REVIEW Open access

Artificial Intelligence in Healthcare Systems (2017–2025): From Predictive Analytics to Autonomous Governance Architectures

Mustafa Demir[1*], Ayse Yilmaz[1], Mehmet Kaya[2], Selin Arslan[1]

Abstract

The integration of artificial intelligence (AI) into healthcare systems marks a fundamental shift from isolated predictive analytics tools to embedded, scalable architectures that support autonomous governance. This narrative review synthesizes 28 peer-reviewed publications from leading journals to examine AI's role across healthcare infrastructure and clinical analytics. Early work established deep learning foundations for risk prediction, diagnostic support, and prognostic modelling using multimodal data. These capabilities rapidly evolved into system-level applications that enhance data ingestion, real-time inference, and operational optimisation across entire care ecosystems.

By the early 2020s, attention turned to deployment realities, including clinician acceptance, cost-effectiveness, and integration into existing workflows. Frameworks for responsible implementation emerged alongside regulatory perspectives that emphasise safety, equity, and continuous oversight. Recent contributions highlight the transition toward closed-loop systems in which predictive outputs inform decisions, trigger interventions, and feed outcome data back for model recalibration. Governance architectures now address ethical challenges, explainability gaps, and the move from generalist to specialised medical AI.

This review organises the literature through an original systems-level lens spanning four interconnected pillars—data foundations, analytic intelligence, deployment mechanisms, and governance layers—rather than replicating prior application-specific taxonomies. Cross-study analysis reveals consistent patterns: predictive analytics serve as the foundational engine, clinical decision support acts as the execution layer, closed-loop feedback enables adaptation, and governance ensures sustainable autonomy. The synthesis demonstrates that AI is no longer an adjunct technology but a core infrastructural element reshaping how healthcare systems ingest, process, act upon, and learn from data at scale.

Trajectory as a coherent progression toward autonomous yet human-centred governance, the review provides clinicians, system architects, and policymakers with a unified understanding of current capabilities and the infrastructural requirements for responsible scaling.

Keywords Clinical decision support, Artificial intelligence, Healthcare systems, Predictive analytics, Closed-loop systems, Autonomous governance

*Correspondence:
Mustafa Demir
mustafa.demir@outlook.com
[1] Department of Healthcare Informatics, Faculty of Medicine, Hacettepe University, Ankara, Turkey
[2] Department of AI in Clinical Systems, Faculty of Engineering, Middle East Technical University, Ankara, Turkey

Introduction

https://doi.org//y495416102

Introduction

Healthcare systems under pressure: the case for AI integration

Modern healthcare systems are operating under sustained and intensifying pressure. Demographic transitions have led to rapidly ageing populations, increasing the prevalence of multimorbidity and long-term care needs. Simultaneously, the global burden of chronic diseases—including cardiovascular conditions, diabetes, cancer, and neurodegenerative disorders—has escalated in both high-income and resource-constrained settings. These trends are compounded by persistent workforce shortages, uneven geographic distribution of clinicians, administrative inefficiencies, and rising expenditure trajectories, all of which challenge fiscal sustainability. Together, these forces have exposed structural limitations in traditional models of care delivery and data management.

Historically, healthcare analytics has been dominated by rule-based systems, retrospective audits, and static reporting dashboards. These approaches, while valuable for quality assurance and compliance, are inherently reactive. They rely on predefined thresholds and limited-variable frameworks that cannot adequately accommodate the complexity, velocity, and heterogeneity of modern healthcare data. Contemporary care environments generate continuous streams of multimodal data—electronic health records (EHRs), diagnostic imaging, laboratory results, genomic data, wearable sensor outputs, claims data, and administrative records. The dimensionality and interdependence of these datasets exceed the analytic capacity of conventional statistical paradigms.

Artificial intelligence (AI) has emerged as a transformative response to this mismatch between system complexity and analytic capability. By leveraging machine learning, deep learning, and reinforcement learning architectures, AI systems can process high-dimensional inputs, identify latent patterns, and generate probabilistic forecasts in real time. Crucially, AI enables the transformation of raw data into actionable intelligence at scale, supporting clinical decision-making, operational optimisation, and strategic planning [1-5]. Between 2017 and 2025, peer-reviewed literature increasingly documented this transition, framing AI not merely as a collection of predictive models but as a foundational infrastructure layer within healthcare systems.

High-impact syntheses emphasised that AI's value extends beyond individual-level risk prediction or diagnostic augmentation. When integrated effectively, AI contributes to system-wide optimisation: dynamic resource allocation, predictive staffing models, early detection of system strain, improved care coordination across fragmented services, and outcome improvement across populations [6, 7]. For example, predictive analytics can identify patients at risk of deterioration, enabling pre-emptive intervention and reducing avoidable hospitalisations. Operational algorithms can anticipate bed occupancy or supply shortages, mitigating bottlenecks before they escalate into crises. At the population level, AI supports risk stratification and targeted prevention strategies, aligning clinical and public health priorities.

However, early enthusiasm for AI's transformative potential was tempered by recognition that technical performance metrics—such as area under the curve (AUC) or accuracy—do not automatically translate into clinical or operational impact. Implementation failures often stemmed from misalignment with existing workflows, inadequate interoperability between data systems, and insufficient consideration of ethical, regulatory, and human factors [8, 9]. Models trained on retrospective datasets frequently encountered degraded performance in real-world settings due to dataset shift, bias, or incomplete data capture. Furthermore, clinician trust emerged as a decisive determinant of adoption; opaque "black-box" outputs, when not accompanied by interpretability mechanisms or clear accountability structures, limited integration into routine practice.

Thus, the contemporary case for AI integration rests not only on computational capability but on systemic alignment. Successful deployment requires embedding AI within clinical pathways, ensuring compliance with regulatory frameworks, safeguarding patient privacy, and instituting governance mechanisms that balance innovation with accountability. The period under review demonstrates a progressive shift from viewing AI as a tool for isolated prediction toward conceptualising it as a structural component of healthcare infrastructure.

Chronological evolution of AI capabilities

The evolution of AI in healthcare reflects a trajectory from technical feasibility to infrastructural integration and governance maturity.

The period opened with foundational demonstrations of deep learning applied to health informatics and medical imaging. Early studies validated convolutional neural networks for radiological image classification, natural language processing pipelines for clinical note extraction, and multimodal architectures capable of integrating structured and unstructured data sources [10-12]. These foundational efforts established proof of concept that AI could match or exceed human-level performance in narrowly defined tasks under controlled conditions.

By 2019, the discourse matured toward system-level implications. Comprehensive overviews articulated the convergence of human and artificial intelligence in what was termed “high-performance medicine,” envisioning hybrid systems in which AI augments clinician cognition rather than replacing it [13-15]. During this phase, emphasis expanded to predictive analytics for disease detection, personalised treatment recommendations, and population health management. The narrative shifted from isolated algorithmic benchmarks to broader clinical integration and value creation.

The years 2020–2022 marked an accelerated period of translation into real-world systems. Global health crises, most notably the COVID-19 pandemic, catalysed the rapid deployment of AI-enabled surveillance, triage, and predictive modelling tools. Concurrently, the maturation of cloud-based infrastructure, federated learning frameworks, and scalable data platforms enabled cross-institutional collaboration without centralising sensitive data. Scholarly attention is increasingly focused on deployment challenges rather than model architecture alone. Key themes included data interoperability across heterogeneous EHR systems, bias detection and mitigation strategies, fairness auditing, and mechanisms for cultivating clinician trust [16-19].

Importantly, systematic examinations of AI in low- and middle-income countries highlighted both the promise and the constraints of implementation in resource-limited contexts [3]. Where digital infrastructure and data governance frameworks were strengthened, AI demonstrated potential for system-level improvements in triage, diagnostics, and supply chain management. However, without foundational investments in connectivity, workforce training, and regulatory oversight, algorithmic interventions risked exacerbating inequities.

From 2023 onward, the literature increasingly coalesced around governance, accountability, and the controlled expansion of autonomy. Randomised controlled evaluations of AI-enabled decision support systems began to appear, offering higher-quality evidence regarding clinical and operational impact [10, 13]. Regulatory perspectives matured, with agencies and professional bodies articulating pathways for approval, post-market surveillance, and lifecycle monitoring of adaptive algorithms [14]. Consensus reports outlined implementation frameworks that integrate technical validation with organisational readiness assessment and ethical oversight [6].

Simultaneously, advances in generative AI and generalist-to-specialist adaptation architectures expanded the frontier of capability [7, 9]. Large foundation models demonstrated the capacity to process multimodal inputs—text, images, structured data—and to adapt across clinical domains with limited fine-tuning. These developments introduced the possibility of semi-autonomous systems capable of orchestrating complex decision pathways under defined guardrails. The conversation consequently shifted from “Can AI predict?” to “Under what conditions can AI act?”

This chronological progression—from predictive foundations to deployment realism and ultimately governance-enabled autonomy—forms the empirical backbone of the present synthesis. It reflects a field transitioning from experimental innovation to infrastructural integration.

Defining AI-enabled healthcare systems and analytics

To clarify the scope of this review, precise definitions are required. “Healthcare systems” refers to the interconnected infrastructure of digital platforms, clinical workflows, operational processes, financial mechanisms, and governance structures that collectively deliver care across settings. This includes hospitals, primary care networks, community health services, public health agencies, and regulatory bodies, all linked through data exchange and organisational coordination.

“Analytics” encompasses the full continuum of data-driven processes:

- Descriptive analytics, which summarises historical patterns and performance indicators.
- Diagnostic analytics, which identifies causal relationships and explanatory factors.
- Predictive analytics, which estimates future events or risk states.
- Prescriptive analytics, which recommends optimal actions based on forecasted scenarios.
- Autonomous or closed-loop systems, which execute decisions and iteratively learn from outcomes.

Within this framework, AI is conceptualised not as standalone software but as an infrastructural layer embedded within the healthcare ecosystem. It ingests heterogeneous data, generates probabilistic or generative intelligence, supports or executes decisions, monitors downstream outcomes, and feeds experiential learning back into the system [1, 20-25]. This cyclical architecture—data ingestion, inference, action, feedback, and adaptation—distinguishes AI-enabled systems from static analytic tools.

This definition intentionally extends beyond narrow clinical applications such as image classification or risk scoring. Instead, it foregrounds system-level properties:

- Scalability – the ability to function across institutions and populations without linear increases in cost.
- Interoperability – seamless integration across disparate data platforms and vendor ecosystems.
- Continuous adaptation – lifecycle monitoring and updating to mitigate performance drift.
- Accountable autonomy – clear delineation of responsibility when AI systems influence or execute decisions.

The reviewed literature consistently demonstrates that isolated predictive tools, when detached from workflow integration and governance structures, achieve limited and often transient impact [5, 23]. In contrast, AI-enabled healthcare systems embed analytics within operational pathways, enabling coordinated optimisation across clinical, administrative, and strategic domains.

Accordingly, this review situates AI not as an adjunct innovation but as an emergent structural layer of healthcare infrastructure. Understanding its evolution, capabilities, and governance requirements is therefore essential for evaluating both its transformative potential and its systemic risks.

Scope and synthesis approach of this review

This narrative review synthesises peer-reviewed publications from high-impact venues. Unlike prior reviews that catalogue clinical applications or taxonomise algorithms, the present work advances an original integrative logic structured around four pillars—data, models, deployment, and governance—and their dynamic interactions. The synthesis logic traces a coherent trajectory: predictive analytics as the foundational engine, clinical decision support as the translational layer, closed-loop architectures as the adaptive mechanism, and autonomous governance as the overarching safeguard.

By cross-analysing perspectives from international health leaders, regulatory bodies, implementation scientists, and clinical researchers, the review identifies recurring infrastructural requirements and emergent design principles. The positioning statement is therefore explicit: AI in healthcare systems has evolved from supplementary predictive capability to core infrastructural architecture capable of autonomous governance, provided that data foundations, model robustness, deployment fidelity, and governance mechanisms are deliberately engineered as an integrated whole. This review maps that evolution and delineates the systems-level requirements for its responsible realisation.

Landscape of AI in healthcare systems & analytics

Data foundations and multimodal analytics pipelines

A robust data infrastructure is the bedrock of AI-enabled healthcare systems. Literature from 2017 onward repeatedly emphasised the necessity of ingesting, harmonising, and governing heterogeneous data streams—structured EHR entries, unstructured clinical notes, high-resolution imaging, physiological waveforms, and patient-generated data [21, 22, 25]. Early deep-learning reviews outlined pipelines capable of handling such multimodality while preserving privacy and provenance [26-28]. Subsequent work documented the real-world deployment of these pipelines at enterprise scale, highlighting challenges in temporal alignment, handling missingness, and federated architectures [1, 27].

Cross-study synthesis reveals a consistent maturation: from siloed dataset training to continuous, system-wide data lakes that support both retrospective analytics and prospective inference. International scoping reviews further illustrate that equitable system strengthening requires attention to data infrastructure disparities across resource settings [3, 17].

Analytic models: from predictive foundations to adaptive intelligence

Predictive analytics dominated the initial wave, with models demonstrating utility in risk stratification, early warning, and treatment response forecasting [24, 27, 28]. The reviewed corpus documents progressive refinement: incorporation of temporal dynamics, uncertainty quantification, and integration of domain knowledge [1, 21]. By the mid-period, attention shifted toward hybrid models that combine predictive outputs with prescriptive recommendations suitable for clinical workflows [22].

Recent contributions extend this lineage to generalist foundations that can be specialised for medical tasks and to generative approaches that augment rather than replace human reasoning [7, 9]. Scoping reviews of randomised trials confirm that model performance in controlled settings does not automatically translate to system-level benefit without accompanying workflow redesign [10].

Deployment and integration into operational healthcare infrastructure

Deployment literature stresses that technical excellence is necessary but insufficient. Frameworks for appropriate implementation underscore the requirement for prospective monitoring, human oversight protocols, and iterative recalibration once models enter live environments [4, 6]. Cost-effectiveness and budget-impact analyses further indicate that sustainable deployment demands clear return-on-investment pathways at the health-system level [5].

Synthesis across studies reveals recurring architectural patterns: edge-cloud hybrids for latency-sensitive applications, API-mediated integration with legacy EHRs, and dashboard interfaces that embed AI outputs within existing clinician cognitive workflows [11, 22, 25]. Acceptance studies highlight the centrality of explainability, trust calibration, and workflow congruence [4, 12].

System-wide analytics for optimisation and resource stewardship

Beyond clinical prediction, AI analytics increasingly address operational dimensions—such as bed management, supply-chain forecasting, workforce scheduling, and population-health risk stratification [22, 23]. The reviewed works illustrate how predictive and prescriptive analytics converge to create self-optimising systems that balance clinical quality with economic sustainability [5, 13]. Global-health perspectives reinforce that such system-level optimisation is especially impactful in resource-constrained environments when appropriately governed [3, 17].

Collectively, the landscape synthesis demonstrates that AI has matured into an infrastructural substrate capable of orchestrating data, models, and operations into coherent, learning healthcare systems. The four-pillar framing—data, models, deployment, governance—emerges as a unifying organisational logic that transcends individual applications and reveals the systemic requirements for progression toward autonomous capability.

Intelligent clinical decision & closed-loop healthcare systems

Intelligent clinical decision support has evolved from rule-based alerts to AI-driven recommendation engines that incorporate patient-specific context, uncertainty estimates, and longitudinal trajectories [1, 11, 26]. The reviewed literature documents architectures in which predictive models surface differential diagnoses, treatment options, or risk trajectories directly within electronic workflows, thereby augmenting rather than replacing clinician cognition [22, 27, 28]. Cognitive-perspective analyses caution that over-reliance risks de-skilling, while under-integration wastes potential; optimal designs therefore embed AI as a collaborative partner [11].

Closed-loop systems represent the next architectural layer, in which AI not only informs but participates in a continuous cycle of observation, decision, action, and evaluation. Randomised evaluations illustrate early examples of AI-triggered interventions followed by automated outcome capture and model updating [6, 10]. Governance frameworks explicitly require human authorisation gates for higher-risk loops while permitting fully autonomous operation in low-risk, well-characterised domains [2, 13, 14].

The transition toward autonomous governance architectures builds upon these closed loops by adding policy-enforcement layers that monitor compliance with regulatory, ethical, and equity standards in real time [7, 9, 15]. Recent perspectives on generalist-to-specialist adaptation and generative augmentation further enable systems that can propose, simulate, and refine care plans with minimal latency [7, 9]. The structural integration of these components into a continuous learning architecture is illustrated in **Figure 1**.

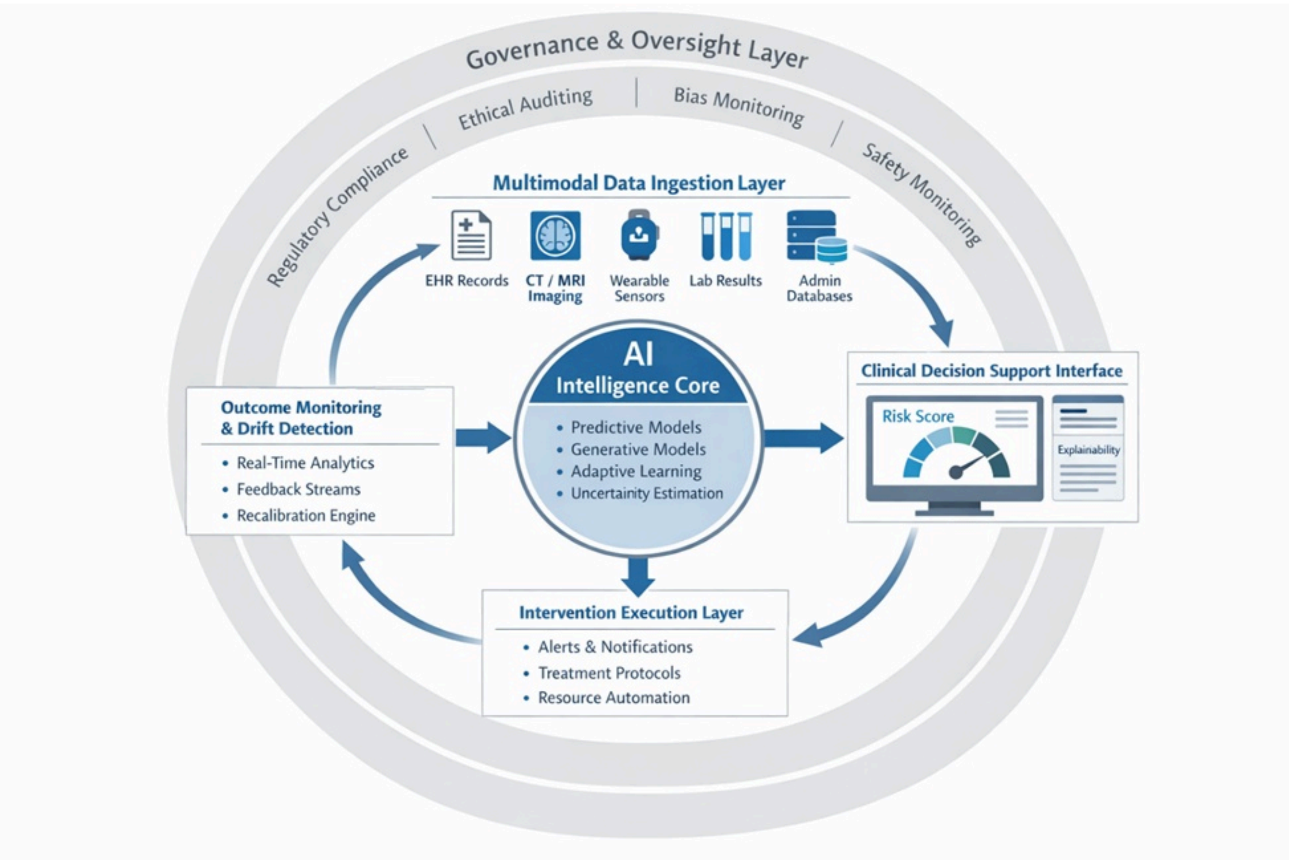

Figure 1. End-to-end closed-loop AI architecture for autonomous governance in healthcare systems

The diagram depicts a clockwise learning cycle beginning with the ingestion of multimodal data (EHR, imaging, physiological streams, and administrative data), which feeds into an AI intelligence core comprising predictive, generative, and adaptive models [1, 28]. Outputs are operationalised through clinical decision support interfaces that embed uncertainty-aware recommendations within workflows [11, 26]. Downstream intervention layers execute alerts, treatment pathways, or operational adjustments [10]. Real-time outcome monitoring enables drift detection and recalibration, closing the learning loop [6]. A concentric governance layer envelops all components, enforcing regulatory compliance, ethical auditing, bias surveillance, and human authorisation gates [2, 9, 13, 14]. The figure illustrates the transition from isolated predictive analytics to fully integrated, governance-enabled autonomous healthcare systems.

These architectures shift the paradigm from episodic decision support to continuous system intelligence. When data, models, deployment, and governance are deliberately co-designed, closed-loop systems achieve the infrastructural conditions necessary for safe progression toward autonomous operation. The reviewed evidence base, while still emergent, consistently supports this systems-level integration as the prerequisite for scalable, equitable, and high-performance healthcare delivery.

Results and Discussion

The synthesis of literature reveals a clear maturational arc in AI for healthcare systems and analytics. Predictive analytics, initially positioned as standalone capabilities for risk stratification and diagnostic augmentation [24, 27, 28], progressively became embedded within broader infrastructural layers [1, 22]. This embedding enabled the emergence of closed-loop architectures, where predictive outputs inform decisions, interventions follow, outcomes feed back for recalibration, and governance oversees the cycle [6, 10, 11, 26].

Cross-study analysis indicates that this progression is not merely additive but multiplicative: robust data foundations enable more reliable models [21, 25], which in turn support safer deployment [4, 12], while governance mechanisms ensure that closed loops remain accountable and adaptive [2, 9, 13, 14]. The four-pillar framing—data, models, deployment, governance—provides an integrative lens absent from prior application-centric reviews. It highlights recurring infrastructural imperatives: continuous data harmonisation, model specialisation from generalist bases [9], workflow-aligned interfaces [11], and policy-enforced oversight [15].

Autonomous governance architectures represent the current frontier, characterised by real-time compliance monitoring, ethical auditing, and bounded autonomy [7, 13, 14]. These architectures shift the emphasis from episodic human validation to systemic safeguards that enable autonomous operation in low-risk domains while retaining human oversight for high-stakes actions. The literature consistently grounds this shift in empirical requirements—drift detection, performance monitoring, and equitable recalibration—rather than speculative futurism [6, 10]. **Table 1** summarises the structural functions, technical requirements, risks, and governance dependencies across the four pillars.

Table 1. Four-pillar systems framework for AI-enabled healthcare infrastructure

Pillar	Primary function	Technical components	System-level risks	Governance dependencies	Representative citations
1. Data foundations	Multimodal ingestion and harmonisation	EHR integration, imaging pipelines, federated learning, temporal alignment engines	Bias propagation, missingness, interoperability gaps	Data standards, privacy regulation, and equity auditing	[1, 3, 21, 22, 25]
2. Analytic intelligence	Predictive, prescriptive, and generative modelling	Deep learning, uncertainty estimation, temporal models, generalist-	Performance drift, hallucination, opacity	Model validation standards, explainability requirements	[7, 9, 24, 27, 28]

		to-specialist adaptation			
3. Deployment mechanisms	Workflow integration and operationalisation	API integration, clinician dashboards, edge-cloud hybrids, and monitoring systems	Workflow misalignment, cognitive overload, cost burden	Human oversight protocols, ROI evaluation, prospective monitoring	[4-6, 11, 12]
4. Governance layer	Ethical, regulatory, and autonomous oversight	Drift detection, compliance engines, bias monitoring, and authorisation gates	Liability ambiguity, regulatory lag, and inequity amplification	Adaptive regulation, audit frameworks, and bounded autonomy models	[1, 2, 7, 13-15, 17]

Importantly, the reviewed works underscore that AI-enabled systems achieve their greatest value when treated as infrastructure rather than applications. Isolated predictive tools yield marginal gains; integrated architectures that orchestrate data-to-intervention cycles deliver systemic optimisation [5, 22, 23]. This systems-level perspective aligns with international consensus on responsible scaling [2, 3, 17], where equity, cost-effectiveness, and clinician acceptance determine long-term viability [4, 5, 12].

The trajectory from predictive analytics to autonomous governance thus reflects a deliberate engineering of healthcare as a learning system—one that ingests heterogeneous inputs, generates intelligence, executes or recommends actions, observes consequences, and refines itself under governed constraints. This framing offers policymakers and architects a blueprint for prioritising investments in interoperable data platforms, adaptive models, deployment protocols, and layered governance to realise sustainable autonomy.

Challenges and limitations

Despite documented progress, the literature identifies persistent barriers across the four pillars that constrain scalable, equitable integration of AI into healthcare systems.

Data foundations remain vulnerable to inconsistencies in quality, fragmentation, and bias propagation [1, 21, 26]. Heterogeneous sources—EHRs, imaging, wearables—often lack standardisation, leading to missingness, temporal misalignment, and representation gaps that degrade model generalisability [3, 22]. Bias in training data perpetuates disparities, particularly in underrepresented populations or low-resource settings [3, 17]. Privacy-preserving techniques like federated learning show promise but face implementation hurdles in real-world interoperability [25].

Analytic models encounter explainability deficits and performance drift [11, 26]. Black-box architectures undermine clinician trust and error attribution [4, 12], while temporal shifts in data distributions degrade performance in the absence of continuous monitoring [6, 10]. Generalist-to-specialist adaptation mitigates some issues but introduces new risks of over-generalisation or hallucination in generative contexts [7, 9].

Deployment challenges centre on workflow misalignment, acceptance barriers, and resource demands [4, 5, 12]. Clinicians report concerns about cognitive overload, deskilling, and liability when AI recommendations conflict with clinical judgment [11]. Cost-effectiveness analyses reveal high upfront investments and uncertain returns, especially for smaller organisations [5]. Integration with legacy systems often requires extensive customisation, delaying scale [22].

Governance layers struggle with regulatory fragmentation, accountability assignment, and ethical voids [13-15]. Oversight mechanisms lag behind autonomous capabilities, complicating liability in closed-loop scenarios [2, 9]. Equity auditing and bias mitigation require standardised protocols that remain underdeveloped [3, 17]. Resource-intensive frameworks disadvantage low- and middle-income settings [3].

Collectively, these limitations underscore that technical maturation outpaces infrastructural, human, and regulatory readiness. Without concerted mitigation—through standardised data pipelines, prospective monitoring, clinician co-design, and adaptive regulation—the risk persists that AI amplifies rather than alleviates systemic inequities and inefficiencies [5, 26]. The literature calls for holistic approaches that address these interdependencies rather than isolated fixes.

Future research directions

Building on identified gaps, future investigations should prioritise the following interconnected agendas to advance AI-enabled healthcare systems toward responsible autonomy.

First, longitudinal studies of closed-loop performance in live environments are essential. Current evidence relies heavily on controlled or short-term deployments [6, 10]; extended evaluations tracking drift, recalibration efficacy, and outcome sustainability across diverse populations would strengthen claims of adaptive intelligence [11, 26].

Second, comparative effectiveness research should benchmark integrated architectures against traditional systems on system-level endpoints—care coordination, resource utilisation, equity metrics—beyond individual prediction accuracy [5, 22, 23]. Multi-site trials incorporating cost-effectiveness and clinician workload would inform scalable investment priorities [5].

Third, governance innovation requires empirical validation. Frameworks for real-time ethical auditing, equity monitoring, and bounded autonomy need prospective testing in heterogeneous settings [2, 9, 13, 14]. Research on hybrid human-AI decision fusion—optimal allocation of authority across risk tiers—would guide safe progression toward greater autonomy [11].

Fourth, data infrastructure research should focus on federated, privacy-preserving platforms that enable equitable model training across resource gradients [3, 17, 25]. Techniques for bias detection and mitigation in dynamic, multimodal streams merit priority [1, 21].

Fifth, implementation science must examine organisational factors enabling sustained adoption—governance maturity models, clinician training pathways, and change management strategies [4, 6, 12]. Studies in low-resource contexts would address global disparities [3, 17].

Finally, interdisciplinary work bridging regulatory science, health economics, and ethics is needed to co-develop adaptive oversight models that evolve with technology [14, 15]. Simulation-based testing of autonomous scenarios could accelerate safe exploration without patient risk.

These directions, grounded in the reviewed corpus, aim to close the gap between demonstrated potential and widespread, equitable realisation of autonomous governance architectures.

Conclusion

The period 2017–2025 witnessed an AI transition in healthcare from supplementary predictive analytics to foundational infrastructural architectures capable of supporting autonomous governance. Predictive models provided the analytic core, deployment mechanisms embedded intelligence in workflows, closed-loop designs enabled continuous adaptation, and governance layers imposed necessary accountability.

This narrative review, synthesising high-impact publications through an original four-pillar lens—data foundations, analytic intelligence, deployment mechanisms, governance—demonstrates that maximal impact emerges when these elements are engineered as an integrated system rather than isolated components. The resulting architectures promise learning healthcare ecosystems that optimise clinical, operational, and equity outcomes at scale.

Persistent challenges—data bias, explainability gaps, deployment friction, regulatory lag—underscore the realisation that realising it depends on deliberate mitigation across technical, human, and policy dimensions. Future progress hinges on rigorous evaluation of closed loops, validation of governance innovations, and equitable infrastructure investments.

Ultimately, AI-enabled healthcare systems hold transformative potential when designed with systemic intent: ingesting diverse data, generating robust intelligence, supporting informed action, learning from outcomes, and governing autonomously yet responsibly. Achieving this vision requires sustained interdisciplinary commitment to bridge current limitations and translate demonstrated capabilities into pervasive, high-performance care delivery.

Acknowledgements
None

Conflict of interest
None

Financial support
None

Ethics statement
None

Received: 28 Aug 2025 Revised: 23 Sep 2025 Accepted: 01 Nov 2025
Published online: 20 January 2026

References

Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-8.
https://doi.org/10.1038/s41591-021-01614-0.

Silcox C, Zimlichmann E, Huber K, Rowen N, Saunders R, McClellan M, et al. The potential for artificial intelligence to transform healthcare: perspectives from international health leaders. NPJ Digit Med. 2024;7:88.
https://doi.org/10.1038/s41746-024-01097-6.

Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. NPJ Digit Med. 2022;5:162.
https://doi.org/10.1038/s41746-022-00700-y.

Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med. 2023;6(1):111.
https://doi.org/10.1038/s41746-023-00852-5.

El Arab RA, Al Moosa OA. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. NPJ Digit Med. 2025;8:548.
https://doi.org/10.1038/s41746-025-01722-y.

Wells BJ, Nguyen HM, McWilliams A, Pallini M, Bovi A, Kuzma A, et al. A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. NPJ Digit Med. 2025;8(1):514.
https://doi.org/10.1038/s41746-025-01714-7.

de Vere Hunt IJ, Jin KX, Linos E. A framework for considering the use of generative AI for health. NPJ Digit Med. 2025;8(1):297.
https://doi.org/10.1038/s41746-025-01695-y.

Yang M, Luo Y, He T, Diao S, Li H, Zou K, et al. Application of artificial intelligence to measure and predict patient values and preferences: a scoping review. NPJ Digit Med. 2025;8:769.
https://doi.org/10.1038/s41746-025-02156-2.

Wang Z, Wang H, Danek B, Li Y, Mack C, Arbuckle L, et al. A perspective for adapting generalist AI to specialized medical AI applications and their challenges. NPJ Digit Med. 2025;8:429.
https://doi.org/10.1038/s41746-025-01789-7.

Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health. 2024;6(5):e367-e373.
https://doi.org/10.1016/S2589-7500(24)00047-5.

Tikhomirov L, Semmler C, McCradden M, Searston R, Ghassemi M, Oakden-Rayner L. Medical artificial intelligence for clinicians: the lost cognitive perspective. Lancet Digit Health. 2024;6(8):e589-e594.
https://doi.org/10.1016/S2589-7500(24)00123-7.

Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health. 2021;3(9):e599-e611.
https://doi.org/10.1016/S2589-7500(21)00132-1.

Angus DC, Khera R, Lieu T, Bibbins-Domingo K, Perlis RH, Ross JS, et al. AI, health, and health care today and tomorrow: the JAMA summit report on artificial intelligence. JAMA. 2025;334(18):1650-64.
https://doi.org/10.1001/jama.2025.18490.

Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA. 2025;333(3):241-7.
https://doi.org/10.1001/jama.2024.21451.

Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. JAMA. 2020;323(6):509-10.
https://doi.org/10.1001/jama.2019.21579.

Howell MD. Generative artificial intelligence, patient safety and healthcare quality: a review. BMJ Qual Saf. 2024;33(11):748-54.
https://doi.org/10.1136/bmjqs-2024-016912.

Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018;3(4):e000798.
https://doi.org/10.1136/bmjgh-2018-000798.

Ashrafian H, Darzi A. Transforming health policy through machine learning. PLoS Med. 2018;15(11):e1002692.
https://doi.org/10.1371/journal.pmed.1002692.

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.

Saria S, Butte A, Sheikh A. Better medicine through machine learning: what's real, and what's artificial? PLoS Med. 2018;15(12):e1002721.
https://doi.org/10.1371/journal.pmed.1002721.

Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform. 2017;21(1):4-21.
https://doi.org/10.1109/JBHI.2016.2636665.

Sahni NR, Stein G, Zemmel R, Cutler DM. Artificial intelligence in US health care delivery. N Engl J Med. 2023;389(4):348-58.
https://doi.org/10.1056/NEJMra2204673.

Kohane IS. Compared with what? Measuring AI against the health care we have. N Engl J Med. 2024;391(17):1645-7.
https://doi.org/10.1056/NEJMp2404691.

Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-43.
https://doi.org/10.1136/svn-2017-000101.

Combi C, Pozzi G. Clinical information systems and artificial intelligence: recent research trends. Yearb Med Inform. 2019;28(1):83-94.
https://doi.org/10.1055/s-0039-1677903.

Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e750.
https://doi.org/10.1016/S2589-7500(21)00208-9.

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.

Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-9.
https://doi.org/10.1038/s41591-018-0316-z.