Mustika Justice
Nomor 1 Volume 5
ISSN: 2776-4257
The article is published at https://jurnal.uic.ac.id/mustikajustice

Justice and Privacy in AI Healthcare

Cecep Mustafa[1], Rita Komalasari[2]

Ibnu Chaldun University, Indonesia[1]
Yarsi University, Indonesia[2]

Correspondent Email:cecepmustafa97@gmail.com

Abstrak
Keadilan merupakan inti dari integrasi etis kecerdasan buatan (AI) dalam bidang kesehatan, menuntut agar inovasi menjunjung tinggi keadilan, kesetaraan, dan hak-hak pasien. Studi ini meneliti bagaimana organisasi pelayanan kesehatan dapat melindungi privasi pasien dan mempromosikan keadilan sambil memanfaatkan kecerdasan buatan (AI) untuk meningkatkan kualitas perawatan. Melalui tinjauan pustaka terhadap studi empiris, kerangka regulasi, dan analisis etika, penelitian ini mengeksplorasi tantangan etis, hukum, dan keadilan yang muncul akibat ketergantungan AI pada data pasien yang bersifat sensitif. Temuan menunjukkan adanya kesenjangan yang signifikan dalam regulasi dan mekanisme perlindungan saat ini, terutama terkait keadilan dan akuntabilitas dalam sistem berbasis AI. Untuk mengatasinya, penelitian ini mengusulkan kerangka kerja holistik yang mengintegrasikan prinsip Privacy by Design (PbD), AI yang etis, dan prinsip berorientasi keadilan. Studi ini menyimpulkan bahwa pendekatan yang seimbang sangat penting untuk mencapai inovasi teknologi sekaligus menegakkan keadilan dalam pelayanan kesehatan berbasis AI.
Kata Kunci: kesehatan berbasis AI, privasi data, AI yang etis, privacy by design, keamanan data pasien, keadilan.

Abstract
Justice lies at the core of ethical AI integration in healthcare, demanding that innovation uphold fairness, equity, and patient rights. This study examines how healthcare organizations can protect patient privacy and promote justice while leveraging artificial intelligence (AI) to improve care. Using a literature review of empirical studies, regulatory frameworks, and ethical analyses, it explores the ethical, legal, and justice-related challenges arising from AI's reliance on sensitive patient data. The findings reveal significant gaps in current regulations and safeguards, particularly regarding fairness and accountability in AI-driven systems. To address these, the study proposes a holistic framework integrating Privacy by Design (PbD), Ethical AI, and justice-oriented principles. It concludes that a balanced approach is essential to achieving both technological innovation and justice in AI-enabled healthcare.

Keywords: *AI-driven healthcare, data privacy, ethical AI, privacy by design, patient data security, Justice*

Accepted Date: 13 November 2025
Publish Date: 10 Februari 2026

Introduction

The integration of artificial intelligence (AI) in healthcare has the potential to revolutionize patient care, streamline diagnostic processes, and enhance treatment outcomes (Reddy et al., 2020; World Health Organization [WHO], 2021). AI systems require vast amounts of data to function effectively, often relying on patient information to learn, adapt, and provide accurate insights (Price & Cohen, 2019). However, as AI becomes increasingly capable of analyzing sensitive patient data, the question arises: how can we ensure the protection of patient privacy and security? The rapid advancements in AI technology present both opportunities and challenges, particularly in terms of maintaining the confidentiality of personal health information while leveraging the power of AI (Balthazar et al., 2018; Alahmad et al., 2023).

As healthcare organizations increasingly adopt AI technologies, the issue of patient data privacy and security has become an urgent concern (Jobin et al., 2019). While AI promises significant improvements in diagnostic accuracy and treatment outcomes, the collection, use, and storage of sensitive patient data raise serious questions about its protection (Beauchamp & Childress, 2019). A core challenge in the use of AI in healthcare lies in ensuring that the vast amounts of patient data required for training these systems are handled securely and ethically (Floridi & Cowls, 2019). Recent studies highlight the growing importance of addressing this issue. According to a 2021 report by the World Health Organization (WHO), over 60% of healthcare institutions worldwide reported that they were using AI systems in clinical settings, which often necessitate access to detailed patient records (WHO, 2021). As AI systems are increasingly integrated into healthcare processes, the risk of data breaches grows proportionally. In 2020 alone, the U.S. Department of Health and Human Services recorded 599 health data breaches affecting 27 million individuals, the highest number on record (U.S. Department of Health and Human Services [HHS], 2020). This statistic underscores the critical need for robust data security measures in healthcare, particularly when handling AI systems that require large datasets (Mittelstadt, 2019).

The problem is not confined to healthcare organizations but extends to patients who are often unaware of how their data may be used, who has access to it, and for what purposes (Leslie, 2019). A survey conducted by the Pew Research Center in 2020 revealed that only 30% of Americans expressed confidence that their health data would be securely managed by healthcare organizations (Pew Research Center, 2020). This statistic illustrates a significant gap between the growing reliance on AI and public trust in the systems handling personal health information (Price & Cohen, 2019). Given these alarming figures, the need for effective privacy and security frameworks to protect patient data in the age of AI is both timely and critical (Goddard, 2017; Voigt & von dem Bussche, 2017).

This study will explore the dual challenge of safeguarding patient data while enabling AI systems to function effectively. It will examine the ethical and legal implications of using AI in healthcare, including concerns about data privacy, informed consent, and the accountability of AI-driven decisions (Mittelstadt, 2019). Furthermore, the study will analyze various strategies for ensuring data protection, including data anonymization, encryption, and regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe (European Parliament & Council of the European Union, 2016; HHS, 2020).

Literature Review

The integration of artificial intelligence (AI) into healthcare systems has generated extensive scholarly attention, particularly concerning data privacy, security, and ethical governance (Alahmad et al., 2023; Price & Cohen, 2019). Existing research broadly recognizes that while AI has the potential to revolutionize patient care, it also introduces unprecedented privacy risks due to its dependence on large volumes of sensitive patient data (Mittelstadt, 2019; Floridi & Cowls, 2019). Traditional studies on healthcare data protection have largely focused on issues such as data breaches, unauthorized access, and compliance with established frameworks like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) (European Parliament & Council of the European Union, 2016; U.S. Department of Health and Human Services [HHS], 2020). However, these frameworks were designed for conventional data systems and do not adequately address the unique vulnerabilities introduced by AI technologies (Goddard, 2017).

Recent literature highlights several emerging threats specific to AI-driven systems, including model inversion attacks, data poisoning, and adversarial manipulation, which can compromise the confidentiality and integrity of patient information (Fredrikson et al., 2015; Finlayson et al., 2019). While some researchers have examined these threats from a technical standpoint, there remains a lack of comprehensive, interdisciplinary studies that integrate technical, ethical, and legal perspectives (Jobin et al., 2019). Most existing works treat AI as an extension of traditional data systems rather than as a transformative paradigm requiring new forms of accountability and oversight (Leslie, 2019).

Another gap identified in the literature concerns the human dimension of AI integration. Studies on AI adoption in healthcare tend to emphasize efficiency and predictive accuracy, often neglecting the critical role of patient trust, informed consent, and transparency (Pew Research Center, 2020; Beauchamp & Childress, 2019). Ethical analyses of AI have drawn attention to issues of bias, fairness, and accountability, yet these concerns are frequently discussed in isolation from technical security measures (Wachter et al., 2017). Recent findings suggest that public trust in AI-powered healthcare systems is declining due to perceived opacity in data handling practices and inadequate communication about data usage (WHO, 2021). A growing body of work argues for the need to align technical safeguards with ethical and social principles, emphasizing that effective data protection must also account for justice, equity, and patient autonomy (Floridi et al., 2018; Alahmad et al., 2023).

Legal scholarship has similarly pointed to significant regulatory gaps. While GDPR and HIPAA establish foundational protections, they fall short in addressing algorithmic transparency, decentralized learning, and cross-border data flows inherent in AI systems (Voigt & von dem Bussche, 2017). Scholars have proposed evolving these frameworks to include AI-specific provisions—such as requirements for explainability, fairness auditing, and data anonymization protocols tailored to machine learning environments (Mittelstadt, 2019; Price & Cohen, 2019). However, few studies have offered actionable, empirically grounded recommendations for implementing such reforms in real-world healthcare contexts (Jobin et al., 2019).

In the technical domain, innovative privacy-preserving methods such as federated learning, homomorphic encryption, and secure multi-party computation have gained attention as promising solutions (Yang et al., 2019; Bonawitz et al., 2019). These approaches enable AI models to learn from distributed data sources without centralizing sensitive information, thereby mitigating privacy risks (Rieke et al., 2020). Nevertheless, the practical deployment of these methods in healthcare remains limited, and empirical evidence on their effectiveness is still emerging (Alahmad et al., 2023). The literature increasingly calls for integrated frameworks that combine technical robustness with ethical governance and patient-centered accountability (Floridi & Cowls, 2019; Beauchamp & Childress, 2019).

Overall, the existing scholarship underscores a growing consensus that safeguarding patient data in AI-driven healthcare requires more than compliance with existing privacy laws. It necessitates a holistic approach that bridges technical innovation, ethical responsibility, and legal reform (Mittelstadt, 2019; Floridi et al., 2018). The literature collectively points toward the urgent need for a comprehensive framework that balances innovation with justice, ensuring that AI technologies enhance healthcare delivery without compromising patient rights, equity, or trust (Alahmad et al., 2023; Jobin et al., 2019).

Method

This study employs a literature review methodology to analyze existing research on data privacy and security in AI-driven healthcare systems. The review synthesizes empirical studies, regulatory frameworks, and ethical considerations to identify gaps in current approaches. The analysis focuses on recent publications (2018 or later) to ensure relevance and timeliness. Data analysis involves examining case studies of AI data breaches, public trust surveys, and regulatory reviews, drawing on both qualitative and quantitative data. By integrating findings from diverse sources, the study provides a comprehensive understanding of the challenges and solutions related to patient data protection in AI healthcare.

A suitable theoretical foundation for comprehensively analyzing AI-driven healthcare systems—addressing both ethical and technical challenges—is the Privacy by Design (PbD) framework combined with the Ethical AI Framework (Cavoukian, 2011; Floridi & Cowls, 2019). These frameworks together provide a robust basis for examining the intersection of technology, ethics, and privacy in healthcare applications of AI. The Privacy by Design (PbD) framework, originally developed by Ann Cavoukian, emphasizes embedding privacy protections into the design phase of systems and technologies rather than treating privacy as an afterthought (Cavoukian, 2011). PbD aligns with the objectives of AI-driven healthcare research by ensuring that patient data security is considered throughout every stage of AI implementation—from data collection and processing to analysis and storage (Goddard, 2017). It promotes proactive strategies to minimize privacy risks, such as data minimization, anonymization, and encryption (European Data Protection Board [EDPB], 2020). These principles can directly inform actionable solutions for safeguarding patient information while enabling the effective use of AI for clinical and operational improvements.

The Ethical AI Framework complements PbD by addressing the broader ethical dimensions of AI, including fairness, accountability, transparency, and the mitigation of bias (Jobin et al., 2019; Floridi et al., 2018). This framework ensures that AI technologies are developed and deployed in ways that respect patient rights, ensure informed consent, and protect data from misuse (Mittelstadt, 2019). Ethical AI also emphasizes explainability and algorithmic transparency, both essential for fostering public trust and ensuring that AI systems align with societal values and regulatory expectations (Leslie, 2019; World Health Organization [WHO], 2021).

Together, these frameworks guide the comprehensive evaluation of AI in healthcare by integrating technical robustness with ethical responsibility. By embedding privacy and ethical considerations from the design stage onward, they enable the creation of AI-driven healthcare systems that not only enhance efficiency and accuracy but also promote justice, equity, and patient autonomy (Floridi & Cowls, 2019; Beauchamp & Childress, 2019). This integrated theoretical approach ensures that both the technical and moral dimensions of data protection are accounted for while developing practical, privacy-preserving solutions for the healthcare sector.

Findings and Discussion

The Role of AI in Healthcare

Artificial intelligence (AI) has the potential to revolutionize healthcare by significantly improving diagnostic accuracy, enhancing treatment planning, and providing personalized patient care (Topol, 2019; Yu et al., 2018). Through advanced algorithms and data analytics, AI can uncover complex patterns in patient data that may be difficult or impossible for clinicians to detect unaided (Rajkomar et al., 2019). Machine learning (ML) algorithms, in particular, are capable of analyzing large and heterogeneous datasets—including medical imaging, laboratory results, and patient histories—to assist in early disease detection, predict outcomes, and recommend individualized treatment strategies (Jiang et al., 2017).

For instance, AI technologies such as deep learning have achieved remarkable accuracy in radiology, enabling the detection of conditions like cancer, bone fractures, and neurological disorders from medical imaging data (Esteva et al., 2019; Lundervold & Lundervold, 2019). Similarly, natural language processing (NLP) tools are increasingly employed to extract clinically meaningful insights from unstructured textual data, such as electronic health records (EHRs), thereby improving the speed and accuracy of medical decision-making (Shickel et al., 2018; Rajkomar et al., 2018). In addition, AI-driven models are transforming drug discovery and development, accelerating the identification of promising compounds and optimizing clinical trial efficiency through predictive modeling (Zhavoronkov et al., 2019; Topol, 2019).

Despite its vast potential, AI in healthcare introduces significant ethical, privacy, and security challenges. AI systems require access to large, diverse datasets—often containing highly sensitive personal health information—which raises concerns regarding data breaches, consent, and misuse (Price & Cohen, 2019). The aggregation and sharing of such data increases the risk of compromising patient confidentiality, especially when existing regulatory frameworks may not fully address AI's unique vulnerabilities (Mittelstadt, 2019; European Data Protection Board [EDPB], 2020). Moreover, the reliance on patient data for algorithmic learning heightens the need for robust data governance, including strong anonymization protocols and secure data access mechanisms (Veale & Binns, 2017).

As AI technologies become increasingly integrated into clinical workflows, establishing comprehensive guidelines and privacy-preserving frameworks—such as federated learning and homomorphic encryption—becomes critical (Li et al., 2020). Balancing innovation with patient privacy, ethical accountability, and legal compliance is therefore a central challenge in realizing the transformative potential of AI in healthcare (Floridi & Cowls, 2019; Beauchamp & Childress, 2019).

Data Privacy and Security Considerations

The integration of artificial intelligence (AI) into healthcare systems raises substantial concerns regarding data privacy and security, given that these technologies depend heavily on the use of sensitive patient information (Mittelstadt, 2019; Price & Cohen, 2019). AI models are frequently trained on extensive datasets that may include personal health records, medical images, and genomic data (Jiang et al., 2017). However, many of these datasets are not consistently anonymized or stored securely, leaving healthcare organizations vulnerable to cyberattacks and unauthorized access (Veale & Binns, 2017; European Data Protection Board [EDPB], 2020). The healthcare sector has repeatedly been identified as one of the most targeted industries for data breaches, with over 27 million individuals affected by healthcare-related breaches in 2020 alone (U.S. Department of Health and Human Services [HHS], 2020). Such incidents expose highly personal information—such as medical history, diagnostic details, and treatment plans—creating risks of identity theft, financial fraud, and potential harm to patient wellbeing (Kruse et al., 2017).

To mitigate these risks, robust data protection strategies are essential. One foundational measure is data anonymization, in which personally identifiable information (PII) is removed or altered so that individuals cannot be reidentified (Rieke et al., 2020). Anonymization enables the continued use of data for AI model training and health research while minimizing privacy risks. Encryption serves as another critical safeguard, ensuring that even if data are intercepted or accessed by malicious actors, they remain unreadable without proper decryption keys (Goddard, 2017). Additionally, secure data storage protocols—including end-to-end encrypted cloud systems, access controls, and multi-factor authentication—are indispensable for protecting patient data during both storage and transmission (Al-Megren et al., 2022).

Beyond these traditional safeguards, emerging privacy-preserving machine learning techniques provide innovative methods for securing patient data in AI systems (Li et al., 2020). One of the most promising is federated learning, which allows AI models to be trained across decentralized data sources without requiring direct data sharing (Yang et al., 2019). In this approach, model updates rather than raw data are exchanged between nodes, ensuring that sensitive patient information remains localized and secure (Brisimi et al., 2018). This distributed learning paradigm is particularly effective in healthcare settings, where data sensitivity and regulatory compliance are critical (Kaissis et al., 2020).

Implementing these technical safeguards—anonymization, encryption, secure storage, and privacy-preserving learning—is vital to ensuring that AI applications in healthcare do not compromise patient privacy or confidentiality (Floridi & Cowls, 2019). As AI technologies continue to evolve and integrate deeper into healthcare infrastructure, these protective measures must be continuously strengthened and adapted to meet emerging threats (Rieke et al., 2020). Only through the consistent application of comprehensive data privacy and security frameworks can healthcare institutions deploy AI responsibly, maintaining patient trust and upholding the ethical principles of transparency, autonomy, and justice (Beauchamp & Childress, 2019).

Ethical Implications of AI in Healthcare

In addition to technical considerations of data privacy and security, the integration of artificial intelligence (AI) into healthcare introduces a range of ethical challenges that must be carefully addressed to ensure responsible, transparent, and just implementation (Floridi & Cowls, 2019; Mittelstadt, 2019). Addressing these concerns is essential for maintaining public trust in AI systems and safeguarding patient rights throughout their lifecycle. Among the most pressing ethical issues are informed consent, bias in AI models, and the opacity of AI decision-making processes (Jobin et al., 2019; Beauchamp & Childress, 2019).

Informed Consent. Patient autonomy—a foundational principle of biomedical ethics—requires that individuals understand and consent to how their data are collected, used, and analyzed (Beauchamp & Childress, 2019). However, AI systems often operate in ways that are complex or opaque, limiting patients' understanding of how their data contribute to algorithmic decision-making (Vayena et al., 2018). Informed consent in the AI context must extend beyond general permission for data usage; it should include clear, comprehensible explanations of how patient data are processed, analyzed, and potentially shared (Morley et al., 2020). Patients should be aware of how AI algorithms may influence diagnosis, treatment recommendations, and clinical decisions. Healthcare institutions must therefore develop transparent consent mechanisms and ensure patients retain the right to opt out of data sharing when desired (World Health Organization [WHO], 2021).

Bias in AI Models. AI systems trained on large datasets are prone to reflect and reinforce the biases embedded in those datasets (Mehrabi et al., 2021). When training data are unrepresentative or skewed toward specific demographic groups, algorithms can generate inequitable outcomes, disproportionately disadvantageing underrepresented populations (Obermeyer et al., 2019). For instance, diagnostic AI models trained primarily on data from white populations have been shown to underperform when applied to minority groups, leading to disparities in medical accuracy and quality of care (Ghassemi et al., 2021). To address this, AI developers and healthcare institutions must prioritize diverse and representative training data, conduct regular fairness audits, and employ bias-mitigation techniques to ensure equitable model performance (Buolamwini & Gebru, 2018; Rajkomar et al., 2018).

Opacity of AI Decision-Making. A major ethical concern in AI is its "black box" nature—particularly in deep learning systems—where the decision-making process is not easily interpretable by humans (Lipton, 2018). This opacity undermines accountability and trust, as clinicians and patients may struggle to understand or justify an AI's recommendation (Doshi-Velez & Kim, 2017). In healthcare, where transparency and explainability are crucial, explainable AI (XAI) approaches should be adopted to make AI outputs interpretable for both practitioners and patients (Samek et al., 2017). Interpretable AI not only enhances clinician confidence but also ensures that life-impacting decisions can be ethically defended and aligned with medical reasoning.

To effectively address these ethical challenges, comprehensive governance frameworks must prioritize fairness, transparency, accountability, and justice throughout the AI development lifecycle (Jobin et al., 2019; Leslie, 2019). Ethical principles should be integrated from the design phase ensuring that informed consent, algorithmic fairness, and explainability are not afterthoughts but core components of responsible AI deployment (Floridi & Cowls, 2019).

Ultimately, ensuring ethical AI in healthcare requires collaboration among healthcare providers, policymakers, technologists, and patient advocates (Vayena et al., 2018). By prioritizing fairness, transparency, and informed consent, AI systems can enhance patient welfare, promote equitable outcomes, and uphold the foundational principles of healthcare ethics and justice (Beauchamp & Childress, 2019; Floridi & Cowls, 2019).

Legal Frameworks and Regulations

Existing legal frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union form the cornerstone of contemporary patient data protection and privacy governance. These regulations were designed to safeguard health data from unauthorized access and to ensure compliance with standards for data collection, storage, and transmission (Greenleaf, 2018; McGraw, 2013). Under HIPAA, healthcare providers, insurers, and affiliated entities are required to implement stringent administrative, technical, and physical safeguards to prevent the misuse of protected health information (U.S. Department of Health & Human Services [HHS], 2020). Similarly, the GDPR establishes robust data protection provisions, mandating explicit consent for data collection, ensuring the right to access and erasure, and enforcing accountability among data controllers (European Parliament & Council, 2016).

However, both HIPAA and GDPR were conceived before the widespread integration of artificial intelligence (AI) technologies into healthcare, leaving significant gaps in addressing the ethical and technical complexities of AI-driven data processing (Brey, 2021; Price & Cohen, 2019). HIPAA, for instance, was primarily developed to manage privacy within traditional healthcare systems and does not directly address AI-specific challenges such as algorithmic transparency, federated data-sharing, or continuous machine learning on patient data (Cohen & Mello, 2018). AI systems often require vast, diverse datasets to train predictive models, leading to increased risks related to data sharing and re-identification—issues not explicitly covered by HIPAA's privacy or security rules (Gerke et al., 2020). The act's limited guidance on secondary data use and de-identification standards further complicates compliance in AI-driven research contexts.

Similarly, while the GDPR is regarded as one of the most comprehensive data protection laws globally, it faces limitations in governing AI's "black box" decision-making processes. Although Article 22 of the GDPR grants individuals the right to obtain "meaningful information about the logic involved" in automated decisions, AI systems often operate in opaque ways that make such explanations difficult to provide (Wachter et al., 2017). This tension between algorithmic opacity and the right to explanation highlights a critical regulatory gap, as patients and healthcare professionals may be unable to fully comprehend how AI systems generate diagnoses or treatment recommendations (Selbst & Powles, 2017). Moreover, the GDPR's consent mechanisms were designed for static data uses, not for continuous or evolving AI models that rely on dynamic data inputs (Veale & Edwards, 2018).

Given the accelerating adoption of AI in healthcare, regulatory modernization is urgently needed to ensure that patient privacy, data integrity, and algorithmic accountability are preserved. Updated legal frameworks should explicitly address data-sharing protocols, anonymization standards, and AI transparency obligations (Leslie, 2019; Floridi et al., 2021). Specifically, anonymization and pseudonymization procedures must be standardized to prevent re-identification risks, while new guidelines should mandate explainability and periodic audits of AI systems to verify compliance with ethical and legal standards (Goodman & Flaxman, 2017).

Furthermore, policymakers should consider developing AI-specific healthcare regulations that complement existing laws. These could include requirements for algorithmic impact assessments, fairness audits, and accountability mechanisms to ensure that healthcare organizations remain responsible for the decisions made by AI systems (European Commission, 2021). By embedding these principles into legislation, governments can ensure that AI-driven healthcare innovation advances patient care while maintaining public trust and upholding fundamental rights.

Ultimately, while HIPAA and GDPR provide a strong foundation for protecting patient data, they must evolve to meet the challenges posed by AI's complexity, scale, and opacity. A revised, AI-aware legal framework—grounded in transparency, fairness, and accountability—will be essential to ensuring that AI technologies in healthcare are both ethically sound and legally compliant (Cohen & Mello, 2019; Brey, 2021).

Developing Justice-Focused, Practical, and Actionable Solutions

The integration of artificial intelligence (AI) into healthcare presents both transformative opportunities and serious challenges concerning patient data privacy, ethical governance, and justice. To ensure that AI systems are deployed safely and equitably, healthcare institutions must adopt practical, actionable solutions that prioritize data protection, ethical accountability, and respect for patient rights. A combination of regulatory innovation, technical safeguards, and ethical design principles offers a pathway toward responsible and just AI adoption in healthcare.

One effective approach is the implementation of Privacy by Design (PbD), a proactive framework that embeds privacy safeguards into the architecture of AI systems from the earliest stages of development (Cavoukian, 2011). Rather than retrofitting privacy features after system deployment, PbD integrates principles such as data minimization, purpose limitation, and anonymization directly into design workflows. This ensures that only the minimum necessary data are collected and that identifiable patient information is either removed or protected through encryption and secure data management practices (Tene & Polonetsky, 2014). By operationalizing privacy as a default setting, PbD enhances both security and patient trust, aligning technological innovation with fundamental ethical commitments to autonomy and confidentiality.

Complementing PbD, the Ethical AI Framework provides essential guidance for ensuring that AI systems uphold fairness, accountability, and transparency throughout their lifecycle (Jobin et al., 2019; Floridi et al., 2021). This framework emphasizes explainability—the ability of AI systems to provide human-understandable justifications for their decisions—and accountability, ensuring that healthcare providers remain responsible for AI-driven outcomes. Regular bias audits and ethical impact assessments are necessary to detect and mitigate algorithmic discrimination, especially against marginalized or underrepresented groups (Leslie, 2019). Integrating ethical AI principles ensures that patient rights and justice are not subordinated to efficiency, reinforcing trust and legitimacy in data-driven healthcare.

Emerging privacy-preserving technologies further enhance the feasibility of ethical AI deployment. Homomorphic encryption allows computation on encrypted data without exposing sensitive information, thereby preserving confidentiality during data processing (Acar et al., 2018). Similarly, federated learning enables distributed model training across multiple institutions without transferring raw patient data, ensuring that sensitive health information remains locally stored (Yang et al., 2019). These innovations allow healthcare organizations to collaborate on AI research while maintaining stringent privacy protections and regulatory compliance.

Together, Privacy by Design, the Ethical AI Framework, and privacy-preserving technologies such as homomorphic encryption and federated learning form a comprehensive strategy for secure and just AI adoption in healthcare. These frameworks operationalize ethical principles into concrete, enforceable measures that protect patient data, reduce bias, and ensure transparency in decision-making. By combining these approaches, healthcare institutions can harness AI's full potential while maintaining public trust, safeguarding privacy, and advancing justice in digital health systems.

Conclusion

Artificial intelligence (AI) offers transformative potential for healthcare by improving diagnosis, treatment, and patient outcomes; however, this advancement must be guided by the principle of justice to ensure that innovation promotes fairness, equity, and respect for patient rights. As AI depends on sensitive health data, protecting privacy and ensuring ethical data use are essential to maintaining both security and social integrity. While existing regulations such as HIPAA and GDPR provide important foundations, they remain insufficient to address AI's unique challenges, including bias, opacity, and accountability. A justice-oriented approach requires evolving these frameworks to guarantee equitable access, prevent discrimination, and safeguard patient autonomy. To realize this vision, healthcare institutions must adopt Privacy by Design (PbD), Ethical AI principles, and privacy-preserving technologies such as federated learning and homomorphic encryption, while policymakers should embed justice into AI governance through transparency mandates, fairness audits, and mechanisms for redress. Centering justice within AI-driven healthcare ensures that technological innovation not only enhances patient care but also upholds equity, dignity, and public trust.

References

Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. ACM Computing Surveys, 51(4), 79–110. [https://doi.org/10.1145/3214303]

Alahmad, G., Al-Mohaimeed, A., & Alzahrani, M. (2023). Artificial intelligence in healthcare: Ethical and legal challenges. Journal of Medical Ethics, 49(2), 134–140. [https://doi.org/10.1136/medethics-2021-108051]

Balthazar, P., Harri, P., Prater, A., & Safdar, N. M. (2018). Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. Journal of the American College of Radiology, 15(3), 580–586. [https://doi.org/10.1016/j.jacr.2017.11.035]

Beauchamp, T. L., & Childress, J. F. (2019). Principles of biomedical ethics (8th ed.). Oxford University Press.

Cavoukian, A. (2011). Privacy by Design: The 7 foundational principles. Information and Privacy Commissioner of Ontario.

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679: General Data Protection Regulation (GDPR). Official Journal of the European Union.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). [https://doi.org/10.1162/99608f92.8cd550d1]

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2021). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 31(1), 1–24. [https://doi.org/10.1007/s11023-020-09585-2]

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European regulation that has a global impact. International Journal of Market Research, 59(6), 703–705. [https://doi.org/10.2501/IJMR-2017-050]

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. [https://doi.org/10.1038/s42256-019-0088-2]

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. [https://doi.org/10.1038/s42256-019-0114-4]

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37–43. [https://doi.org/10.1038/s41591-018-0272-7]

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. Journal of the American Medical Informatics Association, 27(3), 491–497. [https://doi.org/10.1093/jamia/ocz200]

Tene, O., & Polonetsky, J. (2014). A theory of creepy: Technology, privacy, and shifting social norms. Yale Journal of Law and Technology, 16(1), 59–102.

Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. Science Robotics, 2(6), eaap6950. [https://doi.org/10.1126/scirobotics.aap6950]

World Health Organization. (2021). Ethics and governance of artificial intelligence for health. WHO Press. [https://www.who.int/publications/i/item/9789240029200]

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19. [https://doi.org/10.1145/3298981]