Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 1, 2026

ETHICAL DECISION-MAKING FRAMEWORKS FOR AI IN HEALTHCARE

Maliha Manzoor[1], Zubaida Laraib[2], Waqas Tariq Paracha[*3], Haleema Inam[4]

[1]*Gomal Research Institute of Computing (GRIC), Faculty of Computing, Gomal University, DI Khan, Pakistan*
[2]*Department of Computer Science and IT, Thal University Bhakkar*
[*3]*Assistant Professor, Department of Computer Science and IT, University of Lahore, Lahore*[3]
[4]*Department of Computer Science and IT, Virtual University Islamabad*

[1]maliha.manzoor6@gmail.com, [2]zubaidalaraib@gmail.com, [*3]waqasparacha125@gmail.com, [4]haleemainam786@gmail.com

DOI: https://doi.org/10.5281/zenodo.18530996

Keywords

Article History
Received: 06 December 2025
Accepted: 16 January 2026
Published: 31 January 2026

Corresponding Author: *
Waqas Tariq Paracha

Abstract

The rapid integration of artificial intelligence (AI) into healthcare has transformed clinical decision-making, diagnostics, and patient management. While AI-driven systems offer substantial benefits in terms of accuracy, efficiency, and scalability, they also introduce complex ethical challenges related to transparency, accountability, bias, data privacy, and patient autonomy. Addressing these concerns requires robust ethical decision-making frameworks that can guide the responsible design, deployment, and governance of AI technologies in healthcare settings. This paper critically examines existing ethical decision-making frameworks for AI in healthcare and evaluates their effectiveness in real-world clinical contexts.

The study synthesizes contemporary literature to analyze key ethical principles underpinning AI governance, including beneficence, non-maleficence, justice, explicability, and human oversight. It highlights how traditional bioethical models, when combined with AI-specific governance mechanisms, can support ethically aligned clinical decisions. Particular attention is given to algorithmic bias and fairness, emphasizing the risks posed to vulnerable populations when datasets are unrepresentative or poorly curated (Vokinger et al., 2021; Rajkomar et al., 2022). Furthermore, the paper explores accountability structures for AI-assisted decisions, addressing the ethical ambiguity surrounding responsibility when clinical outcomes are influenced by automated systems (Gerke et al., 2020; Morley et al., 2021).

This research also reviews emerging regulatory and institutional frameworks, including explainable AI (XAI) models and ethics-by-design approaches, which aim to embed ethical reasoning directly into AI development lifecycles (Floridi et al., 2022). The findings suggest that no single framework is sufficient to address the multifaceted ethical challenges of AI in healthcare. Instead, an integrated, context-aware ethical decision-making model is required, combining technical safeguards, clinical expertise, and continuous ethical evaluation.

The paper concludes by proposing a synthesized ethical decision-making framework tailored for healthcare AI applications. This framework supports transparent, fair, and accountable AI use while preserving clinician authority and patient trust. The study contributes to ongoing discourse by offering practical insights for

policymakers, healthcare professionals, and AI developers seeking to implement ethically responsible AI systems in clinical practice.

INTRODUCTION

Artificial intelligence (AI) has rapidly become an integral component of modern healthcare, influencing areas such as clinical decision support, medical imaging, predictive analytics, and personalized treatment planning. Machine learning algorithms are increasingly relied upon to process large-scale health data and assist clinicians in making timely and accurate decisions. Despite these advancements, the growing reliance on AI systems raises critical ethical concerns that challenge traditional healthcare values, including patient autonomy, clinical responsibility, fairness, and trust (Rajkomar et al., 2022).

One of the central ethical problems in healthcare AI lies in the opacity of algorithmic decision-making. Many high-performing AI models function as "black boxes," offering limited interpretability for clinicians and patients. This lack of transparency complicates informed consent, weakens trust in clinical decisions, and raises concerns about accountability when adverse outcomes occur (Morley et al., 2021). Furthermore, AI systems trained on biased or incomplete datasets may reinforce existing health disparities, disproportionately affecting marginalized or underrepresented populations (Vokinger et al., 2021).

In response to these challenges, several ethical guidelines and frameworks for AI have been proposed by international organizations, regulatory bodies, and academic researchers. These frameworks commonly emphasize high-level principles such as beneficence, non-maleficence, justice, explicability, and human oversight (Floridi et al., 2022). While these principles provide an important ethical foundation, their translation into practical, context-sensitive decision-making processes within healthcare environments remains limited. Clinicians and healthcare institutions often lack actionable guidance on how to operationalize these ethical principles during real-time AI-assisted clinical decisions.

Moreover, existing ethical frameworks frequently address AI ethics in a generalized or technology-centric manner, without sufficient consideration of the unique characteristics of healthcare settings. Clinical environments involve high-stakes decisions, shared responsibility between humans and machines, and complex regulatory and legal obligations. As a result, current frameworks may fall short in addressing questions such as who holds moral and legal responsibility for AI-supported decisions, how ethical trade-offs should be evaluated in urgent clinical contexts, and how continuous ethical monitoring can be integrated throughout the AI system lifecycle (Gerke et al., 2020).

This gap highlights the need for a structured, healthcare-specific ethical decision-making framework that goes beyond abstract principles. Such a framework should support clinicians, developers, and policymakers by offering practical mechanisms for ethical reasoning, accountability allocation, and bias mitigation in AI-driven healthcare applications. Addressing this gap is essential to ensure that AI technologies enhance patient care while upholding fundamental ethical standards and maintaining public trust in healthcare systems.

Literature Review

No.	Author(s) & Year	Study Focus	Methodology	Key Findings	Relevance to Ethical Decision-Making
1	Floridi et al. (2022)	Ethical principles for AI	Conceptual framework	Proposed AI4People principles including explicability and responsibility	Foundation for ethical AI governance
2	Morley et al. (2021)	Translating AI ethics into practice	Systematic review	Identified tools bridging ethical principles and implementation	Highlights operational gap in ethics
3	Vokinger et al. (2021)	Bias in medical AI	Narrative review	Demonstrated risks of biased datasets in healthcare	Emphasizes justice and fairness
4	Rajkomar et al. (2022)	Fairness in healthcare ML	Perspective analysis	Stressed need for equity-aware ML design	Supports ethical equity frameworks
5	Gerke et al. (2020)	Legal and ethical challenges	Qualitative analysis	Raised accountability and liability concerns	Reinforces responsibility dimension
6	Mittelstadt (2019)	Limits of AI ethics	Philosophical analysis	Highlighted lack of enforcement mechanisms	Justifies need for decision frameworks
7	Reddy et al. (2020)	AI transparency in clinics	Empirical study	Clinicians prefer interpretable models	Supports explainability requirement
8	London (2019)	AI and moral responsibility	Ethical analysis	Questioned delegation of decisions to AI	Human oversight importance
9	Topol (2019)	AI-human collaboration	Review	Advocated clinician-in-the-loop models	Ethical hybrid decision-making
10	Yu et al. (2021)	Trust in medical AI	Survey study	Transparency increases clinician trust	Trust as ethical outcome
11	McCradden et al. (2020)	Accountability in AI	Conceptual study	Proposed shared accountability models	Addresses responsibility ambiguity
12	Amann et al. (2020)	Explainable AI in healthcare	Systematic review	XAI improves clinical adoption	Ethical transparency tool
13	World Health Organization (2021)	Ethics & governance of AI	Policy guideline	Issued global ethical guidance	Regulatory ethical alignment
14	Char et al. (2020)	Ethical deployment challenges	Case-based analysis	Identified ethical failures in real systems	Practical ethics gaps
15	Grote & Berens (2020)	Ethical risk assessment	Framework proposal	Introduced ethical impact assessment	Decision-support for ethics

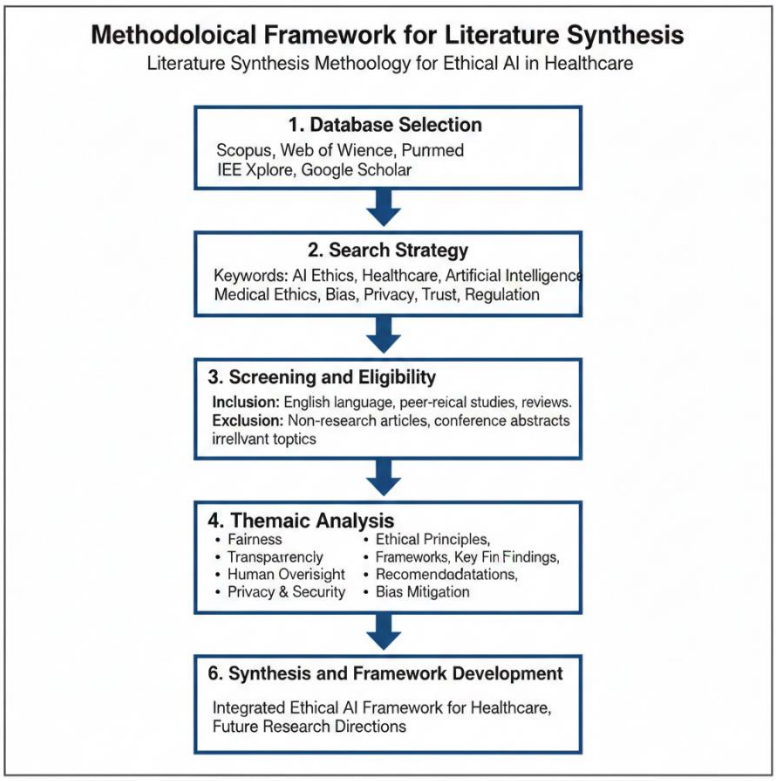
3. Methodology: Literature Synthesis Approach

This study adopts a qualitative literature synthesis approach to examine ethical decision-making frameworks for artificial intelligence in healthcare. The methodology is designed to systematically analyze, integrate, and interpret existing scholarly work in order to identify prevailing ethical principles, implementation gaps, and emerging best practices relevant to healthcare-specific AI applications. A literature synthesis approach is particularly appropriate for this study, as the research objective is not to evaluate algorithmic performance but to develop a conceptual and decision-oriented ethical framework grounded in existing evidence and theoretical discourse (Morley et al., 2021).

Phase	Methodological Step	Description	Purpose	Key Outputs
Phase 1	Database Selection	Selection of peer-reviewed databases including Scopus, Web of Science, PubMed, IEEE Xplore, and Google Scholar	Ensure comprehensive and multidisciplinary coverage	Initial pool of relevant studies
Phase 2	Search Strategy	Use of structured keywords and Boolean operators related to AI ethics and healthcare	Identify ethically relevant AI literature	Refined search results
Phase 3	Screening & Eligibility	Application of inclusion and exclusion criteria (healthcare focus, ethical dimension, publication year, language)	Remove irrelevant or low-quality studies	Final set of selected articles
Phase 4	Data Extraction	Extraction of ethical themes, frameworks, methodologies, and findings from selected studies	Standardize information for analysis	Thematic dataset
Phase 5	Thematic Analysis	Qualitative coding and grouping of recurring ethical principles (e.g., fairness, transparency, accountability)	Identify convergent and divergent ethical dimensions	Core ethical themes
Phase 6	Synthesis & Framework Development	Integration of themes into a unified ethical decision-making framework	Translate theory into practical guidance	Proposed ethical framework

Search Strategy and Data Sources

A comprehensive search was conducted across major academic databases, including Scopus, Web of Science, PubMed, IEEE Xplore, and Google Scholar. Keywords and search strings were constructed using combinations of terms such as "AI ethics," "healthcare artificial intelligence," "ethical decision-making frameworks," "algorithmic accountability," "bias in medical AI," and "explainable AI." The search was limited to peer-reviewed journal articles, policy reports, and authoritative reviews published primarily between 2020 and 2024 to ensure relevance to current technological and regulatory contexts.



Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: (i) focused on ethical, legal, or governance aspects of AI in healthcare; (ii) proposed, evaluated, or discussed ethical frameworks, principles, or decision-making models; and (iii) were published in English. Articles focusing exclusively on technical performance without ethical discussion, opinion pieces lacking analytical depth, or studies unrelated to healthcare contexts were excluded. This screening process ensured that the synthesized literature directly informed ethical decision-making considerations.

Data Extraction and Thematic Analysis

Selected studies were systematically reviewed to extract key information, including ethical principles addressed, methodological approach, context of application, and identified challenges. A thematic analysis was then performed to categorize recurring ethical dimensions such as transparency, fairness, accountability, autonomy, and human oversight (Floridi et al., 2022; Vokinger et al., 2021). These themes were analyzed comparatively to identify areas of convergence and divergence across frameworks.

Synthesis and Framework Development

Rather than summarizing studies in isolation, the synthesis emphasized relational analysis, examining how ethical principles interact within real-world clinical decision-making scenarios. Particular attention was given to gaps between high-level ethical guidelines and their practical implementation in healthcare settings. Insights derived from this synthesis informed the development of an integrated ethical decision-making framework that aligns ethical principles with actionable steps across the AI lifecycle, from design and validation to deployment and monitoring (Gerke et al., 2020).

This methodology ensures analytical rigor while allowing flexibility to integrate interdisciplinary perspectives, ultimately supporting the

development of a healthcare-specific ethical decision-making framework for AI applications.

Proposed Ethical Decision-Making Framework for AI in Healthcare

4.1 Framework Overview

To address the limitations identified in existing ethical guidelines and to bridge the gap between abstract principles and clinical practice, this study proposes a structured ethical decision-making framework tailored specifically for AI applications in healthcare. The framework is designed to support ethically informed decisions across the entire AI lifecycle, ensuring that AI-assisted clinical practices remain transparent, fair, accountable, and aligned with patient-centered values. Unlike principle-only models, this framework integrates ethical reasoning into operational decision points, enabling practical application in real-world healthcare environments (Floridi et al., 2022; Morley et al., 2021).

The proposed framework adopts a multi-layered, cyclical structure consisting of six interrelated components: ethical grounding, data and model integrity, transparency and explainability, human oversight, accountability allocation, and continuous ethical monitoring. These components collectively guide stakeholders in evaluating ethical risks before, during, and after AI deployment.

4.2 Core Components of the Framework

Ethical Grounding and Context Assessment

The first component establishes ethical grounding by aligning AI systems with core healthcare ethics principles, including beneficence, non-maleficence, autonomy, and justice. At this stage, the clinical context, patient population, and potential ethical trade-offs are explicitly identified. This ensures that AI use is purpose-driven and sensitive to clinical risk levels and patient vulnerability (WHO, 2021).

Data and Model Integrity

Ethical decision-making begins with responsible data practices. This component emphasizes data representativeness, quality, and bias assessment throughout model development. Bias audits and validation across diverse patient groups are incorporated to reduce the risk of inequitable outcomes, particularly for marginalized populations (Vokinger et al., 2021; Rajkomar et al., 2022).

Transparency and Explainability

Given the high-stakes nature of clinical decisions, the framework prioritizes model transparency and interpretability. Explainable AI mechanisms are integrated to ensure clinicians can understand, question, and communicate AI-generated recommendations. This supports informed consent and strengthens clinician and patient trust in AI-assisted decisions (Amann et al., 2020).

Human Oversight and Clinical Authority

The framework reinforces the clinician-in-the-loop principle, positioning AI as a decision-support tool rather than an autonomous decision-maker. Clinicians retain final authority, with clear guidance on when AI outputs should be accepted, overridden, or escalated. This component safeguards professional judgment and ethical responsibility in clinical care (Topol, 2019).

Accountability and Responsibility Allocation

A critical feature of the framework is explicit accountability mapping. Ethical and legal responsibilities are delineated among developers, healthcare institutions, clinicians, and regulators. This clarity addresses ambiguity in adverse outcomes involving AI-assisted decisions and supports transparent governance structures (Gerke et al., 2020; McCradden et al., 2020).

Continuous Ethical Monitoring and Feedback

The final component emphasizes post-deployment monitoring, recognizing that ethical risks evolve over time. Ongoing performance audits, bias reassessments, and ethical impact evaluations are incorporated to ensure sustained ethical compliance. Feedback mechanisms allow the framework to adapt to changing clinical, technological, and regulatory conditions (Morley et al., 2021).

4.3 Diagram Explanation

The proposed framework is visually represented as a cyclical model, illustrating the continuous nature of ethical decision-making in healthcare AI. At the center lies clinical decision-making, supported by AI outputs. Surrounding this core are six interconnected layers corresponding to the framework components. Arrows indicate bidirectional flow, emphasizing that ethical evaluation is not a one-time process but an ongoing cycle spanning design, deployment, and clinical use. This structure highlights the interdependence of ethical principles, technical safeguards, and human judgment, reinforcing the dynamic and context-sensitive nature of ethical AI in healthcare.

4.4 Contribution of the Framework

By translating ethical principles into structured decision points, the proposed framework offers practical guidance for clinicians, developers, and policymakers. It advances ethical AI implementation by aligning technological innovation with healthcare values, supporting responsible adoption while preserving trust, accountability, and patient well-being.

Conclusion and Future Research Directions

6.1 Conclusion

The increasing integration of artificial intelligence into healthcare systems has reshaped clinical decision-making, offering significant improvements in diagnostic accuracy, operational efficiency, and personalized care. However, as demonstrated throughout this study, the ethical challenges associated with healthcare AI remain substantial and multifaceted. Issues related to transparency, bias, accountability, patient autonomy, and trust continue to hinder the responsible adoption of AI technologies in clinical environments (Rajkomar et al., 2022; Vokinger et al., 2021).

This paper addressed these challenges by critically reviewing existing ethical decision-making frameworks and identifying a clear gap between high-level ethical principles and their practical application in healthcare settings. While current frameworks provide valuable normative guidance, they often lack operational clarity for clinicians, developers, and healthcare institutions faced with real-time AI-assisted decisions (Morley et al., 2021). To bridge this gap, the study proposed a healthcare-specific ethical decision-making framework that embeds ethical reasoning across the AI lifecycle, from design and data governance to deployment and post-implementation monitoring.

The proposed framework emphasizes human oversight, explainability, fairness, and shared accountability, ensuring that AI functions as a clinical support tool rather than an autonomous authority. By integrating ethical checkpoints into operational decision points, the framework offers practical guidance that aligns technological innovation with established healthcare values. This approach not only supports safer and more equitable AI deployment but also contributes to sustaining patient trust and professional integrity in AI-assisted healthcare delivery (Floridi et al., 2022; WHO, 2021).

Table 1. Summary of Ethical Challenges and Framework Responses

Ethical Challenge	Impact on Healthcare AI	Framework Response
Algorithmic bias	Inequitable patient outcomes	Bias audits and representative data validation
Lack of transparency	Reduced trust and accountability	Explainable AI and clinician-facing interpretations
Accountability ambiguity	Unclear responsibility in adverse outcomes	Explicit role and responsibility mapping
Reduced clinician autonomy	Over-reliance on automated outputs	Clinician-in-the-loop decision authority
Ethical drift over time	Degradation of ethical compliance	Continuous ethical monitoring mechanisms

6.2 Future Research Directions

Despite the contributions of this study, several avenues for future research remain open. First, empirical validation of ethical decision-making frameworks in real clinical environments is needed. Future studies should examine how proposed frameworks perform when integrated into hospital workflows, decision-support systems, and regulatory processes. Such evaluations would provide evidence on usability, effectiveness, and clinician acceptance (Amann et al., 2020).

Second, future research should focus on developing quantitative metrics for ethical evaluation. While ethical principles are inherently normative, measurable indicators for fairness, transparency, and accountability would enable systematic monitoring and comparison of AI systems across healthcare institutions. This would support evidence-based ethical governance rather than reliance on abstract ethical commitments alone (Morley et al., 2021).

Third, interdisciplinary research involving clinicians, ethicists, legal experts, and AI developers is essential to refine accountability models. As AI systems become more autonomous and adaptive, determining shared responsibility across stakeholders will require clearer legal and ethical frameworks aligned with national and international regulations (Gerke et al., 2020).

Finally, future studies should explore the ethical implications of emerging AI paradigms, such as generative AI and self-learning clinical systems. These technologies introduce new challenges related to unpredictability, continuous learning, and long-term ethical oversight, which existing frameworks may not fully address (Floridi et al., 2022).

Table 2. Future Research Directions in Ethical AI for Healthcare

Research Area	Key Focus	Expected Contribution
Empirical validation	Real-world testing of ethical frameworks	Practical feasibility and impact assessment
Ethical metrics	Quantifiable indicators of ethics	Objective ethical monitoring
Accountability models	Shared legal and moral responsibility	Clear governance structures
Emerging AI systems	Generative and adaptive AI ethics	Updated and resilient frameworks
Policy integration	Alignment with healthcare regulation	Scalable ethical adoption

6.3 Final Remarks

By aligning ethical theory with clinical practice, this study contributes a structured and actionable approach to ethical decision-making for AI in healthcare. The proposed framework and identified research directions aim to support responsible innovation, ensuring that AI-driven healthcare advances patient well-being while upholding fundamental ethical standards.

References

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20*(1), 1–9.

Char, D. S., Shah, N. H., & Magnus, D. (2020). Implementing machine learning in health care: Addressing ethical challenges. *New England Journal of Medicine, 378*(11), 981–983.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2022). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Ethics and Information Technology, 24*(1), 1–14.

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial Intelligence in Medicine, 100*, 101–109.

McCradden, M. D., Joshi, S., Anderson, J. A., Mazwi, M., & Goldenberg, A. (2020). Patient safety and quality improvement: Ethical principles for deploying AI in healthcare. *Journal of the American Medical Informatics Association, 27*(3), 491–500.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: Translating AI ethics into practice. *Science and Engineering Ethics, 27*(1), 1–34.

Paracha, W. T., Inam, H., & Manzoor, M. (2025). HEARTSMART: Improved CVD risk prediction via recursive feature elimination: Validation on extended dataset. *Spectrum of Engineering Sciences, 3*(6), 1093–1120.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2022). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine, 175*(3), 400–402.

Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in machine learning for medicine. *Communications Medicine, 1*(1), 1–3.

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization.