ELMER PRESS

Original Article

AI Clin Med. 2026;2:e20

Clinically Aligned AI Governance: Integrating Ethics, Risk, and Regulation in Healthcare

Rajesh Divakaran

Abstract

Background: Artificial intelligence (AI) systems increasingly influence clinical decision-making across healthcare, yet current governance approaches often treat AI as a technical compliance issue rather than an integral component of clinical care, creating critical gaps in patient safety, accountability, and trust. This paper examines how AI can be governed effectively in healthcare, addressing two questions: what distinguishes AI governance in healthcare from other sectors, and how can AI governance be integrated into existing clinical systems to manage risks across the full lifecycle.

Methods: Through systematic analysis of global normative frameworks (WHO, UN), cross-sector governance standards (OECD, NIST), and healthcare-specific regulations (EU AI Act, FDA guidance), alongside synthesis of empirical evidence from deployed clinical AI systems, the paper develops a layered governance architecture for healthcare organizations. Global and regional regulatory documents, published AI governance standards, and empirical case studies from healthcare AI deployments were reviewed and synthesized.

Results: Governance failures in healthcare AI arise less from ethical or regulatory gaps than from weak institutional translation into effective risk management and oversight practices. A three-layer clinically aligned governance architecture is identified for health care organizations, constituted by the three lines of the organizational oversight model. Layer 1—value articulation and clinical purpose, owned by executive and clinical leadership—establishes why AI should be used and under what conditions, bridging ethical intent to existing clinical policy and procurement structures. Layer 2—risk and control integration, anchored in patient safety, quality, and research governance systems—translates clinical intent into operational safeguards across the AI lifecycle. Layer 3—accountability, and assurance, led by internal audit and regulatory oversight functions—sustains institutional trust and continuous learning over time. Each layer bridges directly to an existing clinical governance structure, ensuring AI oversight is embedded within, rather than parallel to the established systems. Healthcare organizations adopting this lifecycle-oriented, clinically integrated approach are better positioned to ensure safe, explainable clinical decision-making while preserving professional judgment and patient trust.

Conclusions: This paper introduces clinically aligned AI governance—a layered governance architecture, embedding AI oversight within existing clinical, research, and patient safety systems—as foundational institutional infrastructure for healthcare. The architecture bridges the gap between generic AI governance frameworks and healthcare-specific needs, addressing persistent translation failures that undermine safe and accountable AI use in clinical settings. These findings have practical implications for healthcare leaders, regulators, and policymakers implementing clinical AI systems.

Keywords: Artificial intelligence; Clinical governance; Healthcare regulation; Patient safety; AI lifecycle; Accountability; Risk management

Manuscript submitted March 13, 2026, accepted March 27, 2026
Published online April 7, 2026

Office of the United Nations High Commissioner for Refugees (UNHCR), Geneva, Switzerland. Email: divakar@unhcr.org

doi: https://doi.org/10.14740/aicm20
AI in Clinical Medicine
2819-7437 (online)

Introduction: Why Clinically Aligned AI Governance Matters

In 2024, over 692 artificial intelligence (AI)-enabled medical devices received regulatory approval globally, with applications spanning diagnostic imaging, genomic analysis, and treatment optimization [1]. Yet the rapid integration of AI into clinical practice has outpaced the capacity of existing governance mechanisms. Health governance refers to the processes, institutions, and norms through which authority, accountability, and decision-making are exercised within the health system to protect public interests, ensure the quality and safety of care, and allocate resources responsibly [2]. Recent analyses document recurring patterns of algorithmic bias, performance degradation, and accountability gaps in deployed systems, often emerging only after clinical adoption [3, 4]. The central question is no longer whether AI requires governance, but how AI can be governed effectively within healthcare's distinctive institutional, professional, and regulatory context.

Across domains such as radiology, oncology, cardiology, genomics, and mental health, AI systems increasingly influence clinical judgment, patient access to care, and the allocation of scarce resources. Errors, bias, opacity, or inappropriate automation in these settings can translate directly into patient harm, inequitable outcomes, or erosion of professional ac-

countability [5–7]. The stakes are higher because healthcare decisions involve vulnerable populations, irreversible consequences, and fundamental rights related to bodily autonomy and informed consent.

In response to these AI-driven risks, a rapidly expanding body of AI governance frameworks has emerged. These span ethical principles, engineering standards, organizational risk management models, and binding legal and regulatory regimes. Recent analyses show growing convergence around lifecycle-based governance, post-deployment accountability, and real-world performance monitoring—developments driven less by abstract ethical concern than by evidence from concrete AI failures [3, 8]. While these frameworks articulate important expectations for "trustworthy" or "responsible" AI, they are largely cross-sectoral in origin and do not fully reflect the distinctive characteristics of healthcare.

Clinical environments are shaped by mature patient safety cultures, established risk management practices, formal evidence-generation regimes, and professional duties of care that differ fundamentally from those in other sectors [9]. Healthcare governance must address AI used in clinical research, where scientific validity and participant protection are central, as well as AI used in routine care, where decision support, efficiency, and resource allocation directly affect clinical judgment and patient care. Moreover, healthcare operates within complex and overlapping regulatory ecosystems governing medical devices, medicines, data protection, and clinical practice, each with distinct approval pathways, evidentiary thresholds, and oversight mechanisms.

This paper examines how AI can be governed effectively in healthcare by addressing two questions: 1) What makes AI governance in healthcare distinct from governance approaches in other sectors? 2) How can AI governance be integrated into existing clinical governance systems to manage AI-specific risks across the lifecycle?

To address these questions, the analysis integrates two bodies of work that have largely evolved in parallel: cross-sector AI governance frameworks and healthcare-specific regulatory and clinical governance regimes. By synthesizing these perspectives, the paper develops an approach to AI governance that is grounded in the realities of clinical practice, rather than treated as a parallel technical or compliance function.

The paper contributes to the field by defining clinically aligned AI governance, an approach that embeds AI oversight within established clinical, research, and patient safety systems. The analysis indicates that governance challenges in healthcare AI arise primarily from limitations in institutional translation rather than from gaps in ethical principles or regulation. Building on this diagnosis, the paper presents a layered governance architecture that clarifies roles, responsibilities, and oversight mechanisms to support safe, accountable, and trustworthy use of clinical AI.

The paper proceeds as follows. The second section maps the macro governance landscape for clinical AI, including global normative frameworks, healthcare-specific regulation, and cross-sector governance standards. The third section examines the distinctive governance challenges that arise when AI is applied in healthcare, focusing on professional accountability, evidentiary standards, lifecycle risk, and empirical harm. The fourth section analyses how existing AI governance frameworks function in practice and why their translation into healthcare settings remains problematic. The fifth section introduces the concept of clinically aligned AI governance and develops a layered organizational architecture aligned with established systems of clinical governance. The sixth section outlines implications and actionable recommendations for healthcare leaders and regulators. The seventh section concludes by reframing AI governance as institutional infrastructure essential to safe, equitable, and accountable healthcare.

Macro Governance Landscape for Clinical AI

AI governance in healthcare is shaped by a multilayered landscape that combines global normative frameworks, binding regional and national regulation, and cross-sector governance standards. Understanding this landscape is essential because healthcare organizations do not implement AI governance in a vacuum: they operate at the intersection of ethical commitments, legal obligations, and professional accountability regimes that together define expectations for safety and transparency.

This section maps that landscape at three levels: global normative frameworks, healthcare-specific regulatory regimes, and cross-sector governance standards that influence organizational practice.

Global normative frameworks: UN, WHO, and OECD

At the global level, AI governance is anchored in human-rights-based and public-interest-oriented frameworks developed by the United Nations system and intergovernmental organizations. For healthcare, the most influential of these is the World Health Organization's *Ethics and Governance of AI for Health*, which articulates six core principles: protection of autonomy; promotion of safety and the public interest; transparency; accountability; inclusiveness and equity; and sustainability. Unlike generic AI ethics frameworks, WHO situates AI governance explicitly within clinical ethics, patient safety, and professional responsibility, emphasizing validation, continuous monitoring, and context-appropriate deployment across both clinical research and service delivery [5].

These health-specific norms align with broader UN commitments to rights-based digital governance, including privacy protection, non-discrimination, and human rights due diligence in digital technologies. Recent UN system work underscores that ethical alignment alone is insufficient: organizations must develop concrete institutional capacity to operationalize these principles through governance structures, skills, and accountability mechanisms [10, 11].

The OECD Principles on Artificial Intelligence, first adopted in 2019 and revised in 2024, provide an internationally agreed baseline for trustworthy AI. They emphasize inclusive growth, human-centered values, transparency, robustness, and accountability, and have strongly influenced national AI

strategies and regulatory approaches [12]. The 2024 revision strengthened provisions related to safety, generative AI, misinformation, intellectual property, privacy, and environmental sustainability, reflecting lessons from real-world deployment.

Crucially, recent OECD work marks a shift from principle-setting toward evidence-driven governance. The OECD AI Incidents and Hazards Monitor documents thousands of AI failures across sectors, including healthcare, demonstrating that many harms arise post-deployment rather than at design stage [7]. This evolution reframes AI governance as a continuous, lifecycle-based process grounded in incident detection, escalation, and learning, rather than as a one-time ethical assessment [3, 8].

While influential, these global frameworks are normative rather than procedural. They align values, risk definitions, and expectations across jurisdictions but deliberately leave implementation to national regulators and organizations. As a result, their effectiveness in healthcare depends on how they are translated into binding regulations and institutional practice.

Regional and national regulatory contexts in healthcare

In healthcare, global normative commitments increasingly intersect with legally binding regulatory regimes. The European Union's Artificial Intelligence Act, adopted in 2024, represents the world's first comprehensive horizontal AI regulation. It establishes a risk-based framework that explicitly classifies most medical AI systems as "high-risk," triggering obligations related to risk management, data governance, transparency, human oversight, and post-market monitoring.

Importantly, the AI Act operates alongside existing EU healthcare regulation, notably the Medical Devices Regulation (MDR) and the *In Vitro Diagnostic* Regulation (IVDR). Joint guidance issued in 2025 by the Artificial Intelligence Board and the Medical Device Coordination Group clarifies that AI Act obligations complement, rather than replace, existing requirements for clinical evaluation, performance validation, quality management systems, and vigilance [13]. This creates a dual accountability model in which AI developers and healthcare providers must demonstrate both technical trustworthiness and clinical safety and effectiveness.

In the United States, AI governance in healthcare remains sector-specific rather than horizontal. The Food and Drug Administration has advanced a Total Product Lifecycle (TPLC) approach for AI-enabled software as a medical device, emphasizing pre-market assurance, real-world performance monitoring, and governance of adaptive systems through Good Machine Learning Practice (GMLP) and Predetermined Change Control Plans (PCCPs) [14, 15]. This approach prioritizes safety and effectiveness while allowing controlled post-deployment learning, contrasting with the EU's broader cross-sector risk-tiered model.

The United Kingdom has adopted a similarly healthcare-focused approach. The Medicines and Healthcare products Regulatory Agency (MHRA) updated guidance in 2025, strengthening lifecycle expectations for AI as a medical device, while the National Institute for Health and Care Excellence (NICE) continues to shape adoption decisions through its Evidence Standards Framework, which assesses clinical effectiveness, economic value, and system impact [16, 17]. Rather than imposing a comprehensive AI statute, the UK relies on sector-specific regulation supported by guidance, regulatory sandboxes, and evidence-based adoption pathways.

Across jurisdictions, a consistent pattern emerges: AI governance in healthcare is increasingly lifecycle-based, risk-proportionate, and evidence-driven. However, responsibility for implementation rests largely with healthcare organizations themselves, which must navigate overlapping ethical expectations, regulatory obligations, research governance requirements, and professional standards—often without a unified internal governance model capable of integrating them coherently.

Cross-sector AI governance frameworks

In parallel with healthcare-specific regulation, a range of cross-sector AI governance frameworks shape organizational practice. As outlined in Table 1 [18–24], these frameworks span voluntary ethical charters, technical standards, risk management models, and certifiable management systems, creating a continuum from principle to enforceable compliance.

Taken together, these frameworks clarify expectations for responsible AI but do not, on their own, resolve how AI should be governed in clinical practice. These frameworks emphasize lifecycle governance.

Distinctive Governance Challenges in Healthcare AI

While the second section outlines the external governance landscape, the challenges below arise when these frameworks are applied within real clinical environments. Gaps remain between normative guidance and procedural execution, particularly where AI systems move from research environments into routine care, fragmenting responsibility across technical, clinical, legal, and research functions [25, 26].

The macro governance landscape for clinical AI is increasingly mature in terms of principles and regulation. What remains underdeveloped is the organizational capacity to integrate these instruments into coherent, clinically grounded governance arrangements. Additionally, there are distinctive features of healthcare—professional accountability, evidence standards, lifecycle risk, and liability—that complicate the translation of AI governance from policy to practice.

AI governance challenges manifest differently in healthcare than in other sectors because clinical environments combine professional accountability, evidence-based decision-making, and regulatory oversight in ways that directly shape patient outcomes. While many AI risks—bias, opacity, system failure—are cross-sectoral, their consequences in healthcare are uniquely consequential, immediate, and often irreversible. This section examines four governance challenges that distinguish clinical AI from AI deployed in other domains.

Table 1. Cross-Sector AI Governance Frameworks

Framework	Type	Core focus	Strengths	Limitations
OECD AI Principles (2019, rev. 2024) [18]	Global policy principles	Human-centered values, transparency, robustness, accountability	Widely adopted; foundation for national strategies	High-level; voluntary
Montréal Declaration (2018) [19]	Ethical charter	Societal values and public engagement	Normative legitimacy	No enforcement
IEEE 7000 Series (2021) [20]	Technical standards	Ethics-by-design	Operationalizes values	Not legally binding
AIGA Framework (2022) [21]	Governance and auditing model	Linking principles to controls	Bridges legal, ethical, technical domains	Requires mature management systems
ISO/IEC 42001:2023 [22]	Management system standard	Enterprise AI governance	Certifiable; institutional accountability	Risk of formalistic compliance
NIST AI RMF (2023) [23]	Risk management framework	Lifecycle risk governance	Flexible and iterative	Voluntary
EU AI Act (2024) [24]	Binding regulation	Risk-based compliance	Strong enforcement	High implementation complexity

Summary of the most widely referenced cross-sector AI governance frameworks, including their type, core focus, strengths, and limitations in the context of healthcare applications.

Clinical context: high stakes and professional accountability

Healthcare AI operates in settings where errors can cause direct physical harm, delayed treatment, or irreversible clinical outcomes. Clinical decision-making is governed by professional accountability frameworks—codes of ethics, licensing regimes, and legal duties of care—that predate AI and remain binding regardless of algorithmic involvement [27]. AI systems, therefore, cannot be treated as neutral decision aids or administrative tools; their outputs directly influence clinical judgment and, by extension, professional responsibility.

Healthcare has also developed mature norms for evaluating evidence quality through randomized controlled trials, clinical validation studies, and practice guidelines. Yet many AI systems enter clinical environments with limited external validation, having been trained and tested on narrow or non-representative datasets under controlled conditions [28]. This creates a mismatch between the evidentiary standards expected of clinical interventions and those often applied to AI systems.

Moreover, healthcare organizations maintain established patient safety cultures—incident reporting systems, root cause analyses, and quality improvement cycles—that are designed to surface, investigate, and learn from harm. When AI governance is managed outside these structures, algorithmic risks remain invisible to safety oversight, weakening institutional learning. AI systems that optimize performance for average populations may also systematically disadvantage patients with lower health literacy, language barriers, or complex social needs, reinforcing inequities that healthcare ethics explicitly seek to mitigate [5].

Lifecycle risk: from static validation to continuous oversight

A defining governance challenge in healthcare AI arises from the dynamic nature of many deployed systems. Traditional medical device regulation assumes stability after approval; however, AI models may degrade as patient populations change, clinical workflows evolve, or data inputs shift. Empirical evidence demonstrates that post-deployment performance decay is common: van der Vorst et al [4] documented clinically significant data drift within 18 months of deployment, while OECD analysis of over 1,100 AI incidents found that most harms emerged after systems entered real-world use rather than at design stage [3].

Some AI systems are explicitly designed to adapt after deployment, updating parameters in response to new data. While such adaptability can improve performance, it complicates validation, accountability, and regulatory oversight [15]. These dynamics challenge governance models based on point-in-time approval. In healthcare, lifecycle governance must extend beyond individual models to encompass ongoing monitoring, change control, and escalation mechanisms capable of detecting and responding to performance degradation, bias, or unsafe interactions as they occur.

Regulatory fragmentation and overlapping obligations

Healthcare AI is subject to multiple, overlapping regulatory regimes that vary across jurisdictions. Medical device regulations govern systems used for diagnosis or treatment, while data protection laws impose constraints on the processing of personal health data. Horizontal AI regulation, such as the EU Artificial Intelligence Act, adds further obligations related to risk management, transparency, and post-market monitoring. These regimes were developed independently, resulting in fragmented oversight responsibilities [29].

Regulatory approaches also diverge internationally. The EU classifies most clinical AI as high-risk under the AI Act, imposing prescriptive compliance obligations. The United States re-

lies on the FDA's adaptive, lifecycle-oriented oversight model, while the United Kingdom emphasizes evidence standards and post-market evaluation through NICE and MHRA guidance [13, 15, 17]. For multinational healthcare organizations and AI developers, this variation complicates compliance and creates uncertainty about accountability expectations.

At the organizational level, regulatory fragmentation often translates into siloed governance: legal teams focus on compliance, technical teams manage system performance, and clinicians remain accountable for patient outcomes without full visibility into algorithmic limitations. Without an integrating governance model, these overlapping obligations can obscure responsibility rather than strengthen oversight.

Accountability gaps and real-world harm

AI complicates traditional models of accountability in healthcare. Medical liability frameworks assume clinician responsibility for independent judgment, yet AI systems increasingly shape diagnostic, triage, and treatment decisions. When errors occur, responsibility may be distributed across clinicians, healthcare organizations, vendors, and data providers, creating ambiguity that undermines learning and remediation [20].

Multiple real-world cases illustrate the consequences of these accountability gaps. Proprietary AI systems used by insurers for prior authorization have denied medically necessary care without meaningful human review, with a high proportion of decisions later overturned on appeal [30, 31]. In clinical care, the widely deployed Epic Sepsis Model failed to identify a majority of sepsis cases during external validation, contributing to delayed intervention and clinician alert fatigue [32]. In another case, unmonitored data drift in a surgical discharge prediction model led to unsafe discharge recommendations and increased readmission risk [33].

The emergence of "shadow AI"—clinicians using unvetted public tools such as generative chatbots for clinical tasks—further demonstrates how governance gaps enable unsafe practices when formal pathways for innovation and oversight are absent [34]. Conversely, institutions that established formal, multidisciplinary AI governance committees reported clearer accountability, earlier detection of risks, and safer adoption of clinical AI systems [35].

Across these cases, four consistent lessons emerge: 1) External validation in diverse clinical settings is essential for high-risk AI. 2) Continuous monitoring for performance drift and bias must be institutionalized. 3) Clear accountability structures are necessary to prevent governance gaps. 4) Proactive governance frameworks are required to manage both sanctioned and unsanctioned AI use.

The governance challenges of healthcare AI arise not from a lack of ethical principles or regulatory attention, but from the interaction of dynamic technologies with clinical accountability, evidentiary standards, and fragmented oversight structures. These challenges cannot be addressed through generic AI governance alone. They require governance approaches that are clinically embedded, lifecycle-oriented, and institutionally integrated—setting the foundation for the clinically aligned governance model developed in the following sections.

Translating AI Governance Frameworks Into Healthcare Practice

Defining clinically aligned AI governance

Clinically aligned AI governance refers to an approach to overseeing artificial intelligence in healthcare that situates AI systems within established clinical, research, and patient safety governance structures. Rather than treating AI as a standalone technical or compliance concern, this approach frames AI-enabled systems as components of clinical practice whose development, deployment, and use must align with clinical values, patient safety obligations, and equity objectives.

Clinically aligned governance is characterized by three interrelated features. First, AI oversight is embedded within existing clinical governance mechanisms such as medical leadership committees, patient safety and quality improvement programs, research ethics oversight, and institutional ethics bodies, ensuring that clinical expertise and professional accountability inform AI-related decisions. Second, governance extends across the full AI lifecycle, encompassing problem definition, validation, deployment, performance monitoring, updating, and decommissioning, recognizing that point-in-time approval is insufficient for systems whose behavior may evolve in use. Third, accountability for AI is distributed across multidisciplinary actors but formalized through clearly defined roles, decision rights, and escalation pathways, reducing ambiguity when systems underperform or cause harm.

Adaptive and learning AI systems bring these requirements into sharper focus. Systems that continue to update after deployment challenge governance and regulatory assumptions premised on static performance, underscoring the need for continuous oversight and explicit accountability. Together, these elements define clinically aligned AI governance as a means of ensuring that AI supports clinical judgment and operates within robust institutional frameworks of safety and responsibility.

Translating governance frameworks into clinical practice

Many healthcare organizations have responded by creating AI ethics boards or technical review committees. While valuable, these structures often operate in parallel to clinical governance, limiting their influence on patient safety, accountability, and professional practice. Clinically aligned AI governance differs by embedding decision rights, escalation pathways, and lifecycle oversight within existing clinical and safety systems rather than adding a separate governance layer.

While the principles underlying clinically aligned AI governance are increasingly reflected in global normative frameworks, cross-sector standards, and healthcare regulation, their

Table 2. Functional Roles of AI Governance Frameworks in Healthcare Settings

Framework category	Representative frameworks	Primary function	What it does not do
Ethical/normative	OECD AI Principles; Montréal Declaration; WHO AI Ethics	Establish legitimacy, values, and expectations	Define clinical safety thresholds or enforce practice
Risk and organizational methods	IEEE 7000 Series; NIST AI RMF; AIGA Framework	Translate values into risk identification, controls, and documentation	Substitute for clinical governance
Management and legal instruments	ISO/IEC 42001; EU AI Act; MDR/IVDR	Institutionalize roles, auditability, and enforcement	Govern bedside clinical decisions

Classification of AI governance frameworks by their primary functional role: ethical/normative frameworks that establish legitimacy and values; risk and organizational methods that translate values into controls; and management and legal instruments that institutionalize accountability and enforcement.

translation into organizational practice remains uneven. Governance failures in healthcare AI rarely arise from an absence of ethical guidance or regulatory expectation; instead, they occur when governance instruments are applied in isolation or treated as parallel requirements rather than integrated into clinical oversight structures.

AI governance instruments perform distinct governance functions. Ethical and normative frameworks articulate values and legitimacy; risk management and organizational methods translate those values into processes and controls; and legal and management instruments institutionalize accountability through documentation, audit, and enforcement. In healthcare settings, effective governance depends on aligning these functions with existing clinical and research oversight arrangements rather than adopting frameworks as standalone solutions.

Functional roles of AI governance frameworks

This functional distinction clarifies why no single framework is sufficient to govern clinical AI and why organizational integration, rather than framework selection, is the central governance challenge (Table 2).

Why cross-sector frameworks fall short in healthcare

Most widely adopted AI governance frameworks are cross-sectoral by design. While this generality supports global alignment, it limits direct applicability in clinical settings.

First, cross-sector frameworks do not define clinical evidentiary standards. Requirements for transparency, robustness, or risk assessment do not specify acceptable thresholds for clinical performance, external validation, or patient safety.

Second, these frameworks rarely align with established clinical oversight mechanisms. Concepts such as "human oversight" or "risk ownership" remain abstract and are not mapped to patient safety committees, morbidity and mortality reviews, research ethics boards, or clinical leadership structures.

Third, cross-sector frameworks do not manage the transition from research to routine care. Ethical approval or technical validation in research contexts does not automatically translate into accountability, liability, or safety oversight in clinical service delivery [5, 7].

As a result, healthcare organizations may formally comply with AI governance frameworks while leaving critical clinical risks unmanaged.

Mapping AI governance to existing healthcare oversight models

Translation into practice requires alignment with governance models already familiar to healthcare organizations. One such model is the "Three Lines" framework used for risk and assurance [36]: 1) First line (clinical, executive, and research leadership) owns AI lifecycle management, including problem definition, deployment decisions, and day-to-day use. 2) Second line (risk management, quality and safety, data governance, IT, research governance) provides guidance, challenge, and structured oversight. 3) Third line (internal audit, regulators, notified bodies) offers independent assurance that governance remains effective over time.

This mapping clarifies that AI governance is an extension of existing accountability arrangements rather than a technical add-on.

Evidence from organizational practice

Empirical evidence from healthcare organizations supports this functional interpretation of governance. Institutions that treated AI governance as a cross-functional capability rather than as an ethics checklist or IT policy demonstrated improved oversight and safer adoption.

Examples include a Canadian health system implementing the People, Process, Technology, and Operations (PPTO) model, which aligned ethics, risk management, and operational readiness through co-design and cross-functional governance [37], and University of Wisconsin Health's multidisciplinary AI Steering Committee, which established shared validation and lifecycle monitoring processes [38]. A two-center surgical discharge study further demonstrated the value of proactive drift detection and continuous monitoring in preventing patient harm [4].

These cases indicate that governance effectiveness depends less on the specific framework adopted than on whether governance functions are clearly assigned, operationalized, and integrated into routine clinical oversight.

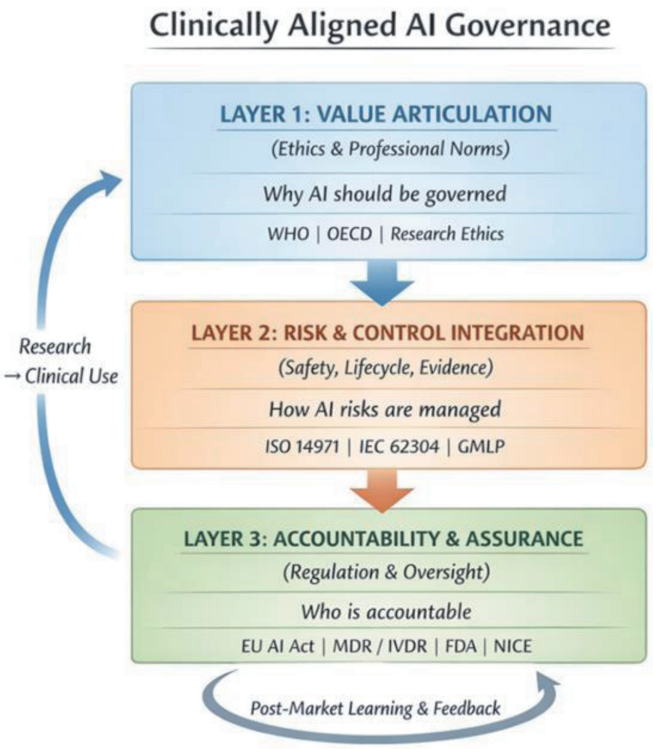

Figure 1. Clinically aligned AI governance: three-layer governance architecture. The architecture comprises three layers constituted by the organizational Three Lines model: Layer 1 (value articulation and clinical purpose; owned by executive and clinical leadership, first line of defense), Layer 2 (risk and control integration; anchored in patient safety, quality, and research governance functions, second line), and Layer 3 (accountability, and assurance; led by internal audit, regulators, and independent reviewers, third line). Bidirectional feedback and escalation pathways connect all three layers. Bidirectional arrows indicate feedback and escalation pathways between layers. Each layer bridges directly to an existing clinical governance structure, embedding AI oversight within, rather than parallel to, established systems of clinical care and accountability.

AI governance frameworks provide essential but partial tools for managing clinical AI. Ethical principles establish legitimacy, risk frameworks translate intent into controls, and legal instruments enforce accountability, but none governs clinical practice on its own. The persistent gap lies in organizational integration. Addressing this gap requires moving from framework adoption to governance capability, a transition developed in the next section.

A Layered Governance Architecture

Clinically aligned AI governance can be understood as a layered organizational architecture comprising three interdependent layers. As outlined in Figure 1, each layer performs a distinct governance function, and together they ensure that AI systems remain aligned with clinical values, safety requirements, and institutional accountability. There may be substantial overlap between these layers, and they should not be viewed as entirely distinct components.

Layer 1: value articulation and clinical purpose

Function

This layer anchors AI use in clinical purpose, ethical boundaries, and duty of care. It establishes why AI should be used, for whom, and under what conditions.

Key activities

Defining acceptable and prohibited AI use cases;

Articulating intended clinical benefit and foreseeable risk;

Aligning AI initiatives with institutional missions, care priorities, and equity objectives;

Embedding ethical expectations into clinical policies, protocols, and procurement decisions.

Ownership

Executive leadership, clinical leadership, and research leadership, supported by ethics committees and clinical governance bodies.

This layer prevents "technology-first" adoption by requiring that AI initiatives demonstrate clinical justification before technical optimization or deployment decisions are made.

Layer 2: risk and control integration

Function

This layer translates clinical intent into operational safeguards that manage risk during development and use. There are several tools available, and the key ones are listed in Table 3.

Key activities

AI-specific risk assessment (e.g., automation bias, data shift, workflow interaction);

Validation and version control appropriate to clinical risk;

Integration of AI monitoring into quality, safety, and research oversight systems;

Change control processes for model updates, retraining, and adaptation;

Incident reporting and corrective action linked to patient safety and quality improvement mechanisms.

Ownership

Clinical and research teams as first-line owners, with structured oversight from risk management, quality and safety, data governance, IT, and research governance functions.

This layer ensures that AI risks are managed using the

Table 3. Risk and Control Standards by Clinical Use Case

Clinical use case	Primary risks	Key standards	Governance focus
AI as medical device	Patient harm, drift	ISO 14971[a]; IEC 62304[b]; GMLP	Safety, traceability
Clinical decision support	Over-reliance, opacity	ISO/IEC 82304-1[c]; ISO 9241[d]; NIST AI RMF	Human oversight
Clinical research AI	Bias, invalid inference	Declaration of Helsinki[e]; ICH-GCP[f]; OECD Health Data	Research integrity
Molecular/genomic AI	Reproducibility, misuse	Domain bioinformatics standards; NIST AI RMF	Scientific validity
Adaptive AI systems	Drift, inequity	GMLP; PCCPs[g]	Continuous oversight and learning

Mapping of primary AI-related risks, applicable governance standards, and governance focus areas across five key clinical AI deployment contexts: AI as medical device, clinical decision support, clinical research AI, molecular/genomic AI, and adaptive AI systems. Superscript letters (a–g) refer to footnotes describing the relevant international standards. [a]ISO 14971: Standard for risk management of medical devices across the lifecycle. [b]IEC 62304: Standard for safe lifecycle management of medical device software. [c]ISO/IEC 82304-1: Requirements for safety and quality of health software products. [d]ISO 9241: Standards on usability and human-centred design of interactive systems. [e]Declaration of Helsinki: Ethical principles for medical research involving human participants. [f]ICH-GCP: International standard for ethical and scientific conduct of clinical trials. [g]PCCPs: Regulatory approach for controlled post-deployment changes to machine learning medical devices.

same infrastructures healthcare organizations already rely on for other high-risk interventions.

Layer 3: accountability, assurance, and learning

Function

This layer sustains trust, compliance, and institutional learning over time.

Key activities

Compliance with binding regulatory regimes and professional standards;

Independent review, audit, and assurance of governance effectiveness;

Evidence generation for adoption, commissioning, and continued use;

External reporting and post-market surveillance where required;

Organizational learning through feedback loops, audits, and governance review.

Ownership

Internal audit functions, regulators, notified bodies, and external reviewers, with ultimate accountability retained by senior leadership.

This layer ensures that governance remains credible as technologies evolve, organizational structures change, and regulatory expectations mature.

Why layering matters

The layered architecture clarifies roles and prevents governance failure modes commonly observed in healthcare AI deployment. Ethical intent without operational controls becomes symbolic. Risk controls without accountability lose authority. Compliance without clinical anchoring risks formal adherence without patient safety impact.

By separating, but aligning, value articulation, risk integration, and assurance, clinically aligned governance enables organizations to manage AI in a way that is both rigorous and clinically credible.

Clinically aligned AI governance is not a new framework but a way of organizing existing governance functions so that AI is governed as part of clinical practice rather than alongside it. The layered architecture provides healthcare organizations with a practical structure for embedding AI oversight within established systems of care, research, and accountability. The next section examines what this organizational capability implies for clinical leaders and regulators responsible for its implementation.

Implications for Clinical Leaders and Regulators

Discussion: from governance architecture to institutional action

The clinically aligned governance model developed in this paper reframes AI oversight from a compliance or technical exercise into a question of institutional design. While ethical principles, regulatory requirements, and technical safeguards are increasingly well articulated, failures in clinical AI governance demonstrate that misalignment between these elements—rather than their absence—is the primary source of risk. In practice, AI-related harm emerges when clinical responsibility, operational control, and institutional accountability are distributed across organizational silos without clear ownership or escalation pathways.

This misalignment is particularly consequential in healthcare because AI systems operate within environments already characterized by complex accountability structures, high-stakes decision-making, and established safety cultures. Treat-

ing AI as an exception governed through parallel committees, ad hoc policies, or vendor assurances undermines these structures and weakens professional accountability. Conversely, embedding AI governance within existing clinical, research, and patient safety systems leverages institutional knowledge, reinforces duty of care, and enables continuous learning.

The implications that follow therefore focus not on introducing new governance instruments, but on clarifying ownership, embedding lifecycle oversight, and strengthening organizational capacity. The recommendations are directed at the points where governance most often fails in practice: leadership accountability, integration into safety systems, operationalization of human oversight, and regulatory alignment across the AI lifecycle.

Recommendations for clinical and executive leadership

Accountability

AI used in patient care must sit under clear clinical ownership, with defined responsibility for validation, safe use, monitoring, and decommissioning. Clinicians should know when to rely on AI, when to override it, and how to escalate concerns.

Three Lines model

AI adoption, updates, and retirement should be based on structured impact assessments and clear clinical value, with the management team driving design and delivery, the risk and oversight teams providing guidance and challenge, and the auditing and evaluating offering independent assurance. The model should be built to integrate seamlessly into existing clinical, risk, and operational processes so that AI decisions follow the same pathways and controls.

Lifecycle governance

AI should be managed as a clinical change across its lifecycle: structured review at deployment, controlled processes for updates, and routine checks on real-world performance. Model and workflow changes must be transparent, auditable, and tied to defined approval thresholds.

Integration into safety and quality

AI-related risks, incidents, and performance signals must flow through existing patient safety reporting, audit, and quality-improvement systems. Continuous monitoring and incident learning should directly inform practice adjustments and design improvements.

Organizational capability

Effective AI governance requires sustained investment in skills, cross-functional oversight, and monitoring infrastructure. Governance should be embedded within institutional risk management, with clear objectives, resourcing, and regular review.

Workforce readiness and equity

Clinicians and staff need practical training on AI's strengths, limitations, and accountability in everyday practice. Support and safeguards should be accessible across all care settings to enable safe, consistent, and equitable adoption.

Taken together, these priorities provide a practical and clinically grounded approach to governing AI in healthcare. By anchoring AI decisions within existing structures for accountability, safety, and organizational oversight, health systems can adopt innovation with confidence while protecting patients and maintaining professional trust.

Conclusions

This paper has argued that effective AI governance in healthcare requires a clinically aligned, layered organizational capability rather than reliance on ethical principles, technical controls, or regulatory compliance in isolation. While global frameworks and healthcare-specific regulations have matured rapidly, persistent governance failures demonstrate that principles and rules alone do not ensure safe, equitable, and accountable use of AI in clinical settings.

Governance maturity emerges when responsibility is institutionalized across three interdependent functions: articulation of clinical purpose and ethical boundaries, integration of AI risks into operational safety systems, and sustained accountability through assurance and learning. Where these functions are misaligned or fragmented, AI-related harms persist despite formal adherence to governance frameworks.

As AI becomes increasingly embedded in diagnosis, treatment, research, and health system decision-making, clinically aligned AI governance should be understood as foundational institutional infrastructure, not optional compliance. Strengthening this capability is essential to protecting patient safety, preserving professional judgment, and maintaining public trust in an increasingly AI-mediated healthcare environment.

The path forward is therefore not the creation of additional principles or frameworks, but the deliberate integration of existing governance tools into clinical practice. Healthcare institutions and regulators that succeed in this integration will be better positioned to realize the benefits of AI while upholding the core values of healthcare: safety, equity, accountability, and patient-centered care.

Acknowledgments

The author thanks colleagues at UNHCR and the broader global health and AI governance community for discussions that informed this work. No specific individuals require acknowledgement.

Financial Disclosure

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author declares no financial relationships with any organizations that might have an interest in the submitted work.

Conflict of Interest

The author declares no conflicts of interest.

Informed Consent

Not applicable. This paper is a narrative review and does not involve human participants, animal experiments, or identifiable patient data.

Author Contributions

In accordance with the ICMJE criteria (www.icmje.org), Rajesh Divakaran (sole author) fulfils all four authorship criteria: 1) substantial contributions to the conception and design of the work, and the acquisition, analysis, and interpretation of data; 2) drafting the work and revising it critically for important intellectual content; 3) final approval of the version to be published; 4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data Availability

The authors declare that data supporting the findings of this study are available within the article.

Disclaimer

The views expressed in this paper are those of the author alone and do not represent the official position of UNHCR or the United Nations.

References

1. Food and Drug Administration. Artificial intelligence/machine learning (AI/ML) action plan. Silver Spring: U.S. Department of Health and Human Services; 2024. [Author note: Verify full citation details].
2. World Health Organization. Health systems governance for universal health coverage: Action plan. Geneva: WHO; 2014.
3. Organisation for economic co-operation and development. Towards a common reporting framework for AI incidents. OECD Artificial Intelligence Papers No. 34. Paris: OECD Publishing; 2025.
4. van der Vorst JP, Smit JM, van de Sande D, van der Ster B, Daams F, Schasfoort R, et al. Importance of model governance in clinical AI models: Case study on the relevance of data drift detection. BMJ Digit Health AI. 2025;1:e000046.
5. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021.
6. Cui S, Traverso A, Niraula D, Zou J, Luo Y, Owen D, El Naqa I, et al. Interpretable artificial intelligence in radiology and radiation oncology. Br J Radiol. 2023;96(1150):20230142. doi pubmed
7. OECD.AI. OECD AI incidents and hazards monitor [Internet]. Paris: OECD; 2025. Available from: https://oecd.ai/en/incidents.
8. Leopard AS, Epstein D. Global AI governance: Five key frameworks explained. Bradley LLP. 2025.
9. Coglianese C, Lehr D. Transparency and algorithmic governance. Admin Law Rev. 2019;71(1):1-42.
10. United Nations. Governing AI for humanity: final report of the Secretary-General's high-level advisory body on artificial intelligence. New York: United Nations. 2024.
11. OHCHR. Human rights due diligence for digital technology use - Guidance of the Secretary-General: Practical Guide. Geneva: Office of the High Commissioner for Human Rights. 2025.
12. Organisation for economic co-operation and development. Recommendation of the council on artificial intelligence (OECD/LEGAL/0449). Paris: OECD Publishing. 2019.
13. European commission, artificial intelligence board, medical device coordination group. Joint guidance on the interplay between the Artificial Intelligence Act and MDR/IVDR. Brussels: European Commission; 2025.
14. Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. Silver Spring: U.S. Department of Health and Human Services; 2021.
15. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Predetermined change control plans for machine learning-enabled medical devices. 2023.
16. Medicines and healthcare products regulatory agency. Software and artificial intelligence as a medical device: Updated guidance. London: MHRA; 2025.
17. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. London: NICE; 2024.
18. OECD AI Principles (2019, rev. 2024). Organisation for Economic Co-operation and Development. OECD Principles on Artificial Intelligence. Paris: OECD; 2019. Revised 2024.
19. Montreal Declaration (2018). Universite de Montreal. Montreal Declaration for a Responsible Development of Artificial Intelligence. Montreal: Universite de Montreal; 2018.
20. IEEE 7000 Series (2021). Institute of Electrical and Elec-

tronics Engineers. IEEE 7000™ Series Standards for Addressing Ethical Concerns in System Design. New York: IEEE; 2021.
21. AIGA Framework (2022). University of Turku; AIGA Consortium. Artificial Intelligence Governance and Auditing (AIGA) Framework. Turku: University of Turku; 2022.
22. ISO/IEC 42001(2023). International Organization for Standardization; International Electrotechnical Commission. ISO/IEC 42001:2023 - Artificial Intelligence Management System. Geneva: ISO/IEC; 2023.
23. NIST AI RMF (2023). National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg (MD): NIST; 2023.
24. EU AI Act (2024). European Parliament; Council of the European Union. Regulation (EU) Artificial Intelligence Act. Brussels: EU; 2024.
25. OECD.AI. OECD AI incidents and hazards monitor. 2024. [Author note: Verify - may refer to the 2025 edition listed above].
26. Plonk A, Perset K. Towards Rights-Based and Trustworthy Digital Governance: OECD Framework for Institutional Capacity in the Public Sector. Paris: OECD; 2025.
27. Kerasidou A, Kerasidou CX. Epistemic authority and medical AI: epistemological differences and challenges in medical practice. Med Health Care Philos. 2026;29(1):89-95. doi pubmed
28. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ. 2020;98(4):251-256. doi pubmed
29. European Commission. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024.
30. Rajkomar A, Dean J, Kohane I. AI pitfalls and what not to do: Mitigating bias in AI. Proc Natl Acad Sci. 2023. Article PMC10546443. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10546443/.
31. 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):447-453. doi pubmed
32. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, Pestrue J, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065-1070. doi pubmed
33. Kelly CJ, Karthikesalingam A, King D. A machine learning-based discharge prediction model and the risks of data drift. Healthcare. 2023;10(6):966.
34. Reddy S, Allan S, Coghlan S, Cooper P. Governance of generative artificial intelligence in clinical care. Lancet Digit Health. 2024;6(2):e123-e130.
35. Liu X, Cruz Rivera S, Moher D, Calvert M, Denniston AK. Reporting guidelines for clinical artificial intelligence research. Nat Med. 2023;29:19-27.
36. Institute of Internal Auditors. The IIA's three lines model. Lake Mary: The Institute of Internal Auditors; 2020.
37. Habib ARR, Lin S, Grant J, et al. A patient-centred governance framework for the use of artificial intelligence in healthcare. Healthc Manage Forum. 2021;34(4):192-197.
38. Liao F, Adelaine S, Afshar M, Patterson BW. Governance of Clinical AI applications to facilitate safe and equitable deployment in a large health system: Key elements and early successes. Front Digit Health. 2022;4:931439. doi pubmed