REVIEW Open access

Ethical, Liability, and Regulatory Governance in AI-Embedded Healthcare Systems

Sanjay Kulkarni[1], Meenal Joshi[1*], Rohan Patil[2], Aniket Deshmukh[2]

Abstract

The integration of artificial intelligence (AI) into healthcare systems and analytics has revolutionized clinical workflows, enabling predictive analytics, diagnostic support, and personalized interventions. However, this embedding raises profound ethical, liability, and regulatory challenges that must be addressed to ensure safe, equitable, and effective deployment. This narrative review synthesizes literature governance frameworks for AI-embedded healthcare, focusing on systems-level infrastructure and clinical analytics.

Ethically, AI systems introduce risks of bias amplification, where algorithms trained on non-representative datasets perpetuate disparities in health outcomes, as seen in racial biases in risk prediction tools. Privacy concerns escalate as data mining from digital phenotyping proliferates, necessitating robust consent mechanisms and transparency in algorithmic decision-making. Liability allocation remains ambiguous, particularly for physicians using AI tools, where harms from opaque “black-box” models complicate accountability among developers, clinicians, and institutions. Regulatory governance demands a shift from product-centric to system-view approaches, incorporating human-AI interactions, ongoing monitoring, and adaptive oversight, as proposed for AI/ML-based software as medical devices (SaMD).

In healthcare systems, AI analytics facilitate end-to-end loops from data ingestion to intervention feedback, but require governance to mitigate distributional shifts and automation complacency. Clinical decision support systems (CDSS) exemplify this, where AI augments human judgment but risks reinforcing outdated practices without ethical recalibration. Radiology is a key domain, and AI in imaging analytics underscores the need for multisociety ethical statements and regulatory vetting.

This review provides an original synthesis that structures AI governance across data ecosystems, model transparency, deployment integrity, and feedback mechanisms. It underscores the imperative for interdisciplinary frameworks that prioritize patient well-being, fairness, and accountability, while avoiding over-speculation. By integrating cross-study insights, we position governance as integral to AI’s infrastructural role in healthcare, advocating for actionable ethics to bridge regulatory gaps and enhance the reliability of clinical analytics. Ultimately, effective governance will enable AI to converge with human expertise, fostering high-performance medicine without compromising equity or safety.

Keywords Artificial intelligence, Healthcare systems, Clinical analytics, Ethical governance, Liability allocation, Regulatory frameworks

*Correspondence:
Meenal Joshi
meenal.joshi@gmail.com
[1] Department of AI in Healthcare Systems, School of Medicine, Savitribai Phule Pune University, Pune, India
[2] Department of Clinical Data Engineering, School of Engineering, Indian Institute of Technology Bombay, Mumbai, India

Introduction

Evolution of AI integration in healthcare infrastructure

The advent of artificial intelligence (AI) in healthcare has marked a transformative shift from traditional rule-based systems to advanced machine learning (ML) paradigms capable of processing vast datasets for predictive and diagnostic purposes. AI applications have increasingly been embedded within healthcare infrastructures, encompassing electronic health records (EHRs), imaging systems, and real-time monitoring tools. This evolution is driven by the convergence of big data analytics and computational power, enabling AI to support clinical decision-making and operational efficiencies. For instance, AI-driven analytics in radiology have progressed from basic image enhancement to sophisticated pattern recognition, reducing diagnostic discrepancies and augmenting radiologist workflows [1-5].

However, this integration extends beyond isolated tools to encompass systemic embeddings, where AI serves as the backbone of healthcare analytics pipelines. Systems-level AI facilitates data aggregation from diverse sources, including wearables and EHRs, to generate actionable insights for population health management and personalized care [6, 7]. Yet, the infrastructural role of AI introduces complexities in data interoperability, where fragmented systems hinder seamless analytics, exacerbating challenges in resource-constrained settings [8, 9]. The literature highlights how AI can address human resource crises by automating administrative tasks and providing diagnostic support, potentially alleviating physician burnout and shortages [10, 11]. Despite these benefits, the rapid pace of AI adoption has outpaced governance structures, prompting calls for ethical recalibration to ensure equitable distribution of benefits [12, 13].

Ethical imperatives in AI-embedded systems

Ethical considerations are paramount in AI-embedded healthcare, as algorithms influence life-altering decisions. Core imperatives include fairness, transparency, and non-maleficence, drawing parallels to medical ethics but adapted to technological contexts [14, 15]. Bias in AI analytics, stemming from skewed training data, poses risks of perpetuating health disparities; for example, algorithms using cost proxies for health needs have demonstrated racial biases, under-identifying care requirements for Black patients [14]. Ethical frameworks emphasize the need for diverse datasets and bias audits to promote health equity [13, 16, 17].

Privacy and consent emerge as critical ethical pillars, particularly in digital phenotyping, where passive data collection from smartphones raises concerns over identifiability and misuse [18-20]. Actionable ethics demand stakeholder engagement to balance innovation with participant protections, including informed consent beyond mere terms of service [18]. In clinical analytics, ethical governance requires transparency in “black-box” models to enable clinician oversight and patient trust, mitigating risks of inscrutable predictions [3, 10]. Multisociety statements in radiology underscore the ethical duty to prioritize patient well-being and ensure that AI tools are reliable and fair [8, 9]. Overall, ethical imperatives advocate for AI as an augmentative force, preserving human qualities such as empathy while addressing systemic biases [19, 21].

Liability landscapes for AI in healthcare analytics

Liability in AI-embedded systems involves navigating accountability for harms arising from algorithmic errors or biases. Physicians using AI face potential liability under negligence standards, where reliance on faulty outputs could constitute malpractice if not critically evaluated [2]. The literature explores shared liability models that distribute responsibility among developers, institutions, and users, especially for adaptive AI that evolves post-deployment [5, 12]. In radiology, liability extends to data privacy breaches and diagnostic failures, with calls for insurance models adapted to AI risks [8, 22].

Challenges arise from AI’s opacity, complicating proof of causation in adverse events; for instance, automation bias may lead clinicians to overlook discrepancies, shifting liability dynamics [10, 16]. Regulatory gaps exacerbate this, as traditional tort doctrines inadequately address AI-induced injuries [19]. Synthesis across studies suggests proactive liability frameworks, including post-market surveillance and clear allocation protocols, to foster innovation without deterring adoption [3, 23-27]. In analytics contexts, liability for biased predictions underscores the need for developer accountability in data curation and model validation [14, 17].

Regulatory frameworks governing AI deployment

Regulatory governance for AI in healthcare has evolved to encompass software as a medical device (SaMD), with bodies such as the FDA and EU authorities adapting frameworks for ML-based tools [5, 6, 28, 29]. A system-view approach is advocated, evaluating AI within broader ecosystems including human factors and workflows, rather than in isolation [5]. This includes precertification programs for developers and ongoing monitoring to address risks associated with adaptive algorithms [12].

In Europe, the AI strategy emphasizes ethical alignment with digital health and integrates GDPR for data protection [6, 7]. US regulations focus on validation and transparency, with radiology-specific guidelines for AI imaging tools [8, 22, 27, 28]. Challenges include jurisdictional variances and the need for harmonized standards to prevent regulatory arbitrage [3, 4]. Literature calls for radiology-centric regulatory boards to vet AI, ensuring safety and efficacy across diverse settings [8, 9]. For analytics, regulations must mandate bias assessments and real-world performance evaluations to mitigate disparities [13, 14, 17].

Review positioning and synthesis logic

This review positions ethical, liability, and regulatory governance as foundational to AI-embedded healthcare systems and analytics, synthesizing literature through an original systems-level framing. We integrate cross-study insights across data ingestion, model deployment, clinical integration, and feedback loops, avoiding replication of existing taxonomies. The scope focuses on narrative synthesis of peer-reviewed works from 2017-2022, emphasizing infrastructural and analytical dimensions without empirical evaluations. This logic enables interpretive structuring, highlighting governance as a dynamic enabler of sustainable AI adoption.

Landscape of AI in healthcare systems and analytics

AI architectures in healthcare infrastructure

AI architectures in healthcare systems have diversified, incorporating ML for data-driven analytics within integrated infrastructures. Deep learning models, particularly convolutional neural networks, dominate imaging analytics, enabling automated detection in radiology and cardiology [7, 25, 28]. Systems-level architectures facilitate scalable data processing, from EHR integration to cloud-based analytics, addressing interoperability challenges [14, 21]. Literature synthesizes how AI infrastructures support predictive modeling, such as sepsis detection and risk stratification, by leveraging heterogeneous data sources [10, 11].

However, infrastructural embeddings require governance to address data quality variations, where poor provenance can undermine analytical reliability [3, 5]. Regulatory frameworks emphasize system-wide evaluations and incorporate human-AI interfaces to prevent workflow disruptions [5, 16]. Ethical considerations in architectures focus on modular designs that allow bias interrogation and transparency, ensuring equitable analytics across populations [13, 17].

Analytics pipelines and data ecosystems

AI analytics pipelines in healthcare synthesize data from ingestion to inference, forming ecosystems that enable real-time insights. Pipelines involve feature extraction, model training, and validation, with applications in phenotyping and population health [15, 20]. Synthesis reveals how pipelines mitigate biases through diverse training sets, as biases in ecosystems can amplify disparities [14, 17]. Governance demands robust data management, including provenance tracking and consent protocols, to protect ecosystems from privacy breaches [3, 18, 20].

Regulatory oversight for pipelines includes standards for data versioning and post-deployment monitoring, adapting to dynamic ecosystems [5, 12]. Liability in analytics ecosystems arises from pipeline failures, necessitating clear accountability chains [2, 19]. The literature integrates insights on pipeline resilience, advocating feedback mechanisms to recalibrate models to account for distributional shifts [10, 16].

Integration of AI in clinical workflows

Integrating AI into clinical workflows transforms healthcare systems by augmenting decision support and operational analytics. In radiology, AI integration reduces error rates by assisting in image interpretation, though perceptual and cognitive discrepancies persist [24-26]. Workflows benefit from AI prioritization, such as triaging urgent cases, enhancing efficiency in resource-limited settings [11, 16].

Ethical integration requires transparency to avoid automation complacency, where over-reliance erodes clinician vigilance [10, 16]. Liability frameworks must account for workflow-induced harms, with shared responsibility models [2, 8]. Regulatory governance promotes validated integrations, including human factors testing to ensure safe AI-clinical synergies [5, 22, 27].

Governance challenges in AI systems

Governance in AI healthcare systems encompasses ethical, liability, and regulatory dimensions to safeguard the integrity of analytics. Ethical governance prioritizes principles operationalized through actionable frameworks, beyond mere declarations [4, 18, 19]. Challenges include ensuring that AI systems promote justice and minimizing harm from biased analytics [13, 14, 17].

Liability governance addresses attribution in system failures, advocating for insurance and legal adaptations [2, 8, 12]. Regulatory challenges involve harmonizing SaMD frameworks and developing system-view regulations to encompass analytics ecosystems [5, 6, 28, 29]. Synthesis highlights the need for interdisciplinary governance that integrates stakeholder inputs to build resilient systems [9, 20, 21].

Emerging trends in AI analytics

Emerging trends in AI analytics focus on adaptive systems and closed-loop feedback, enhancing healthcare responsiveness. Trends include the use of reinforcement learning for personalized interventions and big data analytics to reduce disparities [10, 15, 16]. Governance must evolve with trends, incorporating real-time ethical audits [18, 19].

Liability in trending analytics requires dynamic models that trigger re-evaluations in response to adaptive changes [5, 12]. Regulatory trends emphasize precertification and international alignment, as seen in European strategies [6, 7]. The literature synthesizes trends through systems framing, positioning governance as key to harnessing the potential of analytics without ethical compromise [3, 4, 11].

Intelligent clinical decision and closed-loop healthcare systems

Architectures for intelligent clinical decision support

Intelligent clinical decision support systems (CDSS) leverage AI architectures to fuse data analytics with human expertise, forming the core of AI-embedded healthcare. Architectures typically include modular components for data preprocessing, predictive modeling, and recommendation engines, enabling real-time support in diagnostics and treatment planning [11, 15]. In radiology, CDSS architectures integrate deep learning for image analytics, reducing discrepancies while augmenting radiologist judgment [24-28].

Governance in these architectures demands transparency mechanisms, such as explainable AI, to clarify decision pathways and mitigate black-box risks [3, 10, 19]. Ethical architectures prioritize bias mitigation, ensuring decisions promote equity across demographics [13, 14, 17]. Liability considerations involve architecture designs that log decision traces for accountability, facilitating post-event analysis [2, 12]. Regulatory frameworks require architecture validations, including simulations of clinical scenarios to assess safety [5, 22, 29].

Closed-loop systems in healthcare analytics

Closed-loop healthcare systems represent advanced AI embeddings, where analytics cycles from monitoring to intervention and feedback, automating aspects of care delivery. These systems, akin to reinforcement learning loops, adjust interventions based on continuous data inputs, as in glucose control or ventilator management [10, 16]. In clinical contexts, closed loops enhance precision medicine by recalibrating models with real-world outcomes and addressing distributional shifts [10, 21].

Ethical governance for closed loops emphasizes human oversight to prevent autonomous errors and preserve patient autonomy [18-20]. Liability in closed-loop dynamics arises from feedback-induced harms, necessitating shared models between developers and clinicians [2, 5, 8]. Regulatory approaches advocate for adaptive oversight, with thresholds for human intervention in high-risk loops [6, 12, 27].

Human-AI decision fusion in clinical systems

Human-AI decision fusion integrates AI analytics with clinician expertise, optimizing outcomes through complementary strengths. Fusion dynamics involve AI providing probabilistic insights, which humans contextualize with tacit knowledge, thereby reducing errors such as complacency [10, 16, 24]. Literature synthesizes fusion as essential for high-performance systems, where AI handles data volume and humans manage complexity [11, 26].

A conceptual formula for human-AI decision fusion can formalize this: Let D represent the final decision, where D = f(H, A), with H as human input (judgment, context) and A as AI output (prediction, probability). Governance ensures fusion balances weights to avoid over-reliance and ethically safeguards against bias propagation [4, 13, 17]. Liability frameworks must delineate shared responsibilities, with regulations mandating training to ensure effective integration [5, 8, 22].

Feedback mechanisms and governance integration

Feedback mechanisms in intelligent systems enable iterative improvements, feeding outcomes back into analytics for model refinement. Mechanisms include post-market surveillance and user feedback loops, which are crucial for detecting biases in real time [9, 12, 21]. In closed-loop healthcare, feedback ensures system adaptability and aligns with ethical recalibration needs [18, 20].

Governance integration requires embedding regulatory checks into feedback, such as automated fairness audits [6, 13, 29]. Liability is managed through traceable feedback logs, aiding in harm attribution [2, 19]. Synthesis positions feedback as a governance enabler that fosters resilient clinical systems [3, 4, 5]. **Figure 1** illustrates the closed-loop governance architecture underpinning AI-embedded healthcare analytics systems.

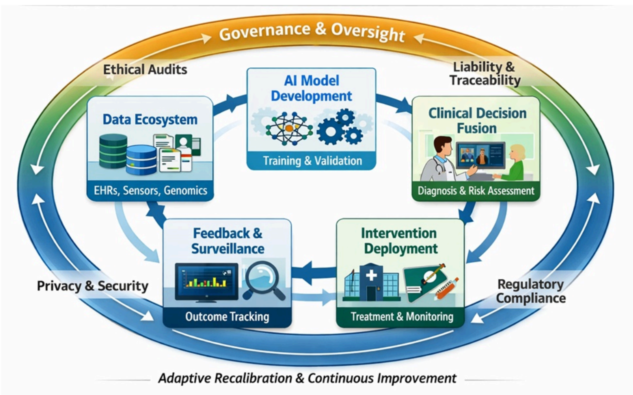

Figure 1. End-to-end governance architecture in AI-embedded healthcare analytics systems.
This figure depicts a cyclical systems architecture integrating data ecosystems, AI intelligence generation, clinical decision fusion, intervention deployment, and feedback surveillance within a continuous analytics loop. A governance oversight halo envelops the operational pipeline, embedding ethical audits, liability traceability, regulatory compliance monitoring, and privacy safeguards across all lifecycle stages. Bidirectional flows represent adaptive recalibration, highlighting governance as an infrastructural control layer ensuring safe, equitable, and accountable AI deployment in healthcare systems.

Results and Discussion

Challenges in ethical governance of AI-embedded healthcare systems

Ethical governance remains a persistent challenge in AI-embedded healthcare systems, particularly as analytics pipelines scale across diverse clinical environments. Bias amplification persists as a core issue, with training data imbalances leading to disparate outcomes in predictive analytics that disproportionately affect underrepresented groups [14, 17]. Cross-study synthesis reveals that while bias detection methods exist, their implementation lags in real-world systems due to resource constraints and lack of standardized protocols [10, 13]. Transparency deficits in model decision pathways exacerbate ethical concerns, as clinicians struggle to interrogate “black-box” outputs, potentially eroding trust and informed consent processes [3, 18, 20].

Privacy vulnerabilities intensify as data ecosystems expand, including digital phenotyping and multimodal inputs, raising the risk of re-identification despite anonymization efforts [17, 18]. Ethical governance must navigate tensions between data utility for analytics and individual protections, often complicated by jurisdictional differences in consent requirements [6, 19]. In closed-loop systems, ethical dilemmas arise from automation risks, where reduced human oversight may compromise patient autonomy or introduce unintended harms [10, 16, 19]. Multisociety efforts in domains like radiology highlight the need for proactive ethical vetting, yet broader adoption remains uneven [8, 9, 22].

Liability allocation and accountability gaps

Liability allocation in AI-embedded systems poses significant challenges, as traditional negligence frameworks inadequately address distributed responsibilities among developers, deployers, and users [2, 5, 12]. When algorithmic errors contribute to adverse events, causation tracing becomes complex due to opacity and adaptive learning, complicating malpractice claims [2, 10]. Shared liability models are proposed, yet implementation varies by jurisdiction, with developers often shielded while clinicians bear frontline accountability [3, 27]. In analytics-driven decision-making, liability extends to institutional procurement and monitoring failures, underscoring the need for traceable audit logs and insurance adaptations [8, 22].

Challenges compound in closed-loop architectures, where feedback mechanisms may alter system behavior post-deployment, blurring initial accountability lines [5, 12]. Regulatory gaps in post-market surveillance further hinder timely liability attribution, as evidenced by calls for adaptive frameworks [6, 29]. Synthesis indicates that without clear allocation protocols, liability fears may deter adoption, particularly in high-stakes analytics [16, 19].

Regulatory fragmentation and harmonization barriers

Regulatory fragmentation represents a major hurdle, with varying approaches across regions impeding seamless AI integration in global healthcare systems [5, 6, 29]. While system-view regulations for SaMD emphasize lifecycle oversight, inconsistencies in classification and validation requirements create compliance burdens [5, 27]. In radiology and imaging analytics, multisociety guidelines advocate context-specific vetting, yet harmonization with broader frameworks remains incomplete [8, 9, 22, 28]. **Table 1** shows the summarized governance risk domains and corresponding oversight mechanisms across AI-embedded healthcare systems.

Table 1. Governance risk domains and oversight mechanisms in AI-embedded healthcare systems

Governance domain	System risk vector	Healthcare manifestation	Oversigh mechanisn
Algorithmic bias	Skewed training datasets	Disparate diagnostic accuracy	Bias audits Dataset diversificati
Model opacity	Black-box inference pathways	Limited clinician interpretability	Explainable modules • Transparen logs
Privacy intrusion	Digital phenotyping and passive data	Re-identification risks	Consent layers • Encryption Federated learning
Liability ambiguity	Distributed decision causality	Malpractice attribution gaps	Audit trails Decision tra logs
Automation complacency	Over-reliance on AI outputs	Reduced clinical vigilance	Human-in-th loop mandates
Regulatory fragmentation	Jurisdictional policy variance	Deployment approval delays	Harmonizati boards • Pr certificatio
Distributional drift	Real-world data shifts	Performance degradation	Post-marke surveillanc
Workflow disruption	Poor system integration	Cognitive overload	Human factors validation

Challenges include adapting regulations to adaptive algorithms, where pre-market approvals may not capture post-deployment shifts [6, 12]. Cross-border data flows in analytics ecosystems complicate compliance with privacy regulations such as the GDPR [6, 20]. Synthesis across literature points to the need for international alignment to prevent regulatory arbitrage while fostering innovation [3-5].

Challenges

Persistent technical and systemic barriers

Despite rapid advances in AI architectures, persistent technical constraints continue to undermine the reliability and scalability of AI-embedded healthcare systems. One of the most consequential barriers is the distributional shift, whereby models trained on curated or institution-specific datasets degrade in performance when exposed to heterogeneous real-world populations [10, 16]. Clinical environments are inherently dynamic: demographic patterns evolve, disease prevalence fluctuates, diagnostic technologies update, and treatment standards change. These shifts introduce covariate drift, label shift, and concept drift, each of which can compromise predictive accuracy and risk calibration. In closed-loop systems, such degradation may propagate silently across iterative cycles, amplifying downstream errors.

Data quality and provenance inconsistencies further compound this problem. Real-world healthcare data are often fragmented, incomplete, and subject to variable documentation practices, limiting the generalizability of predictive analytics [14, 21]. Interoperability deficits between electronic health record (EHR) systems, imaging repositories, and wearable platforms impede seamless data ingestion and feedback integration, weakening the structural coherence of analytics ecosystems. Without standardized ontologies and harmonized metadata architectures, cross-institutional validation remains constrained.

Resource disparities intensify these technical limitations. Large academic medical centers may possess dedicated AI governance units and monitoring infrastructure. In contrast, smaller or rural institutions often lack the computational capacity, workforce expertise, and financial capital necessary for rigorous validation and lifecycle surveillance [16, 17]. This asymmetry risks creating a stratified ecosystem in which governance quality correlates with institutional wealth, thereby reinforcing systemic inequities in AI safety and performance.

Human factors and workflow integration

Human-AI interaction introduces a second category of systemic vulnerability. Automation complacency—where clinicians over-trust algorithmic outputs—can diminish vigilance and reduce critical oversight, particularly in high-throughput environments such as radiology or emergency medicine [10, 16, 24]. Over time, reliance on AI-generated recommendations may erode diagnostic reasoning skills, contributing to skill atrophy and reduced resilience during system failures.

Workflow misalignment represents another persistent barrier. When AI systems are integrated without careful human factors engineering, they may increase cognitive load rather than alleviate it [11, 15]. Poor interface design, unclear confidence metrics, and fragmented alert systems can overwhelm clinicians with redundant or low-specificity prompts. Instead of streamlining care delivery, such disruptions introduce friction into already complex clinical workflows.

Training deficits further inhibit effective human-AI fusion. Many clinicians lack formal training in interpreting probabilistic outputs, calibration curves, or model-uncertainty parameters [22, 26]. Without structured competency development, clinicians may misinterpret predictions or struggle to contextualize AI recommendations within broader clinical narratives. Effective integration, therefore, requires not only technical refinement but also institutional investment in AI literacy and interdisciplinary training.

Equity and access disparities

Equity challenges remain central to governance discourse. Biased analytics, often rooted in historically skewed training datasets, risk perpetuating or amplifying existing health disparities [13, 14, 17]. Algorithms trained predominantly on majority populations may underperform in underrepresented demographic groups, resulting in inequitable risk stratification, delayed diagnoses, or inappropriate treatment recommendations. Such disparities are particularly concerning in predictive analytics systems that influence resource allocation decisions.

Infrastructure inequities exacerbate these concerns. Low-resource settings frequently lack robust digital infrastructure, high-quality data capture systems, and continuous monitoring capabilities [16, 21]. As AI deployment accelerates in well-funded institutions, a widening digital divide may emerge, restricting the benefits of intelligent clinical decision systems to already advantaged populations. Without intentional equity-focused governance, AI integration risks reinforcing systemic stratification rather than mitigating it.

Evolving regulatory and liability pressures

Regulatory and liability landscapes are undergoing parallel transformations. Emerging generative and agentic AI systems challenge traditional regulatory paradigms that assume deterministic outputs and static model behavior [5, 6]. Non-deterministic architectures, capable of adaptive learning and autonomous reasoning, require oversight frameworks that account for probabilistic variability and continuous evolution.

Existing software-as-a-medical-device (SaMD) pathways emphasize pre-market validation, yet adaptive systems demand lifecycle governance that extends beyond initial approval [5, 29]. Jurisdictional variability further complicates compliance, particularly in cross-border data ecosystems subject to divergent regulatory standards [6].

Liability pressures similarly intensify in dynamic closed-loop systems. When harms occur within complex human-AI fusion processes, attribution becomes difficult due to shared decision pathways and evolving model parameters [2, 12, 19]. Traditional malpractice doctrines, designed for human-only clinical judgments, may inadequately capture distributed accountability in AI-mediated environments. Without clarified allocation models, liability uncertainty may deter innovation or shift disproportionate responsibility onto frontline clinicians.

Future research directions

Advancing bias mitigation and fairness metrics

Future research must transition from static fairness assessments to longitudinal bias monitoring embedded within closed-loop systems [13, 14, 17]. Adaptive fairness metrics that recalibrate in response to distributional shifts are essential to prevent drift-induced disparities. Rather than relying solely on pre-deployment audits, governance frameworks should incorporate continuous bias surveillance mechanisms integrated into feedback pipelines.

Federated learning architectures present promising avenues for privacy-preserving equity improvements, enabling collaborative model training across institutions without centralized data pooling [18, 20]. Research should evaluate whether federated approaches reduce representational bias while maintaining performance integrity across heterogeneous populations.

Enhancing explainability in clinical contexts

Explainability research must move beyond generic post-hoc interpretation tools toward context-specific transparency mechanisms aligned with clinical reasoning processes [3, 10, 19]. Clinicians require explanations that map to diagnostic workflows, highlight uncertainty ranges, and clarify actionable implications rather than abstract feature importance rankings.

Empirical studies assessing how explainable outputs influence trust calibration, decision accuracy, and workflow efficiency are critically needed [11, 16]. Research should also examine how explanation granularity affects cognitive load and whether adaptive explanation systems—tailored to user expertise—improve human-AI fusion outcomes.

Developing adaptive regulatory models

Regulatory innovation must prioritize lifecycle governance models incorporating real-time performance monitoring, dynamic risk thresholds, and adaptive oversight triggers [5, 6, 29]. Instead of episodic review processes, future regulatory architectures should embed surveillance infrastructures within operational systems.

Regulatory sandbox environments offer structured testing spaces where emerging AI systems can be evaluated under supervised conditions before large-scale deployment [27, 28]. Comparative research on sandbox effectiveness across jurisdictions may inform harmonized international standards, mitigating fragmentation while encouraging innovation.

Exploring liability and insurance innovations

Liability reform research should model shared accountability regimes that distribute responsibility across developers, institutions, and clinicians based on system design and operational roles [2, 8, 12]. Simulation-based analyses of adverse-event attribution within closed-loop systems could clarify causal pathways and inform policy design.

AI-specific insurance products represent another area for exploration, potentially buffering innovation risks while incentivizing compliance with governance best practices [3, 19]. Longitudinal studies evaluating how liability frameworks influence adoption rates, safety outcomes, and transparency in reporting would provide empirical grounding for legal reform.

Strengthening feedback and recalibration mechanisms

Future investigations should focus on automated feedback infrastructures capable of detecting performance drift, bias emergence, and workflow disruption in near real time [9, 12, 21]. Governance overlays that integrate ethical audits directly into recalibration pipelines may enhance resilience.

Multi-stakeholder governance models—incorporating clinicians, patients, developers, and regulators—should be empirically evaluated to determine optimal configurations for oversight design [4, 18]. Research into scalable recalibration protocols, including threshold-based human override triggers, may strengthen safety in high-risk closed-loop systems.

Conclusion

The embedding of artificial intelligence in healthcare systems and analytics offers transformative potential for clinical decision support, predictive modeling, and closed-loop interventions. Yet, its realization hinges on robust ethical, liability, and regulatory governance. This narrative review synthesizes key literature from 2017–2022, framing governance across data ecosystems, model deployment, human-AI fusion, and feedback mechanisms. Ethical imperatives demand proactive bias mitigation, transparency, and privacy safeguards to uphold equity and non-maleficence. Liability challenges necessitate clear allocation models to foster accountability without stifling innovation. Regulatory evolution toward system-view and adaptive approaches is critical for safe integration.

The proposed end-to-end analytics loop, with governance overlay, underscores the interconnectedness of technical and oversight elements. Persistent challenges—bias, opacity, fragmentation—highlight the urgency of interdisciplinary action. Future directions in fairness, explainability, and harmonized models offer pathways forward. Ultimately, effective governance will enable AI to augment human expertise, advancing high-performance, equitable healthcare while safeguarding patient well-being and societal trust.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 16 Feb 2023 Revised: 30 Mar 2023 Accepted: 07 May 2023
Published online: 20 July 2023

References

Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med. 2018;378(11):981-3. https://doi.org/10.1056/NEJMp1714229.

Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA. 2019;322(18):1765-6. https://doi.org/10.1001/jama.2019.15064.

Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018;15(11):e1002689. https://doi.org/10.1371/journal.pmed.1002689.

Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7. https://doi.org/10.1038/s42256-019-0114-4.

Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. NPJ Digit Med. 2020;3:53. https://doi.org/10.1038/s41746-020-0262-2.

Cohen IG, Evgeniou T, Gerke S, Minssen T. The European artificial intelligence strategy: implications and challenges for digital health. Lancet Digit Health. 2020;2(7):e376-e379. https://doi.org/10.1016/S2589-7500(20)30112-6.

Fenech ME, Buston O. AI in Cardiac Imaging: A UK-Based Perspective on Addressing Governance, Ethics, and Data Protection. J Med Artif Intell. 2019;2:15. https://doi.org/10.21037/jmai.2019.08.02.

Geis JR, Brady A, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Radiology. 2019;293(2):436-40. https://doi.org/10.1148/radiol.2019191586.

Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of using and sharing clinical imaging datasets for artificial intelligence: a perspective from the RSNA ethics of AI in radiology committee. Radiol Artif Intell. 2020;2(2):e190269. https://doi.org/10.1148/ryai.2020190269.

Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7. https://doi.org/10.1136/bmjqs-2018-008370.

Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7.

Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Das R, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. https://doi.org/10.1038/s41591-019-0548-6.

Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. https://doi.org/10.7326/M18-1990.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. https://doi.org/10.1126/science.aax2342.

Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med. 2019;112(1):22-8. https://doi.org/10.1177/0141076818816910.

Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res. 2018;18(1):545. https://doi.org/10.1186/s12913-018-3359-4.

Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics. 2019;21(2):E167-79. https://doi.org/10.1001/amajethics.2019.167.

Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med. 2019;17(1):137. https://doi.org/10.1186/s12916-019-1377-7.

Rigby MJ. Ethical dimensions of using artificial intelligence in health care. AMA J Ethics. 2019;21(2):E121-4. https://doi.org/10.1001/amajethics.2019.121.

Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med. 2018;1:68. https://doi.org/10.1038/s41746-018-0075-8.

Panch T, Mattie H, Celi LA. The “inconvenient truth” about AI in healthcare. NPJ Digit Med. 2019;2:77. https://doi.org/10.1038/s41746-019-0155-4.

Jaremko JL, Azar M, Bromwich R, Lum W, Cheong L, Gibert M, et al. Canadian Association of Radiologists white paper on ethical and legal issues related to artificial intelligence in radiology. Can Assoc Radiol J. 2019;70(2):107-18. https://doi.org/10.1016/j.carj.2019.03.003.

Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 2017;8(1):171-82. https://doi.org/10.1007/s13244-016-0534-1.

Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018;2(1):35. https://doi.org/10.1186/s41747-018-0061-6.

Rajpurkar P, Chen PJ, Ng AY, Lungren MP. Deep learning in radiology: now the real work begins. J Biomed Inform. 2019;89:1-2. https://doi.org/10.1016/j.jbi.2018.11.009.

Langlotz CP. Will artificial intelligence replace radiologists? Radiol Artif Intell. 2019;1(3):e190058. https://doi.org/10.1148/ryai.2019190058.

Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging. 2018;9(5):745-53. https://doi.org/10.1007/s13244-018-0649-y.

Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol.

2018;15(3 Pt B):504-8. https://doi.org/10.1016/j.jacr.2017.12.026.

Kohli A, Mahajan V, Seals JK, Klein A. Concepts in US Food and Drug Administration Oversight of Artificial Intelligence for Medical Imaging. AJR Am J Roentgenol. 2021;216(4):885-7. https://doi.org/10.2214/AJR.20.24678.