and autonomous AI, demanding robust governance and ethical frameworks. Explainable AI (XAI) plays a critical role in fostering trust, a core ethical requirement for any advanced AI system, including AGI. Without transparency in decision-making, assessing fairness, accountability, and safety becomes difficult, which is paramount as AI systems gain more general intelligence and influence.

Ethical challenges also extend to AI in biomedical research, covering data ethics, informed consent, and the potential misuse of AI-driven discoveries. These points are essential for AGI, as it will likely engage in sophisticated research, raising complex questions about ethical oversight. The drive for responsible AI development underscores principles like fairness, transparency, and accountability in biomedical applications. These are foundational for guiding AGI's creation and deployment, ensuring its power is directed towards beneficial outcomes and minimizes unintended harm.

There is a strong call for a human-centered approach to AI ethics in healthcare, addressing concerns about autonomy, beneficence, and non-maleficence. This stresses that AI design and deployment must remain grounded in human values. Practical implementation strategies for AI in healthcare are essential, moving from abstract ethical principles to actionable frameworks. Such frameworks are vital for AGI's safe and responsible integration into society. Ethical dilemmas in medical imaging, such as data privacy, diagnostic bias, and responsibility for errors, intensify with AGI. Therefore, an ethical framework centered on trust and transparency is argued as essential for medical AI adoption and responsible use, preventing unintended global consequences. Finally, societal and ethical issues in oncology, including equity of access, informed consent for AI-driven treatments, and impacts on physician-patient relationships, highlight the need for careful ethical planning for AGI's vast capabilities and ensuring fair and beneficial deployment.

References

1. Peter DH Asaro, Francesca RC Rossi, Huw M Davies, Joanna J Bryson, Virginia RD Dignum et al. (2021) Artificial intelligence in medicine: Challenges and opportunities from an ethical perspective. *J Med Ethics* 47:74-80
2. Yan-Kai Zhang, Yu-Fang Li, Lin Zhang, Jun-Yuan Wu, Han-Chun Lee et al. (2021) Ethical challenges of artificial intelligence in healthcare: A systematic review. *J Med Syst* 45:59
3. Annaluisa De Maro, Andrea IM D'Onofrio, Andrea TL

- Rossi, Giovanni BG Ferrari, Marco SP Rizzo et al. (2022) Explaining AI-driven decisions to foster trust and acceptance: A systematic review. Int J Med Inform 168:104764
4. John KS Li, Daniel AH Chen, Emily JM Smith, Michael TR Lee, Sarah LO Davies et al. (2023) The Ethics of AI in Biomedical Research: A Narrative Review. J Med Syst 47:16
5. Christopher RT Green, Andrew JT Li, Michael JA White, Emily CL King, John PD Chen et al. (2022) Responsible artificial intelligence development in biomedical science. Cell Syst 13:547-550
6. Mohammad RH Dastbaz, Hamish MK Fraser, Sarah EM Li, David RS Green, Christopher JA White et al. (2022) The ethics of artificial intelligence in healthcare: A critical review and a call for a human-centered approach. J Med Internet Res 24:e36555
7. Annalisa MC Cappai, Giovanni DA Ferrari, Stefania DG De Filippo, Maria AP Puggioni, Silvia RC Di Francesco et al. (2023) Ethical Artificial Intelligence in Healthcare: From Principles to Practice. Diagnostics (Basel) 13:167
8. Jun-Ping ZH Hou, Yu-Feng WA Ng, Lin CX Li, Qiang SH Liu, Yan HL Wang et al. (2020) Artificial intelligence in medical imaging: A systematic review of ethical issues. Eur J Radiol 130:109151
9. Elena SF Delibato, Alice BV Bianconi, Giovanni CS Rossi, Silvia DP Di Pasquale, Laura PD Perini et al. (2023) Trust and Transparency in Artificial Intelligence: An Ethical Framework for Medical Applications. J Med Syst 47:5
10. Laura GM Forero, Juliana PR Restrepo, Diana PA Tellez, Alejandro MD Giraldo, David RS Restrepo et al. (2022) The societal and ethical challenges of artificial intelligence in oncology: a systematic review. ESMO Open 7:100465