DERLEME MAKALE / REVIEW ARTICLE
İstanbul Esenyurt Üniversitesi Sağlık Bilimleri Dergisi / İESÜN Sağ Bil Derg
Istanbul Esenyurt Universty Journal of Health Sciences / IESUN Health Sci J
ISSN: 3023-5049
Doi: https://doi.org/

Artificial Intelligence Ethics In Healthcare Management: A Review Of Patient Privacy, Bias, And Transparency

Ufuk Burak KARCIOĞLU[1]

[1] Ufuk Burak KARCIOĞLU, PhD (Candidate), İstinye University, Institute of Health Sciences, Department of Health Management, Istanbul, Turkey, 0009-0006-9131-6407

Geliş Tarihi / Received: *31.11.2025,* ***Kabul Tarihi / Accepted:*** *24.03.2026,* ***Yayın Tarihi/ Published:*** *30.04.2026*

ABSTRACT

Aim: This study examines the major ethical challenges associated with the use of artificial intelligence (AI) in healthcare management and identifies frameworks necessary for fair, transparent, and trustworthy AI implementation.
Materials and Methods: This study adopts a narrative and conceptual review approach to explore the ethical implications of AI in healthcare management. Rather than attempting an exhaustive review, the study synthesizes influential and thematically relevant literature addressing patient privacy, algorithmic bias, and transparency. Peer-reviewed articles, policy documents, and ethical frameworks published primarily between 2018 and 2025 were examined to identify dominant themes, areas of convergence, and ongoing debates. Selected studies were analyzed comparatively to develop a literature-informed ethical framework relevant to healthcare management.
Results: AI-driven predictive analytics and decision-support systems rely heavily on sensitive data, increasing the risk of data breaches and unauthorized access. Privacy-preserving approaches such as federated learning and differential privacy are therefore critical. Additionally, algorithms trained on non-representative datasets may perpetuate healthcare inequities, while the opacity of complex AI models can reduce transparency and accountability.
Conclusion: The absence of standardized ethical metrics and comprehensive governance frameworks limits the sustainable adoption of AI in healthcare management. Integrating technical, organizational, and regulatory measures, supported by stakeholder engagement, is essential for achieving ethical and trustworthy AI use in healthcare.
Keywords: Artificial Intelligence, Health Services Administration, Confidentiality, Algorithms

Sağlık Yönetiminde Yapay Zekânın Etik Boyutları: Hasta Gizliliği, Algoritmik Önyargı ve Şeffaflık Üzerine Derleme

ÖZET
Amaç: Bu çalışmanın amacı, sağlık yönetiminde yapay zekâ kullanımına eşlik eden etik sorunları hasta gizliliği, algoritmik önyargı ve şeffaflık boyutları çerçevesinde kavramsal ve analitik olarak incelemektir. Çalışma, mevcut literatürde ağırlıklı olarak klinik uygulamalara odaklanan etik tartışmaların, sağlık yönetimi karar alma süreçlerine yeterince yansıtılmadığı varsayımından hareket etmektedir.
Gereç ve Yöntem: Bu çalışma, sistematik derleme iddiası taşımayan, anlatısal ve kavramsal bir derleme niteliğindedir. Sağlık yönetimi, yapay zekâ etiği ve dijital yönetişim alanlarında yayımlanmış güncel ulusal ve uluslararası çalışmalar karşılaştırmalı ve yorumlayıcı bir yaklaşımla analiz edilmiştir.
Bulgular: Bulgular, yapay zekâ destekli yönetimsel karar sistemlerinin veri gizliliği risklerini artırabildiğini, temsili olmayan veri setlerinin sağlık eşitsizliklerini kurumsallaştırabildiğini ve opak algoritmaların yönetsel hesap verebilirliği zayıflattığını göstermektedir. Literatürde bu risklerin teknik çözümlerle giderilebileceğine dair ortak bir görüş bulunmakla birlikte, örgütsel ve yönetişim boyutlarının yeterince ele alınmadığı görülmektedir.
Sonuç: Sağlık yönetiminde yapay zekânın etik kullanımı, yalnızca teknik önlemlerle değil; bütüncül yönetişim, şeffaflık ve kurumsal sorumluluk yaklaşımlarıyla mümkün olabilir. Bu çalışma, etik risklerin sağlık yönetimi bağlamında birlikte ele alınması gerektiğini vurgulayan kavramsal bir çerçeve sunmaktadır.
Anahtar Kelimeler: Yapay Zekâ, Sağlık Hizmetleri Yönetimi, Gizlilik, Algoritmalar

Sorumlu Yazar / Corresponding Author: *Ufuk Burak KARCIOĞLU*, PhD (Candidate), İstinye University, Institute of Health Sciences, Department of Health Management, Istanbul, Turkey, 0009-0006-9131-6407
E-mail: burakkarci23@gmail.com

Bu makaleye atıf yapmak için / Cite this article:

INTRODUCTION

The integration of artificial intelligence (AI) into healthcare management has transformed healthcare delivery, offering opportunities to enhance operational efficiency, optimize resource allocation, and improve patient outcomes. AI-driven tools, such as predictive analytics for patient flow, machine learning models for supply chain optimization, and decision support systems for administrative processes, are now integral to modern healthcare systems (Maslej et al., 2025). However, rapid AI adoption introduces complex ethical challenges, notably patient privacy, algorithmic bias, and transparency, which impact trust, equity, and accountability (Al-Amiery, 2025). This review examines these ethical dimensions and evaluates strategies to mitigate associated risks.

Patient privacy is a cornerstone of ethical healthcare. AI systems often rely on large datasets containing sensitive health information from EHRs, wearable devices, and telehealth platforms to predict patient demand, optimize staffing, or streamline logistics (Singh et al., 2025). This reliance raises privacy concerns regarding data security, consent, and misuse. Breaches or unauthorized sharing can erode trust and violate regulations such as GDPR or HIPAA (Pham, 2025). Emerging solutions, including federated learning and differential privacy, balance data-driven insights with patient privacy protection.

Algorithmic bias is another ethical challenge. AI models trained on historical data can perpetuate inequalities, such as disparities in access or treatment outcomes across demographics (Rajkomar et al., 2018). Biased resource allocation algorithms may favor certain populations, worsening inequities. Mitigation strategies include algorithmic auditing and inclusive design. Transparency further complicates ethics in AI. Many models, particularly deep learning systems, act as "black boxes," limiting understanding of decision-making (Kiseleva et al., 2022). Explainable AI frameworks and standardized reporting are critical to fostering trust and accountability.

Existing literature mainly focuses on clinical applications, with administrative decision-making, resource allocation, workforce planning, and operational governance remaining underexplored. Prior reviews often address privacy, bias, or transparency in isolation, lacking an integrated framework that captures their interdependencies in healthcare management. This review provides a conceptual synthesis of ethical challenges in AI for healthcare management and proposes a literature-informed framework to support ethical, accountable, and context-sensitive AI governance.

Patient Privacy in AI-Driven Healthcare Management

The integration of AI into healthcare management relies on large, sensitive datasets, including EHRs, patient demographics, and operational data, to drive predictive models and optimize administrative processes. While enhancing efficiency in patient flow, resource allocation, and supply chain logistics, these approaches introduce significant privacy risks (Pham, 2025). Regulatory frameworks such as GDPR and HIPAA mandate the protection of patient data, yet the scale and complexity of AI challenge traditional safeguards, raising concerns about data security, informed consent, and unauthorized sharing (Weiner et al., 2025).

Privacy risks extend beyond patient–provider interactions, particularly when administrative and clinical datasets are integrated (Pham, 2025; Weiner et al., 2025). Three categories emerge: (1) risks from large-scale data aggregation; (2) governance risks from secondary data use and third-party vendors; and (3) re-identification or inference risks via advanced machine learning. While GDPR emphasizes data minimization and consent, these measures may be insufficient to prevent inference-based privacy violations in management-oriented AI systems (Bouderhem, 2024; Wells et al., 2025), highlighting a tension between regulatory compliance and actual privacy protection.

Privacy Challenges in AI Applications

AI systems in healthcare management require large datasets to train models predicting patient demand, optimize bed allocation, or streamline supply chains. Machine learning may analyze EHRs to forecast emergency admissions or medication usage patterns (Singh et al., 2025). However, aggregating such data increases breach risks, as centralized databases are targets for cyberattacks; the 2017 WannaCry attack illustrates this vulnerability (Bouderhem, 2024). Lack of granular consent raises ethical concerns about autonomy and trust (Osnat, 2025), and sharing data with third-party developers or cloud platforms adds further risk. From a management perspective, privacy challenges affect decision-making and organizational performance. Restrictive policies may limit predictive analytics, while insufficient protections can cause reputational damage, legal sanctions, and loss of patient trust. Managers face a dilemma between maximizing data utility and ensuring ethical compliance, emphasizing that privacy is both a technical and strategic management concern.

Regulatory Frameworks and Their Limitations

Several regulatory frameworks aim to protect patient privacy in the context of AI-driven healthcare management. Table 1 compares key frameworks, including GDPR, HIPAA, and the World Health Organization's (WHO) AI Ethics and Governance Guidelines, highlighting their focus areas, scope, and limitations. GDPR emphasizes data minimization and explicit consent, imposing strict penalties for non-compliance, but its enforcement is limited to the European Union, creating challenges for global healthcare systems (Weiner et al., 2025). HIPAA, while effective in regulating data security within the United States, does not explicitly address the complexities of AI- driven data processing, such as algorithmic re-identification risks. The WHO guidelines provide a global perspective but lack binding authority, relying on voluntary adoption by healthcare organizations (World Health Organization, 2024). These frameworks, while essential, often lag behind the rapid evolution of AI technologies, necessitating complementary technological solutions.

Table 1. Comparative Analysis of Ethical Frameworks Governing AI in Healthcare Management

Framework	Region/ Scope	Key Focus Areas	Strengths	Limitations
GDPR	EU	Data minimization, explicit consent, individual rights, data breach notification	Strong enforcement mechanisms,high fines, user-centric privacy approach	Limited to EU; lacks specific provisions for AI; cross-border data sharing is complex
HIPAA	United States	Safeguarding medical information security standards for covere dentities	Strong within traditional healthcare system s; clear rules for data access and storage	Does not directly address AI- specific risks like algorithmic reidentificati on

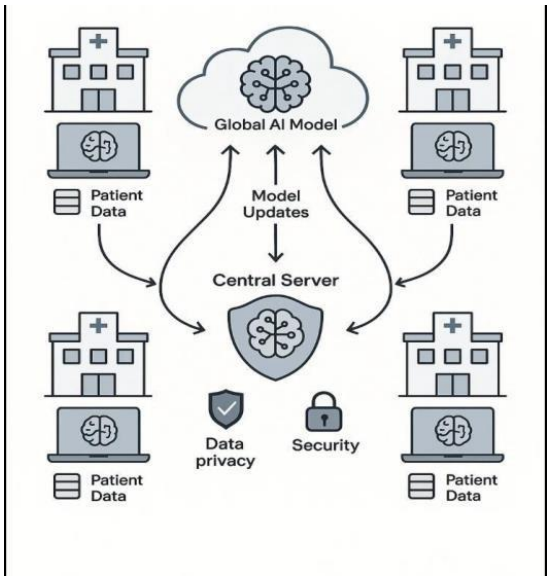
WHO AI Ethics & Governance Guidelines	Global (non-binding)	Ethical AI use, Human rights, accountability, transparency	Offers holistic, Globally İn formed principles; promotes responsible innovation	Non-binding; relies on voluntary compliance; lacks enforcement

Table 1 presents a comparative synthesis of ethical and regulatory frameworks based on recurring themes identified across the reviewed literature, rather than a direct reproduction of any single source.

Technological Solutions for Privacy Protection

To address these privacy challenges, innovative technologies such as federated learning and differential privacy have emerged as promising solutions. Federated learning enables AI models to be trained across decentralized datasets without transferring sensitive patient data to a central server (Wells et al., 2025). As illustrated in Figure 1, this approach involves training models locally on individual healthcare institutions' servers and aggregating only the model updates, thereby minimizing the risk of data exposure. Federated learning has been successfully applied in healthcare management to predict patient admission rates while preserving privacy (El-Geneedy et al., 2025). However, challenges such as computational complexity and the need for standardized protocols limit its widespread adoption.

Figure 1. Conceptual framework of privacy-preserving AI approaches in healthcare management, developed through literature synthesis



Source: Author-developed conceptual framework based on a synthesis of the reviewed literature.

Differential privacy adds controlled noise to datasets, preventing individual patient data from being reverse-engineered from AI outputs (Bouderhem, 2024). This method has been applied in AI-driven resource allocation, though it may reduce model accuracy, creating a trade-off between privacy and performance (Wells et al., 2025). Other approaches, including homomorphic encryption and blockchain-based data management, are being explored, but their scalability in healthcare management is limited (El-Geneed et al., 2025).

Privacy challenges in AI extend beyond technical issues, affecting patient trust and organizational accountability. Failure to address them can erode confidence and hinder AI adoption (Osnat, 2025). Regional disparities in privacy protections underscore the need for global standards accommodating diverse regulatory environments. Future research should develop scalable, privacy-preserving AI frameworks and evaluate their effectiveness in real-world healthcare management. Collaboration among policymakers, healthcare administrators, and AI developers is essential to safeguard patient privacy while maintaining the benefits of AI-driven innovation.

Algorithmic Bias in Healthcare Management

AI in healthcare management has improved resource allocation, patient demand prediction, and administrative efficiency. However, algorithmic bias remains a key ethical challenge, where models trained on historical or non-representative data disadvantage certain groups, such as racial minorities, low-income populations, or women (Agarwal et al., 2023). Bias affects patient prioritization, staff scheduling, and resource distribution, leading to inequitable outcomes (Mackin et al., 2025). This section reviews sources of bias, its impact on equity and trust, and mitigation strategies (see Table 2). Beyond clinical prediction, bias shapes managerial decision-making. Biased demand forecasting influences staffing, bed allocation, and budgets, indirectly affecting patient access and workforce equity (Joseph, 2025; Mackin et al., 2025).

When biased algorithms are embedded in administrative workflows, effects become institutionalized rather than episodic. For example, biased patient flow models may systematically under-resource facilities serving disadvantaged populations, reinforcing structural inequities. This shift from individual- to system-level governance bias is a critical yet underexplored ethical challenge in healthcare management.

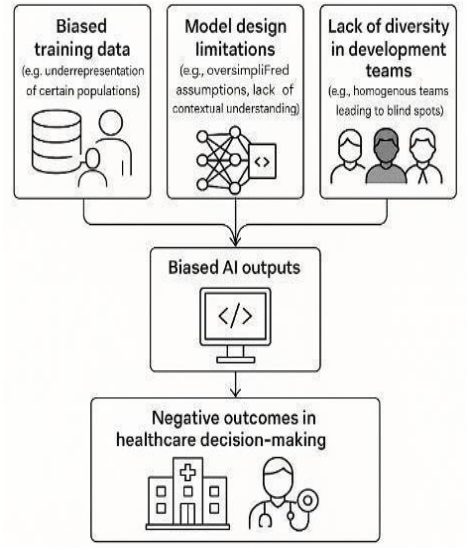
Sources of Algorithmic Bias

Algorithmic bias in healthcare management arises from three sources: biased training data, model design limitations, and lack of diversity in AI development teams (see Figure 2). Training data often reflects historical inequities; underrepresented groups or unequal treatment patterns can lead AI models to prioritize populations with better historical access (Bosco et al., 2025). A landmark study showed a risk prediction algorithm assigned lower scores to Black patients due to cost-based proxies, reducing access (Mackin et al., 2025). Model design, such as algorithm or feature selection, may favor populations with higher historical spending, disadvantaging underserved groups (Aquino et al., 2025). Lack of diversity in development teams can exacerbate bias by overlooking cultural or contextual factors (Nouis et al., 2025).

A synthesis identifies three patterns: (1) biased training data as the primary source; (2) organizational and institutional factors shape outcomes; (3) mitigation requires technical interventions and governance reforms.

These patterns show that algorithmic bias in healthcare management is multidimensional and cannot be resolved through purely technical solutions

Figure 2. Sources of Algorithmic Bias in AI-Driven Healthcare Management



Impacts of Algorithmic Bias

Algorithmic bias in healthcare management affects equity, patient outcomes, and trust. Biased AI systems can produce inequitable resource allocation, prioritizing certain patient groups for hospital beds or staff time, exacerbating disparities in access (Joseph, 2025). For instance, biased predictive models in emergency department triage may underpredict minority needs, causing delayed or inadequate care (Altamirano et al., 2025). These disparities worsen health outcomes and erode trust, particularly in communities historically mistrustful of healthcare systems (Nouis et al., 2025).

Table 2. Case Studies Illustrating Algorithmic Bias in AI-Driven Healthcare Management

Source of Bias	Descrp-tion	Real-World Example	Impact on Healthcare Manageme nt	Mitigation Strategy
Biased Training Data	Historical data reflecting existing social inequalities or missing Representati on of minorities.	An AI tool underestimat ed health needs of Black patients due to cost rights, data breach notification	Unequal resource allocation, lower prioritizatio n for minority groups.	Use representati ve datasets; apply fairness-aware training.
Model Design Limitatio ns	Algorithms not Accounting for social determinants or contextual healthcare differences.	Predictive model failing to consider environment al risk factors affecting low- income patients.	Inaccurate demand forecasting and Misallocati on of services.	Incorporate social context variables into models.
Lack of Team Diversity	Homogeneo us development teams overlooking systemic disparities.	Lack of gender sensitive variables in staff scheduling algorithms.	Gender-biased staffing outcomes	Build multi disciplinary

Strategies to Mitigate Algorithmic Bias

Addressing algorithmic bias requires a multifaceted approach, including data diversification, algorithmic auditing, and inclusive development practices. Ensuring training datasets represent diverse populations, including underrepresented groups, is critical (Agarwal et al., 2023). Stratified sampling can help balance datasets. Algorithmic auditing evaluates models for bias before and after deployment, identifying disparities and enabling recalibration (Altamirano et al., 2025). Standardized frameworks, advocated by organizations like the Algorithmic Justice League, support fairness (Gorelik et al., 2025).

Diverse AI development teams enhance cultural competence and reduce blind spots, while stakeholder engagement, including patients and communities, ensures alignment with ethical principles (Nouis et al., 2025). Fairness-aware machine learning incorporates equity constraints during training (Mackin et al., 2025), though trade-offs between fairness and performance and the need for standardized metrics remain (Joseph, 2025). Models using cost-based proxies tend to underestimate the needs of Black patients, whereas incorporating social determinants improves equity. Data selection and modeling strategies are pivotal, and solutions must extend beyond technical measures to include inclusive design and governance.

Future Directions

Mitigating algorithmic bias in healthcare management requires ongoing research and collaboration among AI developers, healthcare administrators, and policymakers. Future studies should focus on developing scalable fairness frameworks and testing their effectiveness in real-world settings. Additionally, global standards for bias mitigation, similar to those proposed for privacy, could harmonize efforts across diverse healthcare systems. By addressing algorithmic bias, healthcare organizations can enhance equity, improve patient outcomes, and build trust in AI-driven management systems.

Transparency and Explainability in AI Systems

The integration of AI into healthcare management has introduced powerful tools for optimizing administrative processes, such as patient scheduling, resource allocation, and supply chain management. However, the complexity of many AI models, particularly those based on deep learning, often results in "black box" systems, where the decision-making processes are opaque to both healthcare administrators and patients (Kiseleva et al., 2022). This lack of transparency and explainability poses significant ethical challenges, as it undermines accountability, trust, and the ability to validate AI-driven decisions in critical healthcare management contexts. Transparent and explainable AI systems are essential to ensure that stakeholders can understand, scrutinize, and trust the outputs of AI applications (Mohammed, 2025). This section reviews the challenges associated with transparency and explainability in AI- driven healthcare management, examines their implications for ethical governance, and evaluates strategies to enhance the interpretability of AI systems.

In healthcare management contexts, the lack of explainability not only undermines patient trust but also constrains managerial accountability, as administrators may be unable to justify AI-supported decisions during audits, legal reviews, or public scrutiny.

Challenges of Transparency and Explainability

Transparency in AI refers to how understandable a model's internal workings are, while explainability provides clear justifications for its outputs (Mohammed, 2025). In healthcare management, AI systems for predicting admissions or optimizing staff schedules often rely on complex algorithms, such as neural

networks, which are difficult to interpret (El-Geneedy et al., 2025). This opacity creates challenges. Administrators may struggle to justify AI-driven decisions, like prioritizing patients for resources, to stakeholders. For instance, an AI model allocating hospital beds based on opaque risk scores may produce decisions that seem arbitrary or unfair, undermining trust among patients and staff (Osnat, 2025).

Implications for Healthcare Management

The lack of transparency and explainability in AI has significant implications for healthcare management. Opaque decision-making can create mistrust among healthcare professionals, who may hesitate to rely on AI recommendations without understanding their rationale (Nouis et al., 2025), hindering adoption and limiting efficiency gains. At the patient level, opaque AI can erode trust, especially when patients or families do not understand how care decisions are made. For example, a patient prioritization system lacking clear justification may seem arbitrary, causing dissatisfaction or legal challenges (Osnat, 2025). Lack of transparency also complicates accountability. When AI informs critical decisions, such as resource allocation during public health crises, stakeholders need to understand the rationale to ensure fairness and ethical compliance (Ratti et al., 2025). Without explainability, healthcare organizations may struggle to defend AI use under regulatory scrutiny or public criticism, particularly following adverse outcomes. For instance, if an AI model incorrectly prioritizes patients for elective surgeries due to opaque decision-making, it may result in legal liabilities or reputational harm (Pham, 2025).

Technical Approaches

Technical solutions focus on developing interpretable AI models or providing post-hoc explanations for complex models. Rule-based models rely on predefined decision rules, making them suitable for staff scheduling or patient triage (Mohammed, 2025), though their simplicity limits handling complex healthcare tasks. Advanced approaches like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc explanations by identifying features most influential to predictions (El-Geneedy et al., 2025). These methods have been applied to predict patient admission rates, clarifying how variables such as demographics or historical utilization influence outcomes (Osnat, 2025). Attention mechanisms in deep learning highlight which input features the model focuses on during decisions (Ratti et al., 2025). Organizational strategies emphasize transparent AI governance frameworks. Standardized reporting protocols, such as CONSORT-AI, encourage documentation of model design, training, and validation (Gorelik et al., 2025), enhancing transparency for healthcare administrators. Involving multidisciplinary teams, including ethicists, clinicians, and patient representatives, improves contextual relevance and interpretability (Ratti et al., 2025). Engaging community stakeholders in AI-driven resource allocation ensures that systems reflect local healthcare needs and values.

Regulatory Approaches

Regulatory frameworks are critical for mandating transparency and explainability. GDPR includes a "right to explanation" for automated decisions, requiring meaningful information about AI logic (Pham, 2025). Similarly, the U.S. FDA has proposed guidelines for AI-based medical software, emphasizing explainable outputs in healthcare (Provo et al., 2025). However, these regulations mainly target clinical AI, underscoring the need for guidelines tailored to healthcare management applications.

Future Directions

Enhancing transparency and explainability in AI-driven healthcare management requires innovation and collaboration. Future research should develop hybrid models that combine the accuracy of complex algorithms with the interpretability of simpler systems (Mohammed, 2025). Standardized metrics for assessing explainability, such as user comprehension or decision validation accuracy, could benchmark performance. Policymakers should extend regulatory frameworks to address AI's unique challenges in healthcare management, ensuring transparency requirements are practical and enforceable. Training programs for administrators and AI developers can foster a shared understanding of explainability, enabling more effective AI integration into healthcare systems.

Figure 3. Framework for Enhancing Transparency in AI-Driven Healthcare Management

CONCLUSION

Patient privacy, algorithmic bias, and transparency are interrelated ethical concerns in AI-driven healthcare management. Large-scale datasets for predictive analytics introduce privacy risks, including breaches and unauthorized sharing (Pham, 2025). Algorithmic bias from non-representative or historically biased data perpetuates inequities in resource allocation or patient prioritization (Mackin et al., 2025). The opacity of AI systems hinders bias detection, decision validation, and regulatory compliance (Kiseleva et al., 2022). Privacy-preserving techniques like federated learning reduce exposure but may not address embedded biases (Wells et al., 2025), while explainable AI methods such as SHAP or LIME require robust auditing (El-Geneedy et al., 2025). Addressing one challenge alone is insufficient; a comprehensive approach is essential.

Gaps remain. Standardized metrics for evaluating privacy, fairness, and transparency are lacking. GDPR and HIPAA offer guidance but are not tailored to AI-driven administrative applications (Weiner et al., 2025). Scalability of privacy-preserving and explainable AI techniques is limited by computational and cost constraints (El-Geneedy et al., 2025). Research on long-term impacts on trust and equity, especially in low-resource settings, is insufficient, and stakeholder involvement is often overlooked (Ratti et al., 2025; Nouis et al., 2025).

Algorithmic bias is driven by non-representative data and cost-based proxies. While fairness-aware modeling and audits are promoted, structural governance bias is less addressed. Transparency and explainability are critical, though debates persist on the sufficiency of post-hoc methods. Ethical challenges

are interconnected, requiring integrated governance. Administrators must treat ethical AI adoption as both managerial and organizational responsibility. This review provides a conceptual synthesis and proposes a framework for responsible AI governance. Future research should empirically evaluate integrated models, particularly in resource-constrained settings, and examine long-term impacts on equity and institutional trust.

Practical Recommendations for Healthcare AI Implementation

In light of these findings, the following checklist is recommended to guide the implementation of artificial intelligence applications in healthcare organizations in a more ethical, safe, and accountable manner. The key practical recommendations for ethical and responsible AI implementation in healthcare organizations are summarized in Table 3.

Table 3. Practical Ethical Checklist for AI Applications in Healthcare Management

Ethical Focus Area	Description
Patient Consent	Is explicit patient consent for AI-driven data usage and data sharing obtained?
Fairness	Has model fairness been assessed and reported across relevant demographic groups?
Transparency	Is transparency documentation available (e.g., model explanations, version control, decision logic)?
Validation	Has the AI model undergone external or independent validation?
Ethical Oversight	Is an independent AI ethics audit or oversight mechanism planned within the organization?

Conflict of Interest

There is no conflict of interest with any institution or individual within the scope of this study.

Author Contributions

The author contributed to the study design, data collection, analysis, interpretation, and manuscript preparation.

Funding

No Funding.

REFERENCES

Agarwal, R., Bjarnadottir, M., Rhue, L., Dugas, M., Crowley, K., Clark, J., & Gao, G. (2023). Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework. Health Policy and Technology, 12(1), 100702.

Al-Amiery, A. (2025). The ethical implications of emerging artificial intelligence technologies in healthcare. Med Mat, 2(2), 85-100.

Altamirano, S., Vreeken, A., & Ghebreab, S. (2025, October). Machine Learning and Public Health: Identifying and Mitigating Algorithmic Bias through a Systematic Review. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 8, No. 1, pp. 138-151).

Aquino, Y. S. J., Carter, S. M., Houssami, N., Braunack-Mayer, A., Win, K. T., Degeling, C., ... & Rogers, W. A. (2025). Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives. Journal of Medical Ethics, 51(6), 420-428.

Bosco, C. M., Chin, M. H., & Parker, W. F. Addressing Bias in Clinical Algorithms to Advance Health Equity.

Bouderhem, R. (2024). Shaping the future of AI in healthcare through ethics and governance. Humanities and social sciences communications, 11(1), 1-12.

Gorelik, A. J., Li, M., Hahne, J., Wang, J., Ren, Y., Yang, L., ... & Carpenter, B. D. (2025). Ethics of AI in healthcare: a scoping review demonstrating applicability of a foundational framework. Frontiers in Digital Health, 7, 1662642. https://doi.org/10.3389/fdgth.2025.1662642

El-Geneedy, M., El-Din Moustafa, H., Khater, H., Abd-Elsamee, S., & Gamel, S. A. (2025). A comprehensive explainable AI approach for enhancing transparency and interpretability in stroke prediction. Scientific Reports, 15(1), 26048. https://doi.org/10.1038/s41598-025-11263-9

Joseph, J. (2025). Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. Frontiers in Public Health, 13, 1643180. https://doi.org/10.3389/fpubh.2025.1643180

Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. Frontiers in artificial intelligence, 5, 879603. https://doi.org/10.3389/frai.2022.879603

Mackin, S., Major, V. J., Chunara, R., & Newton-Dame, R. (2025). Identifying and mitigating algorithmic bias in the safety net. npj Digital Medicine, 8(1), 335. https://doi.org/10.1038/s41746-025-01732-w

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., ... & Oak, S. (2025). Artificial intelligence index report 2025. arXiv preprint arXiv:2504.07139.

Mohammed, Z. K. (2025). Explainable AI in Health Care: Trust and Transparency in AI-Powered Medical Diagnosis. https://doi.org/10.5772/intechopen.1011279

Nazar, M., Alam, M. M., Yafi, E., & Su'ud, M. M. (2021). A systematic review of human– computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. IEEE Access, 9, 153316-153348. https://doi.org/10.1109/ACCESS.2021.3127881

Nouis, S. C., Uren, V., & Jariwala, S. (2025). Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: a qualitative study of healthcare professionals' perspectives in the UK. BMC Medical Ethics, 26(1), 89. https://doi.org/10.1186/s12910-025- 01243-z

Pham, T. (2025). Ethical and legal considerations in healthcare AI: innovation and policy for safe and fair use. Royal Society Open Science, 12(5), 241873. https://doi.org/10.1098/rsos.241873

Provo, D. H., CCM, P. B., & Cahims, F. (n.d.). The ethical use of artificial intelligence in healthcare: Integrating case management, utilization review, and human collaboration. Retrieved December 11, 2026, from

Ratti, E., Morrison, M., & Jakab, I. (2025). Ethical and social considerations of applying artificial intelligence in healthcare a two-pronged scoping review. BMC Medical Ethics, 26(1), 68. https://doi.org/10.1186/s12910-025-01198-1

Singh, M. P., Keche, Y. N., & Keche, Y. (2025). Ethical Integration of Artificial Intelligence in Healthcare: Narrative Review of Global Challenges and Strategic

Solutions. Cureus, 17(5). https://doi.org/10.7759/cureus.84804

Osnat, B. (2025). Patient perspectives on artificial intelligence in healthcare: A global scoping review of benefits, ethical concerns, and implementation strategies. International Journal of Medical Informatics, 106007. https://doi.org/10.1016/j.ijmedinf.2025.106007

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Weiner, E. B., Dankwa-Mullan, I., Nelson, W. A., & Hassanpour, S. (2025). Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. PLOS digital health, 4(4), e0000810. https://doi.org/10.1371/journal.pdig.0000810

Wells, B. J., Nguyen, H. M., McWilliams, A., Pallini, M., Bovi, A., Kuzma, A., ... & Isreal, M. (2025). A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. NPJ digital medicine, 8(1), 514. https://doi.org/10.1038/s41746-025- 01900-y

World Health Organization. (2024). Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization.

Yu, S., Lee, S. S., & Hwang, H. (2024). The ethics of using artificial intelligence in medical research. Kosin Medical Journal, 39(4), 229-237. https://doi.org/10.7180/kmj.24.140