International Journal of Research in Medical Sciences
Falayi KO et al. Int J Res Med Sci. 2026 Oct;14(10):4735-4745
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20263547

Systematic Review

Patient safety governance in artificial intelligence-enabled healthcare: a systematic review and the artificial intelligence patient safety governance model

Kehinde Oluwagbenga Falayi[1]*, Prince Chukwuemeka Ekeocha[2], Oladayo Mary Akintola[3], Ome Valentina Akpughe[4], Louis Chidozie Nlewam[5], Olanrewaju Olufemi Olawumi[6], Oluwatoyin Aanu Ayeni[7]

[1]Department of Medicine, Obafemi Awolowo University, Ile-Ife, Nigeria
[2]Robert H. Smith School of Business, University of Maryland, Maryland, USA
[3]Department of Computer Science, University of Hertfordshire, Hertfordshire, UK
[4]Health Informatics, Harrisburg University of Science and Technology, Harrisburg, USA
[5]Department of Public Health, Missouri State University, Springfield, MO 65897, USA
[6]University Hospital of Derby and Burton NHS Foundation Trust, Derby, UK
[7]Department of Social Works, University of Ibadan, Nigeria

Received: 24 March 2026
Revised: 21 May 2026
Accepted: 09 September 2026

***Correspondence:**
Dr. Kehinde Oluwagbenga Falayi,
E-mail: gbengakehinde@gmail.com

ABSTRACT

Artificial intelligence (AI) is rapidly becoming part of everyday healthcare practice, helping hospitals and clinics diagnose diseases, recommend treatments, monitor patients and manage medical data. While these technologies offer significant opportunities to improve the quality and efficiency of healthcare, they also introduce new risks that could affect patient safety. This study examines how patient safety can be safeguarded in the era of AI-enabled healthcare. Using a systematic literature review (SLR) guided by the PRISMA (Preferred reporting items for systematic reviews and meta-analyses) approach, the study identifies recurring challenges related to unclear accountability in AI-supported decisions, overreliance on automated recommendations and operational monitoring that may expose patients to potential harm if not properly managed. Building on these insights, the study proposes the AI patient safety governance model (AIPSGM), a structured framework designed to ensure that patient safety is maintained throughout the lifecycle of AI deployment in healthcare institutions. The model highlights five governance pillars: safe AI design and validation, human oversight and clinical accountability, data governance and cybersecurity protection, operational monitoring and risk management and ethical and regulatory governance. In addition, the study introduces the AI Patient safety governance index (AIPSGI) as a practical tool that translates governance principles into measurable indicators that can help healthcare institutions assess their readiness for safe AI adoption. The findings suggest that patient safety in the AI era cannot be achieved through technological accuracy alone but requires continuous governance that integrates technological organizational, ethical and regulatory safeguards. Therefore, this study contributes to the emerging field of responsible AI in healthcare and concludes that patient safety in the AI era depends not only on technological capability but also on the strength of governance structures that guide how these technologies are designed, implemented and monitored.

Keywords: Patient safety, AI adoption, Safety governance, Operational monitoring, Risk management

INTRODUCTION

Healthcare systems around the world are currently experiencing a profound technological transformation driven by the rapid integration of AI. AI technologies are increasingly being applied across healthcare settings to support disease diagnosis, treatment planning, patient monitoring, clinical decision-making and hospital management processes. Recent studies show that AI-powered predictive analytics can analyse large volumes of medical data, including electronic health records, medical imaging and genetic information, to improve diagnostic accuracy, enable early disease detection and support personalized treatment strategies.[1,2] The growing adoption of AI is largely motivated by the need to improve healthcare quality while addressing persistent systemic challenges such as medical errors, rising healthcare costs, workforce shortages and increasing patient demand. AI-supported systems have demonstrated strong performance in areas such as cancer diagnosis, neurodegenerative disease prediction, clinical imaging analysis and remote patient monitoring, contributing to improved clinical outcomes and operational efficiency.[3,4]

Similarly, AI-driven telemedicine and Internet-of-Medical-Things (IoMT) platforms now enable continuous patient monitoring and early intervention beyond traditional hospital environments.[5,6] However, alongside these benefits comes an important concern: technological advancement does not automatically guarantee patient safety. While AI systems can reduce human error through automation and predictive decision support, they may simultaneously introduce new categories of risk. Research increasingly highlights challenges related to algorithmic bias, data privacy vulnerabilities, lack of transparency and overreliance on automated recommendations.[7,8]

For example, AI-assisted diagnostic systems may produce accurate predictions yet remain difficult for clinicians to interpret, raising concerns about accountability and informed clinical judgement.[9,10] This tension reflects a broader shift in how patient safety must be understood in AI-enabled healthcare environments. Traditionally, patient safety initiatives focused on preventing human or procedural errors within clinical workflows. In contrast, modern healthcare increasingly operates as a sociotechnical system in which safety outcomes depend on interactions between clinicians, algorithms, institutional policies and regulatory structures. Studies examining smart hospitals emphasize that although technical knowledge about AI adoption exists, governance frameworks remain fragmented and poorly integrated across stakeholders.[11]

In the same way, investigations into AI deployment in hospital settings reveal growing uncertainty regarding liability, ethical responsibility and patient consent when AI-supported decisions lead to adverse outcomes.[12] Emerging evidence further shows that trust remains conditional among both healthcare professionals and patients. Patients often recognize AI's efficiency but continue to demand human oversight, transparency and empathy in clinical decision-making.[13,14] Healthcare professionals likewise express concerns regarding system integration, workflow compatibility and organizational readiness for AI adoption.[15] Even where AI improves efficiency such as reducing documentation burden or optimizing hospital resource allocation, new risks such as algorithmic inaccuracies or loss of clinical nuance may emerge.[16]

Consequently, a critical governance gap has begun to surface. Much of the existing literature focuses either on AI performance improvements or ethical principles in isolation, while relatively limited attention has been given to how healthcare institutions should systematically govern AI systems to ensure continuous patient protection. Scholars increasingly argue that responsible AI deployment requires embedded ethical alignment, explainability, human oversight and institutional accountability mechanisms rather than standalone algorithmic excellence.[17,18] Recent reviews across specialized domains, including obstetric emergencies, dermatology, oncology and nursing practice, consistently conclude that while AI enhances diagnostic precision and operational efficiency, unresolved challenges relating to data governance, professional responsibility and ethical oversight persist.[19,9,20,21] These findings suggest that patient safety in the AI era depends not only on technological capability but also on institutional readiness and governance maturity.

Against this background, the present study responds to the growing need for a structured governance-oriented perspective on AI-enabled patient safety. Specifically, this study develops an AIPSGM that explains how healthcare institutions can align technological innovation with safety oversight, accountability and ethical responsibility. To operationalize this framework, the study further proposes the AIPSGI as a measurable tool for assessing institutional preparedness and governance effectiveness in AI-driven healthcare environments. By shifting the conversation from what AI can do to how AI should be governed, this study contributes to the emerging field of responsible and human-centered AI in healthcare.

Patient safety has always been a central pillar of healthcare. At its simplest, patient safety means protecting patients from preventable harm during medical care.[23] Traditionally, this has focused on reducing medical errors, preventing hospital-acquired infections, avoiding medication mistakes and strengthening communication among healthcare professionals. Over the years, structured reporting systems, safety protocols and quality improvement initiatives have been developed to ensure that healthcare delivery minimizes harm as much as possible. However, as healthcare systems become increasingly digital and technology-driven, the concept of patient safety must evolve beyond human and procedural

errors to include technological influences on care delivery.[8,18]

AI refers to computer systems designed to perform tasks that typically require human intelligence, such as pattern recognition, prediction, classification and decision-making.[3] In healthcare, AI tools are used to analyse medical images, predict disease risks, recommend treatments, monitor patient vital signs, detect medication errors and optimize hospital operations.[2] The appeal of AI lies in its ability to process vast amounts of data quickly and identify patterns that may not be immediately visible to clinicians. This has led to improvements in diagnostic accuracy, early detection of complications, personalized medicine and operational efficiency in hospitals and clinics.[3] However, AI systems are not neutral tools. They learn from historical data, which may contain biases or incomplete information. They can produce highly accurate predictions, yet still operate as "black boxes" where the reasoning behind decisions is unclear. Unlike traditional medical devices, AI systems may continuously update or adapt, which makes their behaviour dynamic rather than fixed. This complexity introduces new dimensions to patient safety that traditional safety frameworks were not designed to manage.[4]

While AI can reduce certain types of medical errors, it may also introduce new risks. Algorithmic bias may lead to unequal outcomes across patient groups if training data are not representative.[10] Overreliance on automated recommendations may weaken critical clinical judgement. Data privacy and cybersecurity vulnerabilities may expose sensitive patient information. When adverse outcomes occur, responsibility may be unclear, whether it lies with clinicians, developers or healthcare institutions. These concerns demonstrate that patient safety in the AI era is no longer limited to preventing individual mistakes; it now requires careful management of complex interactions between humans, algorithms and institutional systems.[12]

This shift highlights the growing importance of governance in AI-enabled healthcare environments. Governance refers to the structures, policies, accountability mechanisms and oversight processes that guide how technologies are implemented and monitored. In the context of AI, governance involves ensuring transparency, maintaining human oversight, monitoring algorithm performance, safeguarding patient data and clarifying responsibility for AI-assisted decisions.[20] Another critical concept emerging from recent discussions is human-AI collaboration. AI is not intended to replace clinicians but to support them. Safe implementation depends on maintaining a balance in which AI enhances clinical judgement rather than overriding it. Effective collaboration requires training healthcare professionals to understand both the strengths and limitations of AI systems, embedding explainability tools into clinical workflows and preserving the ability of clinicians to question or override automated outputs. When human oversight remains central, AI functions as a decision-support tool rather than an autonomous authority.[22,17]

METHODS

This study adopts the SLR approach to examine the growing body of research on AI and patient safety in healthcare. A systematic review was selected because it allows researchers to gather and evaluate existing studies in a structured and transparent way. To guide the review process, the study followed the PRISMA framework. PRISMA is widely used in health and technology research because it provides a clear and transparent method for documenting how studies are identified, screened and selected for inclusion in a review.

Relevant literature was identified through a comprehensive search of major academic databases including Scopus, Web of Science, PubMed and Google Scholar, which collectively cover a broad range of peer-reviewed studies in medicine, health technology and data science. The search strategy used combinations of keywords related to the central themes of the study such as "AI in healthcare," "AI and patient safety," "AI governance in healthcare," "clinical decision support systems" and "ethical AI in healthcare." Boolean operators such as AND and OR were used to connect these keywords and broaden the search results. The initial search produced 512 records related to AI and healthcare. After removing 78 duplicate articles, 434 unique studies remained for further evaluation. The titles and abstracts of the studies were then reviewed to determine whether they were relevant to the focus of this study.

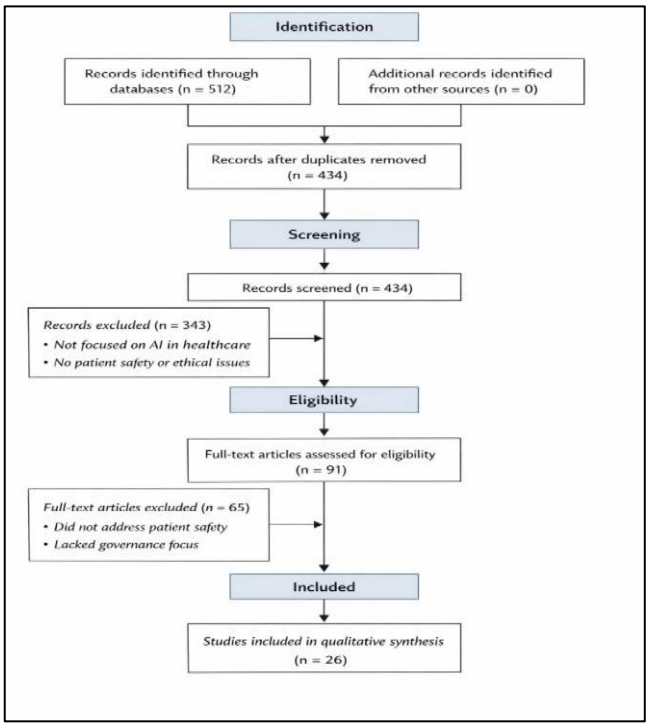

Figure 1: Study selection process diagram.

At this stage, 343 studies were excluded because they did not examine AI within healthcare settings or did not discuss issues related to patient safety, governance, or ethical concerns. The remaining 91 studies were then assessed through full-text review to evaluate their eligibility more carefully. During this stage, 65 articles were excluded because they did not sufficiently address patient safety implications or governance mechanisms related to the use of AI in healthcare environments. As a result, 26 studies met all the inclusion criteria and were retained for the final review. The included studies formed the foundation for the qualitative synthesis and analysis that informed the development of the AIPSGM and the AIPSGI proposed in this study. Once the final set of relevant studies was selected, key information was extracted from each article including details such as authors, year of publication, research focus, methodology used and findings. Through this synthesis process, several recurring themes became clear across the reviewed studies including the growing role of AI in diagnostic accuracy, predictive analytics, telemedicine and clinical decision support, as well as persistent concerns related to data privacy, algorithmic bias, transparency and regulatory oversight, the importance of maintaining human oversight in AI-assisted healthcare systems to ensure patient safety and ethical decision-making.

Table 1: Summary characteristics and quality assessment.

Designs	AI area	Patient safety focus	Governance/ethics focus	Quality rating	Studies
Review study	Predictive analytics and disease diagnosis	Early disease detection and improved treatment accuracy	Data privacy and AI integration challenges	High	1
Conceptual/ ethical review	AI-assisted diagnosis and treatment decisions	Risks of overreliance on AI and patient trust concerns	Transparency, accountability, autonomy and ethical standards	High	7
Review study	AI in patient safety systems	Reduction of medical errors and adverse events	Regulation, transparency and policy concerns	High	8
Review study	AI in Alzheimer's disease detection	Early diagnosis and remote monitoring	Human oversight, bias and privacy concerns	Moderate	10
Narrative review	Clinical AI and predictive analytics	Reduction of diagnostic errors and improved monitoring	Regulatory oversight and patient trust	High	2
Review study	AI in dermatological diagnosis	Improved diagnostic accuracy and personalized treatment	Data privacy and model transparency	Moderate	4
Narrative review	AI in obstetric emergencies	Early detection of maternal complications	Ethical barriers and technological governance	High	19
Review study	Precision medicine and deep learning	Improved diagnostics and treatment personalization	Regulatory compliance and interpretability concerns	High	3
Mixed methods study	AI implementation in smart hospitals	Safe implementation of AI systems	Ethical frameworks and stakeholder coordination	High	11
Experimental study	AI hospital management systems	Reduced waiting time and improved operational efficiency	Data privacy and clinician acceptance	Moderate	15
Mixed methods study	AI-driven telemedicine	Remote monitoring and improved patient engagement	Interoperability, data security and clinician training	High	6
Legal and Policy Analysis	AI-assisted diagnosis systems	Misdiagnosis risks and patient consent issues	Liability, accountability and ethics regulation	High	12
Experimental study	IoMT and AI healthcare monitoring	Early disease diagnosis and real-time intervention	Data protection and monitoring security	Moderate	5
Qualitative study	Patient trust in AI healthcare	Human oversight and patient confidence	Privacy, transparency and cultural acceptance	High	13
Perspective paper	Predictive maintenance systems	Prevention of equipment-related patient harm	Human-centered governance and organizational readiness	Moderate	18
Systematic review	AI nursing documentation systems	Reduction of medication and documentation errors	Human oversight and AI-generated inaccuracies	High	16
Experimental study	AI-assisted cancer diagnosis	Improved diagnosis and reduced treatment risks	Data security, fairness and explainable AI	High	9
Randomized clinical trial	AI clinical decision support	Reduction of hospital-acquired complications	Monitoring of adverse events and implementation safety	High	26

Continued.

Designs	AI area	Patient safety focus	Governance/ethics focus	Quality rating	Studies
Retrospective mixed methods	Technology-related treatment systems	Radiation therapy error prevention	Need for proactive governance mechanisms	High	23
Framework and experimental study	Edge-AI patient monitoring	Real-time monitoring and emergency response	Data privacy and explainable AI concerns	High	25
Scoping review	AI in nursing practice	Protection of patient autonomy and safety	Accountability, transparency and governance readiness	High	20
Systems review	AI-enabled dispensing systems	Prevention of medication-related harm	Continuous monitoring and ethical oversight	High	24
Cross-domain analytical study	Human-aligned AI systems	Reduction of harmful AI decisions	Ethical governance and explainability	High	17

Synthesis of literature and emerging governance gap

The reviewed literature consistently shows that AI is transforming healthcare delivery through improved diagnosis, predictive analytics, telemedicine, personalized medicine, nursing care, and hospital operations. Studies such as found that AI improves diagnostic accuracy, treatment planning, and clinical efficiency, while highlighted its value in specialized areas such as neurological care and dermatology. Similarly, showed that AI-powered telemedicine, wearable technologies, and IoMT systems are improving remote healthcare and real-time patient monitoring.[1-6,10,15,25] further demonstrated AI's role in improving scheduling, bed management, and resource allocation within healthcare institutions.

Despite these benefits, the literature also identified major concerns relating to ethics, patient safety, accountability, transparency, and cybersecurity.[7] emphasized risks such as algorithmic bias and overdependence on automated systems, while noted that AI complicates traditional medical liability structures.[12,13] found that patient trust in AI depends heavily on transparency, human oversight, and data security, whereas [20] highlighted concerns about accountability and depersonalization of care in AI-assisted nursing environments.

A major pattern across the studies is that most research focuses more on AI performance and efficiency than on governance systems capable of managing AI-related risks. Although several studies acknowledged ethical and regulatory concerns, few proposed practical and comprehensive governance frameworks for healthcare institutions.[11] further observed that AI governance knowledge remains fragmented across stakeholders and regulatory systems, while emphasized the need for continuous monitoring, human oversight, and proactive risk management to prevent patient harm.[16,23,24]

The literature therefore reveals a significant governance gap: the absence of a unified framework that integrates patient safety, ethical governance, accountability, cybersecurity, operational monitoring, and regulatory oversight into a single coordinated structure. This gap provides the foundation for the present study, which proposes the AIPSGM and the AIPSGI. These frameworks aim to provide a structured, measurable, and patient-centered governance approach that strengthens accountability, improves transparency, reduces patient safety risks, and supports the responsible adoption of AI in healthcare systems.

AIPSGM

AIPSGM is proposed as a structured framework designed to protect patients throughout the entire lifecycle of AI deployment in healthcare. Model recognizes that patient safety cannot depend on technological accuracy alone. Although AI systems may achieve high diagnostic/ predictive performance, safe healthcare outcomes also depend on proper supervision, implementation, monitoring, and regulatory control within real clinical environments.[1,17] As result, AIPSGM views patient safety as a continuous governance process shaped by interaction between AI technologies, healthcare professionals, institutional policies and regulatory oversight.

Figure 2: AI patient safety governance lifecycle.

The first pillar of the model is Safe AI Design and Validation, which focuses on ensuring that AI systems are responsibly developed and rigorously tested before clinical use. This includes algorithm validation using diverse datasets, bias detection, fairness testing, explainable AI structures, clinical certification, and pre-deployment risk assessments to minimize patient safety risks. The second pillar, Human Oversight and Clinical Accountability, emphasizes that AI should support rather than replace professional clinical judgement. Clinicians remain responsible for final treatment decisions through human-in-the-loop processes, supported by AI literacy training, clear accountability structures, and escalation procedures that allow healthcare professionals to question or override unsafe recommendations.[12,20]

The third pillar is data governance and cybersecurity protection, which focuses on protecting sensitive patient information through secure storage systems, encryption, access controls, regulatory compliance, and continuous cybersecurity monitoring. Ethical data usage policies also ensure that patient confidentiality and autonomy are maintained throughout AI operations. Fourth pillar, operational monitoring and risk management, recognizes that AI systems require continuous supervision after deployment. Real-time performance monitoring, periodic auditing, incident reporting systems, and ongoing feedback mechanisms help healthcare institutions detect errors early, manage performance drift, and reduce the risk of patient harm.[16,23]

The final pillar, Ethical and Regulatory Governance, ensures that AI adoption aligns with principles of fairness, transparency, accountability, and patient rights. This includes ethical review processes, regulatory compliance frameworks, transparency measures, liability structures, and patient consent mechanisms that inform individuals about the role of AI in their care.[12,13] Together, these five pillars provide an integrated and patient-centered governance framework for the safe, responsible, and accountable use of AI in healthcare institutions.

Operationalizing the AIPSGM

To ensure practical applicability, the proposed AIPSGM is operationalized by translating its governance dimensions into measurable institutional indicators that can be assessed within hospitals and clinical environments. The model conceptualizes patient safety in the AI era as a function of five interrelated governance pillars: AI transparency, human oversight, accountability structures, risk management and monitoring and ethical-regulatory compliance.

Each pillar represents an organizational capability required to minimize technology-induced harm while maintaining safe clinical decision-making. The AIPSGM can be implemented using a composite AIPSGI constructed from standardized scores across governance dimensions.

$$AIPSGI = \sum_{i=1}^{n} \frac{w_i G_i}{n}$$

Where: G_i = governance dimension score; w_i = assigned weight of each governance pillar; and n = number of governance dimensions.

The index enables hospitals to quantitatively evaluate their readiness to safely deploy AI technologies across diagnosis, treatment decision-making and patient monitoring systems. The AIPSGI aggregates scores across the governance dimensions of the AIPSGM into a single, interpretable index. Each dimension is rated on a scale from 1 (low readiness) to 5 (high readiness) based on measurable indicators.

Step by step illustration

Step 1: Scoring each dimension

Step 1 scoring each dimension shown in the Table 2.

Table 2: Scoring each dimension.

Governance dimension	Indicators	Score (1-5)	Notes
AI transparency	Explainable AI outputs; clarity of rationale	4	Clear decision explanations
Human oversight	Clinician review and override capability	5	Strong human-in-the-loop policies
Accountability structures	Liability and reporting mechanisms	3	Policies exist but not consistently applied
Risk management and monitoring	Continuous audit and bias checks	4	AI performance monitoring
Ethical and regulatory compliance	Data privacy, informed consent	5	Full compliance with HIPAA/GDPR
Clinician-AI collaboration	Training and usability perception	4	Clinicians trained
Patient trust and safety perception	Patient confidence and engagement	3	Mixed patient trust

Step 2: Assign weights and calculate the composite index

The AIPSGI is calculated as, if equal weights are assumed (w=1 for each)

$$AIPSGI = \sum_{i=1}^{n} \frac{w_i G_i}{n}$$

$$AIPSGI = \frac{4+5+3+4+5+4+3}{7} = \frac{28}{7} = 4.0$$

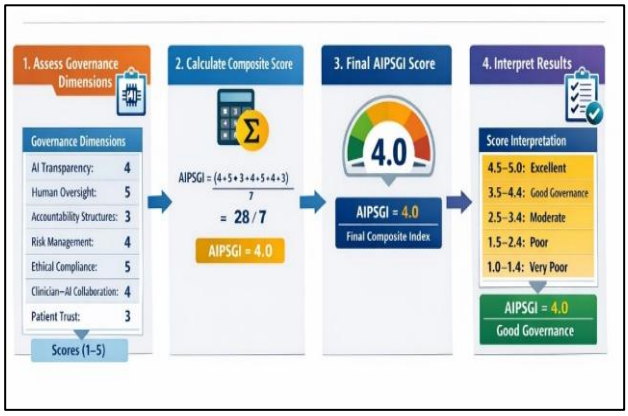

Figure 3: AI patient safety governance workflow.

Step 3: Interpret the score

Step 3 interpret the score shown in the Table 3 below.

Table 3: Interpret the score.

Index range	Interpretation
4.5-5.0	Excellent AI patient safety governance
3.5-4.4	Good governance with minor gaps
2.5-3.4	Moderate governance, requires improvement
1.5-2.4	Poor governance, high risk for AI-related harm
1.0-1.4	Very poor governance, critical risks present

Example: AIPSGI=4.0 → Good governance: the hospital demonstrates strong AI safety practices but may need to improve accountability and patient engagement measures.

Step 4: Practical use

Benchmarking includes compare hospitals or departments. Gap analysis-Identify weak dimensions (e.g., accountability and patient trust). Monitoring over time-Track improvements after AI safety interventions. Policy evaluation included assess regulatory compliance effectiveness.

RESULTS

The included studies examined the use of AI across a wide range of healthcare settings, including disease diagnosis, precision medicine, cancer care, dermatology, neurology, maternal healthcare, and hospital management (Table 4).

Table 4: Summary tables of synthesized findings.

Healthcare area	What the study examined	Main findings	Studies
AI applications in healthcare			
Predictive analytics	AI-driven predictive analytics using EHRs, medical images and genetic data	Machine learning models supported early disease detection and personalized treatment planning.	1
Clinical AI	Machine learning and deep learning in clinical care	AI improved disease detection, medical imaging analysis and clinical decision-making.	2
Precision medicine	AI applications in oncology, cardiology, neurology and pharmacogenomics	AI supported diagnosis and personalized treatment across multiple specialties.	3
Alzheimer's disease	AI-assisted neuroimaging and speech analysis	AI enabled earlier diagnosis and remote patient monitoring.	10
Dermatology	Deep learning for skin lesion recognition	AI accurately identified skin lesions and supported individualized treatment planning.	4
Obstetric emergencies	AI-assisted prediction of maternal complications	AI improved early identification of maternal complications.	19
Hospital management	AI-based scheduling and resource management	AI improved patient scheduling, bed allocation, workflow efficiency and reduced waiting times.	15
AI-driven telemedicine	Wearable devices, smart sensors and machine learning	AI supported remote diagnosis, chronic disease management, patient engagement and continuous monitoring.	6
IoMT framework	AI-enabled IoMT	The proposed model achieved 98.62% disease-classification accuracy and supported early diagnosis through continuous biosensor monitoring.	5
Edge-enabled AI	Edge computing for healthcare monitoring	AI improved emergency response, anomaly detection and real-time monitoring for cardiac events, respiratory distress and elderly fall detection.	25

Continued.

Healthcare area	What the study examined	Main findings	Studies
Patient safety, ethics and governance			
Ethical AI	Ethical issues in AI-assisted healthcare	Reported concerns relating to algorithmic bias, transparency, accountability and overreliance on AI-generated recommendations.	7
Human-centred AI	Ethical governance framework	Reported improvements in fairness, interpretability, trust calibration and human-aligned AI.	17
Nursing practice	AI in nursing	Reported concerns regarding professional autonomy, accountability, privacy, equity, professional identity and depersonalized care.	20
Patient trust	Interviews with 32 patients	Participants viewed AI positively but preferred continued human oversight because of concerns about privacy, transparency and trust.	13
Smart hospitals	AI implementation in hospitals	Reported challenges relating to stakeholder coordination, ethical governance and regulatory guidance.	11
Medical liability	Legal issues in AI-assisted diagnosis	Found that responsibility for patient harm becomes less clear when clinicians and AI jointly contribute to decisions.	12
Medication safety	AI-supported medication safety systems	Reported that safe implementation requires human–AI collaboration, interoperability, continuous monitoring, bias assessment and governance mechanisms.	24
AI monitoring systems	AI-enabled patient monitoring	Reported the need for secure data management, cybersecurity protection and continuous monitoring after deployment.	5 and 25

DISCUSSION

AI-Driven transformation of healthcare delivery

The findings of this study highlight an important reality in modern healthcare: while AI is rapidly improving diagnostic accuracy, treatment planning and patient monitoring, the systems that ensure patient safety have not evolved at the same pace.[1-3] The systematic review revealed that most existing studies focus primarily on the technical benefits of AI, such as improving clinical efficiency, enabling early disease detection and supporting precision medicine.[4,10] However, far fewer studies examine how healthcare institutions should govern these technologies to ensure that patients remain protected from harm, errors, bias or misuse.[11,24]

AI applications in clinical diagnosis and predictive analytics

Across the reviewed studies, AI systems were consistently shown to enhance clinical performance in several areas of healthcare. For example, AI-assisted diagnostic tools have demonstrated strong capabilities in detecting diseases earlier and more accurately by analyzing large clinical datasets and medical images.[1,2] Similarly, predictive analytics tools are increasingly being used to identify potential complications before they become severe, allowing healthcare providers to intervene earlier and improve treatment outcomes.[3] AI systems have also shown promise in supporting clinical decision-making by helping physicians analyze complex patient data and recommending possible treatment pathways.[4,10]

Patient safety risks and ethical concerns

Despite these benefits, the literature also reveals growing concerns related to patient safety. Several studies highlight the risk of algorithmic bias, which may occur when AI systems are trained on datasets that do not adequately represent diverse patient populations.[7,17] When such systems are applied in real-world clinical environments, they may produce inaccurate or unfair predictions for certain demographic groups. This raises serious ethical and safety concerns, particularly in healthcare systems that serve diverse populations. In addition, some studies report the risk of overreliance on AI-generated recommendations, where clinicians may trust algorithmic outputs without adequate verification.[7,20] In such situations, errors produced by the system may go unnoticed and potentially affect patient outcomes.

Data governance and cybersecurity risks also emerged as significant concerns. AI systems rely heavily on large volumes of patient data, including electronic health records, medical images and real-time monitoring data.[5,25] While these datasets enable advanced analytics and predictive capabilities, also create vulnerabilities related to data breaches, unauthorized access and misuse of sensitive health information.[5,6] Without strong governance frameworks to protect patient data, widespread adoption of AI technologies could expose healthcare institutions to significant privacy and security risks.

Governance, accountability, and regulatory gaps

Another important issue identified in the literature relates to accountability and responsibility in AI-supported

clinical decision-making. Traditional medical practice clearly defines the responsibility of healthcare professionals for treatment decisions. However, when AI systems contribute to diagnostic or treatment recommendations, determining responsibility becomes more complex.[12] Several scholars note that it is often unclear who should be held accountable if an AI system makes an incorrect recommendation that leads to patient harm.[11,12]

Need for an integrated patient safety governance framework

The review indicates that most existing studies examine AI applications in isolation rather than from a broader systems perspective. Many studies focus on specific technologies, disease areas or clinical tasks, such as cancer diagnosis, neurological disease prediction or hospital workflow optimization.[10,15] While these contributions are valuable, they often do not consider how AI technologies interact with institutional policies, clinical workflows, regulatory structures and human decision-makers.[11,24]

The findings of this study therefore suggest that patient safety in the AI era cannot be achieved through technological improvements alone. Instead, it requires a comprehensive governance framework that integrates technological design, human oversight, data protection, operational monitoring and ethical accountability.[17,24] In response to this gap, this study proposes the AIPSGM, which provides a structured approach to managing patient safety throughout the lifecycle of AI deployment in healthcare institutions. The AIPSGM emphasizes that safe AI implementation is not a single event but an ongoing governance process. It begins with responsible AI design and validation, ensuring that algorithms are tested for accuracy, fairness and reliability before they are introduced into clinical settings.[1,17] The model then highlights the importance of human oversight and clinical accountability, recognizing that AI systems should support rather than replace medical professionals.[12,20]

The model also emphasizes the critical role of data governance and cybersecurity in protecting patient information. Because AI systems rely heavily on patient data, strong data protection policies are necessary to ensure confidentiality, integrity and ethical use of medical information.[5,6] In addition, the model incorporates continuous operational monitoring, which allows healthcare institutions to track AI system performance after deployment. Through regular audits, incident reporting systems and performance evaluations, potential risks can be identified and addressed before they result in patient harm.[16,23] Finally, the AIPSGM highlights the importance of ethical and regulatory governance. Transparency, accountability and regulatory oversight are essential for maintaining public trust in AI-enabled healthcare systems.[11,17] Clear policies regarding responsibility, liability and patient consent help ensure that both healthcare providers and patients understand how AI technologies are used in clinical decision-making.[12,13]

Industry implications and contribution to knowledge

This study contributes to the growing literature on AI in healthcare by shifting attention from technological performance to patient safety governance. A major contribution is the development of the AIPSGM, which provides a structured framework for understanding how patient safety can be maintained throughout the lifecycle of AI deployment in healthcare institutions. Unlike many previous studies that focus mainly on isolated AI applications, the AIPSGM adopts a broader systems perspective by integrating technology design, human oversight, institutional policies, data governance, and regulatory accountability into a single governance structure.

The study also introduces the AIPSGI, which translates governance principles into measurable dimensions such as algorithm validation, clinician oversight, data protection, operational monitoring, and ethical accountability. This provides healthcare institutions and policymakers with a practical tool for assessing their readiness for safe and responsible AI adoption. Beyond theoretical contributions, the findings offer important industry implications for healthcare institutions, healthcare professionals, AI developers, and policymakers. For hospital administrators, the governance model provides guidance for establishing policies on AI validation, risk management, cybersecurity, and continuous monitoring before AI systems are integrated into clinical practice. For healthcare professionals, the study emphasizes the importance of AI literacy and continuous training so clinicians can critically evaluate AI-generated recommendations rather than rely on them blindly.

The findings also highlight the responsibility of AI developers to design systems that are transparent, explainable, fair, and trained on diverse datasets to minimize bias and improve reliability across patient populations. At the policy level, the study emphasizes the need for stronger regulatory frameworks addressing algorithm certification, accountability, liability, data protection, and ethical oversight. Overall, the study reinforces the idea that successful AI adoption in healthcare depends not only on technological innovation but also on effective governance systems that prioritize patient safety, transparency, and ethical responsibility.

Limitations and future research directions

This study provides useful insights into patient safety governance in the AI era, but several limitations should be acknowledged. First, the study is based on a systematic review of existing literature rather than primary empirical data collected directly from healthcare institutions or professionals. As a result, the findings reflect patterns identified in published studies and may not fully capture

real-world implementation experiences. In addition, the AIPSGM and the AIPSGI are conceptual models developed from literature synthesis and have not yet been empirically tested in healthcare settings.

Future research should therefore focus on empirically validating the proposed governance model and index within real healthcare institutions. Studies could examine how hospitals implement AI governance mechanisms and whether stronger governance systems improve patient safety outcomes. Future researchers may also apply the AIPSGI across different healthcare systems to compare governance readiness and identify best practices. In addition, further studies should explore patients' perceptions of AI governance, trust, and acceptance of AI-assisted healthcare, as public confidence will remain essential for the responsible adoption of AI technologies in healthcare systems.

CONCLUSION

As the use of AI tools grow, ensuring that patients remain protected from harm becomes an equally important concern. Therefore, this study concludes that patient safety in the AI era depends not only on technological capability but also on the strength of governance structures that guide how these technologies are designed, implemented and monitored.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Babu SR, Kumar NBSV, Divya AS, Thanuja B. AI-Driven Healthcare: Predictive Analytics for Disease Diagnosis and Treatment. Int J Modern Trends Sci Technol. 2024;10(06):5-9.
2. Haque A. Artificial Intelligence in Healthcare: Transforming Diagnosis, Treatment, and Patient Care. Academica Global. J Computer Sci Technol Studies. 2025;3(2):73-84.
3. Ali SA, Ali S, Shamim S. Role of Artificial Intelligence in Transforming Diagnosis, Treatment, and Patient Care: A Review. Curr Pharmacogenomics Personalized Med. 2025;XX(X):1-19.
4. Cao F, Yang Y, Guo C, Zhang H, Yu Q, Guo J. Advancements in artificial intelligence for atopic dermatitis: diagnosis, treatment, and patient management. Ann Med. 2025;57(1):2484665.
5. Damera VK, Cheripelli R, Putta N, Sirisha G, Kalavala D. Enhancing remote patient monitoring with AI-driven IoMT and cloud computing technologies. Scientific Rep. 2025;15:24088.
6. Kumar A. AI-Driven Telemedicine: Enhancing Remote Diagnostics and Patient Monitoring. Int J Res Innovat Social Sci. 2025;IX(III):6033-7.
7. Agarwal V, Priyanka BL, Reddy P, Sidhu A, Gupta SG, Rastogi S, et al. Analyzing Ethical Dilemmas in AI-Assisted Diagnostics and Treatment Decisions for Patient Safety. Sem Med Writing Educat. 2024;3:497.
8. Kaduba GS, Yugam, Ranganathan V, Seth N, Chunara F, Chilla S. Artificial Intelligence in Healthcare: Improving Patient Safety Through Innovation. J Med Dental Front. 2024;1(2):1-6.
9. Raiyana J, Alam R. Assuring the Safety of Patient Data in AI-Cancer Diagnosis and Treatment. Int J Med Health Res. 2025;3(4):70-82.
10. Sirigineedi A. Exploring the Benefits and Risks Associated with Incorporating Artificial Intelligence into the Diagnosis, Treatment, and Monitoring of Alzheimer's Disease. Am J Student Res. 2025;3(6):40-52.
11. de Graaf Y, Ahmed A, Sanges C, Herbst L, Vrijhoef M. Societal factors influencing the implementation of AI-driven technologies in (smart) hospitals. PLOS One. 2025;20(6):e0325718.
12. Wang Y, Zhou Z. Medical damage liability risk of medical AI: from the perspective of DeepSeek's large-scale deployment in Chinese hospitals. Front Public Health. 2025;13:1726205.
13. Khan IU, Rehman U, Hanif S, Mehwish H, Almagharbeh WT, Shahid M, et al. Patient perspectives on the use of AI in medical decision-making: Exploring patient trust and acceptance of AI-driven healthcare services. Insights-J Health Rehabilitat. 2025;3(3):416-22.
14. Falayi KO, Osagie MU, Akinola KO, Igwemezie PC, John SA. Building Trust in Smart Hospitals in Developing Countries: A PRISMA Review of Blockchain-Based Health Data Security. Int J Life Sci Res Arch. 2026;10(01):77-86.
15. Khare A, Penubaka KK, Chithrakumar T, Geetha M, Kamalavalli D, Jadhav AB. AI-Driven Patient Flow Management in Hospitals: Reducing Wait Times and Enhancing Care. J Neonatal Surg. 2025;14(10s):696-708.
16. Teimouri GA, Taheri AA, Moeinipou Y, Fathabadi MA, Bagheri F. The Impact of AI-based Nursing Documentation on Time Management and Patient Safety: A Systematic Review. J Res Health. 2025;5:641-60.
17. Mahmood M. Ethical and Human-Aligned Artificial Intelligence for Public Welfare, Financial Integrity, and Pediatric Healthcare Decision Systems. Front Computer Sci Artificial Intelligence. 2025;5(1):13-8.
18. Alkhatib SA, Katmah R, Kosaji D, Bin Afzal SU, Tariq MH, Simsekler MCE, et al. AI-Driven Decision Support Framework for Preventing Medical Equipment Failure and Enhancing Patient Safety: A New Perspective. J Multidiscip Healthcare. 2025;18:6299-313.
19. do Breviário AG, Gagnin ERG, Fidalgo PMM, Giehl L, Nadif B, Albó JA. Artificial intelligence in the early diagnosis of obstetric emergencies. Curr Scientific J. 2025;5(13):516-51.
20. Anazia CN. The ethical dimensions of artificial intelligence in nursing practice: Balancing innovation

and human care. World J Pharmacy Med Sci. 2026;2(2):01-14.
21. Otunlape O. Governance Frameworks for Enterprise AI Systems Operating in Regulated Environments. Int J Computer Applicat. 2026;187(74):13-21.
22. Okolo AO, Ogbeche QE, Abe R, Akinola KO, John SA. Building Products that Think: Integrating AI Intuition into Design Strategy. World J Innovat Modern Technol. 2023;7(2):109-30.
23. Taylor MA, Symington S, Bakhtiari M. Risk Factors and Mitigation Strategies in Radiation Therapy: A Study of Patient Safety Events from 40 Hospitals. Patient Safety. 2026;8(2):154896.
24. Ouda E, Chaabi I, Abualola H, Ramadan MA, Patro PK, Kaya GK, et al. The Role of Artificial Intelligence in Reducing Dispensing Errors for Patient Safety and Quality: A Systems Approach. Risk Management Healthcare Policy. 2025;19:573762.
25. Kumar M, Verma SK. Edge-Enabled IoT Healthcare Monitoring: An AI-Driven Framework for Real-Time Patient Analytics. Int J All Res Educat Scientific Meth. 2026;14(2):391-9.
26. Walsh CG, Long Y, Novak LL, Salwei ME, Tillman B, French B, et al. AI-Driven Clinical Decision Support to Reduce Hospital-Acquired Venous Thromboembolism: A Trial Protocol. JAMA Network Open. 2025;8(10):e2535137.

Cite this article as: Falayi KO, Ekeocha PC, Akintola OM, Akpughe OV, Nlewam LC, Olawumi OO, et al. Patient safety governance in artificial intelligence-enabled healthcare: a systematic review and the artificial intelligence patient safety governance model. Int J Res Med Sci 2026;14:4735-45.