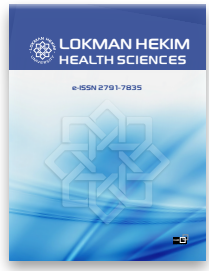
LOKMAN HEKIM HEALTH SCIENCES

DOI: 10.14744/lhhs.2026.37118
Lokman Hekim Health Sci 2026;6(2):372–382



lokmanhekimhs.com

REVIEW

The Governance of Artificial Intelligence in Healthcare: Ethical Foundations, Legal Challenges, and Implementation Realities

Mahmut Yılmaz

Department of Intensive Care, İzmir City Hospital, İzmir, Türkiye

Abstract

Artificial intelligence (AI) is transforming healthcare across diagnostics, decision-making, and clinical workflows, yet its integration raises complex ethical, legal, and operational challenges. This narrative review synthesizes three traditionally fragmented domains: Ethical principles, legal accountability, and implementation realities. We draw on literature from major databases alongside regulatory frameworks, including the World Health Organization, the Organisation for Economic Co-operation and Development, the National Institute of Standards and Technology, the European Union (EU), the Food and Drug Administration (FDA), and the International Medical Device Regulators Forum, and examine Türkiye's policies (e.g., Personal Data Protection Law No. 6698) to provide a middle-income country perspective. This review makes three contributions. First, we reconceptualize core bioethical principles – autonomy, beneficence, non-maleficence, and justice – in AI-mediated settings, emphasizing transparency, human oversight, and equity-sensitive design. Second, we frame legal accountability as a distributed system involving developers, institutions, and clinicians. Third, we bridge theory and practice through real-world cases (sepsis prediction vs. proprietary algorithms) and propose an integrated lifecycle governance model. Comparative analysis of the EU AI Act, FDA's 2026 guidance, and Türkiye's regulatory landscape shows convergence toward risk-based governance, alongside persistent gaps, particularly in middle-income settings. Responsible AI governance requires not only regulatory compliance but also continuous evaluation, transparency, and human-centered oversight. Despite global convergence on high-level principles, significant gaps remain in translating these into enforceable mechanisms and clinical practice. Future research should prioritize empirically validated governance models that ensure AI augments – rather than undermines – clinical judgment and patient trust.

Keywords: Artificial intelligence; Clinical implementation; Governance; Healthcare ethics; Legal accountability; Regulatory frameworks; Türkiye

Artificial intelligence (AI) has moved from an experimental technology to an increasingly routine component of healthcare systems. It now shapes diagnostic pathways, prognostic assessments, clinical decision-support tools, and even the administrative and logistical functions of health institutions. The appeal of AI – particularly machine learning models and, more recently, large language models – lies in its ability to process complex data, recognize patterns at scale, and potentially tailor care to individual patients. However, healthcare is not a neutral technological environment. Decisions made in clinical settings carry immediate and sometimes irreversible consequences, meaning that errors,

Cite this article as: *Yilmaz M. The Governance of Artificial Intelligence in Healthcare: Ethical Foundations, Legal Challenges, and Implementation Realities. Lokman Hekim Health Sci 2026;6(2):372–382.*

Correspondence: Mahmut Yılmaz, M.D. İzmir Şehir Hastanesi, Yoğun Bakım Kliniği, İzmir, Türkiye
E-mail: mahmutyilmazmd@gmail.com **Submitted:** 01.05.2026 **Revised:** 01.05.2026 **Accepted:** 04.05.2026 **Available Online:** 16.06.2026

hidden biases, or opaque decision-making processes can directly translate into patient harm. For this reason, AI in healthcare must be understood not simply as a technical advancement but as a socio-technical intervention that reshapes patient rights, professional responsibilities, and institutional accountability. International guidance increasingly stresses that the benefits of AI can only be sustained when systems are governed with explicit attention to safety, effectiveness, transparency, and responsible oversight throughout their lifecycle.[1,2]

Recent discussions on trustworthy AI in healthcare have further emphasized that adoption is increasingly recognized to be slowed not by technical limitations but by unresolved ethical concerns. Issues such as algorithmic bias, lack of transparency, privacy risks, and ambiguous accountability structures continue to undermine trust. These concerns point toward the need for multidimensional governance models that integrate fairness, robustness, safety, transparency, privacy, and accountability as interconnected rather than isolated principles.[3,4] Recent meta-analyses of global AI ethics guidelines further demonstrate a growing convergence around core principles – including transparency, justice, non-maleficence, responsibility, and privacy – while simultaneously highlighting substantial variation in their implementation across jurisdictions and sectors.[5] Similarly, unified ethical frameworks have proposed adding explicability as a core principle, linking transparency directly to accountability and interpretability in algorithmic systems.[6] At the same time, governance discussions remain heavily concentrated on European Union (EU) and United States (US) regulatory discourse, whereas implementation in middle-income jurisdictions often occurs within fragmented legal and institutional environments.

This review, therefore, synthesizes recent literature and regulatory documents to examine the ethical foundations, legal accountability structures, and emerging governance approaches for responsible AI integration into healthcare systems.

Materials and Methods

This narrative review critically synthesized contemporary ethical and legal challenges associated with the implementation of AI across healthcare systems. The review was designed as a structured narrative review rather than a systematic review or meta-analysis. Accordingly, its objective was not the exhaustive retrieval of all published studies, but the identification and critical synthesis of conceptually influential, regulatorily authoritative, and clinically relevant literature.

We searched the following electronic databases: PubMed/MEDLINE, Scopus, and Google Scholar using a Boolean search strategy developed across four conceptual domains: AI technologies, healthcare context, ethical principles, and legal/regulatory governance. The structured database search focused on literature published between January 2021 and February 2026. To ensure conceptual completeness, seminal earlier publications identified through hand-searching and backward reference-list screening were also incorporated when they were considered foundational to the ethical, legal, and governance discourse on AI in healthcare.

The Boolean Search Syntax Used was as Follows

("Artificial intelligence" OR "machine learning" OR "deep learning" OR "clinical decision support" OR "generative AI" OR "large language model") AND (healthcare OR medicine OR clinical) AND (ethics OR fairness OR bias OR accountability OR transparency OR explainability OR privacy OR justice) AND (law OR regulation OR governance OR liability OR "risk management" OR "medical device regulation" OR "malpractice").

Inclusion criteria comprised peer-reviewed original research, policy analyses, regulatory frameworks, consensus reports, and governance guidelines addressing the ethical or legal implications of AI in healthcare delivery. Articles limited exclusively to technical algorithm development without substantial ethical or regulatory discussion were excluded. To ensure the currency and authority of the synthesized material, particular emphasis was placed on literature published in high-impact, peer-reviewed journals and on documents issued by authoritative international bodies, including the World Health Organization (WHO), the Organization for Economic Co-operation and Development (OECD), the European Commission, and the National Institute of Standards and Technology (NIST). The reference lists of eligible publications were additionally screened to identify further relevant institutional documents. Consistent with narrative review methodology, conceptual relevance, regulatory authority, and contemporary applicability were prioritized over quantitative synthesis. No formal preferred reporting items for systematic reviews and meta-analyses flow diagram, duplicate screening workflow, or risk-of-bias assessment tool was applied, because the review was not designed as a systematic evidence synthesis. This methodological choice was deliberate: The aim was not to quantify effect sizes or aggregate study outcomes, but to critically synthesize

conceptually influential and regulatorily authoritative sources across a rapidly evolving field where randomized controlled trials remain scarce.

Accordingly, this review synthesizes selected literature and regulatory materials to examine the core ethical principles, legal accountability structures, and emerging governance approaches guiding responsible AI integration into healthcare systems.

Ethical and Regulatory Analysis of AI in Healthcare

The foundational bioethical principles – autonomy, beneficence, non-maleficence, and justice – remain central to medical practice, yet their practical meaning shifts when clinical decisions are influenced by algorithmic systems. Autonomy, for instance, becomes difficult to safeguard when patients are asked to consent to recommendations generated by models whose internal logic may be inaccessible even to clinicians. In such contexts, informed consent risks becoming a procedural formality rather than a meaningful exchange of information. Respecting autonomy in algorithm-mediated care, therefore, requires more than technical accuracy; it demands transparency about the role of AI, honest communication about uncertainty, and the preservation of a human decision-maker capable of contextualizing algorithmic outputs within the patient's personal values and clinical circumstances. WHO guidance reinforces this perspective by emphasizing that humans must remain in control of medical decisions, rather than delegating authority to opaque systems.[2]

The principles of beneficence and non-maleficence similarly acquire new dimensions. AI systems are often justified on the grounds that they improve efficiency or diagnostic accuracy, yet these benefits must be weighed against new categories of harm. Dataset shift, limited external validity, spurious correlations, and automation bias can all lead to inappropriate clinical decisions. In particular, automation bias may encourage clinicians to defer to algorithmic outputs even when these conflict with clinical intuition. Another emerging concern is alert fatigue, where excessive algorithm-generated warnings may desensitize clinicians and increase the likelihood of critical oversights. Such risks become especially pronounced when models are deployed in populations or environments different from those used in training, or when performance deteriorates over time. A safety-oriented ethical approach, therefore, requires continuous evaluation, real-world validation, and clear mechanisms for retraining or withdrawal. WHO regulatory considerations highlight safety, effectiveness, and stakeholder dialogue as key elements in balancing risks and benefits.[1]

Justice and fairness are perhaps the most complex ethical challenges in AI-mediated healthcare. Clinical data often reflect longstanding disparities in access to care, diagnosis, and treatment. When training datasets fail to adequately represent certain populations, models may systematically underperform for those groups, thereby reinforcing existing inequities. Contemporary analyses have already documented such patterns in clinical AI systems, with potential implications for unequal treatment outcomes.[7] This concern is strongly supported by empirical evidence demonstrating that widely used healthcare algorithms may systematically disadvantage certain populations; for example, a large-scale study revealed that a commonly used risk prediction algorithm significantly underestimated the health needs of Black patients due to biased proxy variables.[8] Such findings illustrate that algorithmic bias is not merely theoretical but a measurable, clinically relevant source of inequity. It is therefore important to distinguish between algorithmic bias, which originates from technical flaws in data or model design, and the broader concept of fairness, which concerns the just distribution of healthcare benefits and burdens. Bias may arise not only from incomplete datasets but also from proxy variables, measurement errors, or structural inequities embedded in healthcare systems. Ethical deployment of AI thus requires subgroup performance evaluation, transparent reporting of limitations, and proactive mitigation strategies. WHO guidance explicitly identifies inclusiveness and equity as core ethical imperatives.[2]

Transparency and explainability occupy a central position at the intersection of ethics, law, and clinical practice. Although black-box models may achieve impressive predictive accuracy, their lack of interpretability can undermine shared decision-making and weaken professional accountability. Clinicians cannot reasonably justify decisions they do not understand, nor can patients provide meaningful consent in the absence of intelligible explanations. From a legal perspective, this challenge aligns closely with emerging data protection principles, such as the "right to explanation," which restricts fully automated decision-making and requires individuals to obtain meaningful information about algorithmic outputs.[9] This reinforces the necessity of interpretability not only as an ethical requirement but also as a developing legal standard. Trustworthy AI frameworks, therefore, identify transparency, explainability, privacy, fairness, and accountability as essential dimensions for safe implementation.[3] This emphasis is echoed in governance initiatives such as the NIST AI risk management

framework, which characterizes trustworthy AI through attributes including validity, reliability, safety, transparency, explainability, accountability, and privacy.[10]

Data privacy and confidentiality have long been central ethical obligations in healthcare, but the data-intensive nature of AI introduces additional complexities. The development and deployment of AI systems typically require large volumes of clinical and imaging data, both of which are highly sensitive and subject to strict regulatory protections. Healthcare institutions must therefore navigate complex legal frameworks designed to safeguard personal health information and ensure responsible data use.

The increasing reliance on cloud-based infrastructures and cross-border data transfers further complicates this landscape by introducing jurisdictional challenges and potential regulatory conflicts.[2,3,11] Legal instruments such as the Health Insurance Portability and Accountability Act and the general data protection regulation (GDPR) establish safeguards for personal health information, mandate lawful data processing, and restrict certain forms of automated decision-making. Ethical oversight mechanisms, including Institutional Review Boards, also play a crucial role in assessing risks, ensuring informed consent, and protecting privacy and fairness in AI-related healthcare research.[11] Ethical data governance, therefore, requires data minimization, clear purpose limitation, secure storage, and transparency regarding secondary data use. WHO guidance warns that inadequate safeguards may erode public trust and ultimately result in patient harm.[1,2]

Legal Accountability and Regulatory Frameworks: International and National Perspectives

Questions of legal accountability represent another major challenge. AI systems diffuse causal responsibility across multiple actors, including developers, vendors, and healthcare institutions, clinicians, and data providers. When harm occurs, identifying the responsible party becomes significantly more complex than in traditional malpractice scenarios. Conventional liability models focus primarily on clinician negligence, yet AI-related harm may stem from flawed model design, insufficient validation, poor integration into clinical workflows, or inadequate performance monitoring. A pragmatic approach is therefore to conceptualize accountability as shared but clearly delineated, with each stakeholder assuming responsibilities appropriate to their role. This perspective aligns with the OECD principles on AI accountability.[12,13]

International Regulatory Frameworks

Regulatory frameworks are evolving in response to these challenges. In the EU, the Artificial Intelligence Act (AI Act) entered into force in August 2024, introducing a risk-based approach to AI regulation. Its application is phased, with some provisions taking effect from February 2025 and the main body of obligations from August 2026. Many medical AI applications are classified as high-risk systems, subject to stringent requirements concerning training data quality, conformity assessments, risk management, transparency obligations, and human oversight.[14,15]

Additional guidance clarifies how the AI Act interacts with existing medical device regulations, underscoring the need for coordinated compliance across regulatory regimes.[16] The interplay between the EU AI Act and established medical device regulatory frameworks further complicates compliance processes by introducing overlapping obligations related to risk classification, data governance, and post-market monitoring, highlighting emerging challenges of regulatory fragmentation in AI-enabled healthcare systems.[17] The phased implementation timeline further highlights the importance of aligning institutional deployment strategies with regulatory milestones.

In the US, regulatory oversight largely depends on whether a software function is classified as a medical device. The Food and Drug Administration (FDA) guidance on clinical decision-support software distinguishes between exempt and regulated systems based on functionality and transparency criteria.[18] Meanwhile, the concept of Software as a Medical Device, developed by the International Medical Device Regulators Forum, provides a widely accepted framework for risk categorization and regulatory approaches.[19] Ethical oversight bodies, such as Institutional Review Boards, continue to play a central role in ensuring that AI research involving human data complies with standards for informed consent, privacy protection, and risk minimization.[11]

National Regulatory Perspective: The Case of Türkiye

While the majority of international governance discourse focuses on the EU and the US, a comprehensive legal analysis requires examination of national frameworks in other jurisdictions. Türkiye represents an instructive case, as it has developed a multi-layered regulatory approach to AI in healthcare without yet adopting a comprehensive AI-specific law comparable to the EU AI Act.[20–23]

First, data protection constitutes the foundational layer. Türkiye's Personal Data Protection Law No. 6698, enacted

in 2016 and substantially amended in 2024 to align more closely with GDPR, governs the processing of personal data, including health data.[21,23,24] Under Article 6 of the Personal Data Protection Law No. 6698, health data are classified as special categories of personal data and are subject to heightened protection.[24] For AI development, this means that training datasets derived from electronic health records or other clinical repositories require either patient consent or a lawful processing basis under the law. Importantly, Article 11 of the same law grants data subjects the right to object to a result arising against them through analysis performed solely by automated systems; however, the law does not establish a fully articulated AI-specific "right to explanation" comparable to broader debates under the GDPR framework.[21,23,24]

Second, medical device regulation provides the product safety layer. The Turkish Medicines and Medical Devices Agency has adopted the Medical Device Regulation (MDR) (harmonized with EU 2017/745), which defines software used for medical purposes as a medical device.[21–23,25] Consequently, AI-based software intended for diagnostic or therapeutic purposes is classified as an active medical device.[21–23,25] The risk classification rules (Annex VIII of the MDR) categorize most clinical AI applications as Class IIa or IIb devices, which require conformity assessment by notified bodies.[21,23,25] However, unlike the EU AI Act, Türkiye's MDR does not yet impose specific requirements for training data quality, bias mitigation, or human oversight beyond general safety and performance requirements.[21–23]

Third, digital health infrastructure is governed by the Regulation on the Provision of Remote Healthcare Services.[26] This regulation primarily structures the legal framework for remote healthcare delivery and related digital service processes. Although it is relevant to the broader digital health ecosystem, it does not expressly provide a dedicated framework for AI model development, algorithm validation, or post-deployment monitoring. This creates uncertainty for developers and institutions seeking to access and use real-world clinical data for AI innovation within a clearly defined regulatory pathway.[21,23,26]

Recent Turkish scholarship has also emphasized that AI applications in healthcare - particularly in high-risk contexts such as pre-hospital emergency services- raise unresolved legal questions regarding malpractice liability, patient autonomy, and the protection of sensitive health data.[27] These analyses underscore that, despite existing regulatory instruments, there remains a need for more detailed, AI-specific legal guidance addressing autonomous or semi-autonomous clinical decision-making systems.

Fourth, liability and malpractice law in Türkiye has not yet specifically addressed AI-related harm. Under the Turkish Code of Obligations (Law No. 6098), medical malpractice liability is primarily fault-based.[20,21,23,28] In cases involving AI, potential liable parties include: The clinician (for over-reliance on erroneous AI output), the hospital (for inadequate validation or training), and the AI developer (for product defect).[20,21,23,28] However, no binding Turkish court precedent has yet established apportionment of liability in AI-related medical injury.[20,21,23] Legal scholars have noted this gap, calling for either legislative clarification or judicial development of AI-specific liability rules.[20,21,23]

Table 1 provides a comparative overview of regulatory approaches to AI in healthcare across the EU, the US, and Türkiye.

Human Oversight as a Cross-Cutting Principle

Across ethical guidelines and regulatory frameworks, human oversight remains a consistent and central principle. AI systems are intended to assist, rather than replace, clinical decision-making. However, meaningful oversight requires more than nominal human involvement. Clinicians must be adequately trained, workflows must be transparent, and systems must include mechanisms for auditing and review. Trustworthy AI frameworks emphasize human-in-the-loop approaches, auditability, and traceability as essential governance tools.[3] European regulatory analyses also suggest that healthcare institutions and clinicians deploying high-risk AI systems may bear specific compliance obligations, including monitoring and documentation responsibilities.[14,16,22]

Implementation Realities: Case-Based Analysis of Clinical Deployment

While normative, ethical, and legal frameworks are essential, the title's promise of "implementation realities" requires examination of actual clinical deployments. This section analyzes two real-world cases – one successful and one problematic – to derive concrete governance lessons.

Case 1: Successful Deployment – Sepsis Prediction at Duke University Health System

Sendak et al.[29] (2020) documented the integration of a deep learning-based sepsis prediction model (the "Sepsis Watch") into routine clinical care at Duke University Health System. The model was designed to identify patients at risk of sepsis up to 12 h before clinical recognition. Key governance elements that contributed to success included:

Table 1. Comparative regulatory approaches to AI in healthcare

Dimension	European Union (AI Act)	United States (FDA)	Türkiye (Personal Data Protection Law No. 6698 + Turkish Medicines and Medical Devices Agency)
Risk classification	High-risk where AI is itself a regulated medical device or a safety component of one, generally requiring third-party conformity assessment	Risk-based; classified as medical device if intended for diagnosis/treatment; pathways: 510(k), De Novo, or Premarket Approval	Class IIa/IIb under MDR (harmonized with EU); no AI-specific statutory risk tier yet
Transparency obligations	Article 13–technical documentation and user information	CDS transparency is relevant to whether the software remains outside the device definition or becomes subject to FDA oversight	Article 11 of the Personal Data Protection Law No. 6698–right to object to automated processing outcomes; no explicit "right to explanation."
Human oversight	Mandatory for high-risk systems (Article 14)	Context-dependent and function-specific; not framed as a horizontal AI-law obligation comparable to the EU AI Act	Not explicitly required in MDR or Personal Data Protection Law No. 6698
Data quality requirements	Training/validation data must be relevant, representative, and bias-free	General software validation and quality system expectations	General data protection and medical device safety obligations; no AI-specific statutory dataset quality requirements
Liability/ malpractice	No standalone AI-specific civil liability regime currently in force; existing product liability and national tort frameworks remain relevant. The proposed revision of the EU Product Liability Directive (2024) explicitly includes software and AI systems, which may reshape liability allocation for medical AI.	Traditional malpractice+product liability	Turkish Code of Obligations No. 6098 (fault-based); no AI-specific judicial precedent; scholarly debate ongoing.[15,16,18]
Post-market monitoring	Mandatory for high-risk systems	Post-market obligations arise through medical device oversight, where applicable	General vigilance under MDR; not AI-specific

SaMD: Software as a medical device; CDS: Clinical decision support; MDR: Medical device regulation; AI: Artificial intelligence; MDR: Medical device regulation; EU: European Union; FDA: Food and Drug Administration.

- Prospective validation: The model was validated on local data before deployment, addressing dataset shift concerns.
- Clinical workflow integration: A dedicated rapid response team received alerts, rather than adding notifications to already-overburdened physicians.
- Human-in-the-loop design: Nurses reviewed each alert before team activation, maintaining human judgment as a filter.
- Continuous monitoring: Performance was tracked weekly, with predefined thresholds for retraining or suspension.
- Clinician training: All users completed simulation-based training on AI limitations and appropriate override conditions.

From a legal accountability perspective, Duke established a clear governance structure: The health system assumed responsibility for model validation and monitoring, while clinicians remained accountable for final treatment decisions. This delineation reduced ambiguity about liability in the event of adverse outcomes.

Case 2: Problematic Deployment – Proprietary Sepsis Algorithm (Epic)

In contrast, Wong et al.[30] (2021) published an external validation study of Epic Systems' proprietary sepsis prediction model (the Epic Sepsis Model, ESM), which was deployed at hundreds of US hospitals. The study found that the model performed poorly on external validation, with an area under the receiver operating characteristic curve of 0.60–0.64, only marginally better than chance. More concerning, the model generated alerts for only 7% of sepsis patients while producing high false-positive rates.

Table 2. Comparative governance lessons from real-world AI deployment cases

Governance function	Successful (Duke Sepsis Watch)	Problematic (Epic ESM)
Local validation	Mandatory before deployment	Often skipped or superficial
Transparency	Model details shared with users	Proprietary, black-box
Human oversight	Nurse filter + team activation	Direct alert to physician
Performance monitoring	Weekly, with suspension thresholds	None documented
Liability clarity	Defined (hospital + clinician)	Ambiguous (vendor vs. hospital vs. clinician)

This table is derived from the case analyses presented in the Implementation Realities section.

Ethical Failures Identified Included

- Lack of transparency: Epic did not disclose detailed model architecture or validation data, preventing independent assessment.
- Potential reinforcement of automation bias: Clinicians at deploying hospitals reported feeling pressured to respond to alerts despite low positive predictive value.
- Inadequate local validation: Many hospitals deployed the model without rigorous local performance testing.
- Absence of post-market surveillance: No systematic tracking of missed sepsis cases or false alarm harms.

Legal implications from this case include potential product liability claims against Epic (for defective software) and institutional negligence claims against hospitals (for failing to validate before deployment). The case illustrates how proprietary black-box models can undermine the ethical principle of transparency and create diffuse accountability. Table 2 provides a comparative summary of governance lessons derived from the Duke Sepsis Watch.[29] and the Epic Sepsis Model[30] case analyses. Together, these cases demonstrate that responsible AI governance is not merely about compliance with regulatory checklists, but about embedding continuous evaluation, transparency, and human judgment into clinical workflows.

Special Populations and Clinical Domains

Certain patient groups and clinical domains raise additional ethical concerns. Pediatric populations, for example, rely on proxy consent and are frequently under-represented in training datasets, increasing the likelihood of diagnostic inaccuracies.[2,4,31]

Rare diseases and genomic medicine pose additional challenges, including the difficulty of scaling phenotype data collection and the presence of complex phenotypic overlap across distinct syndromes, which may complicate accurate diagnosis.[32] Emerging applications of AI in genomic diagnostics highlight both the potential to improve diagnostic performance and the importance of carefully addressing limitations in data quality, bias, and generalizability.[32]

In mental health contexts, the use of AI for suicide risk prediction or sentiment analysis introduces complex questions about autonomy, stigma, and the therapeutic relationship.[2,4] Recent literature also emphasizes that algorithmic predictions in psychiatry may lack contextual interpretability and risk reinforcing existing biases, particularly when trained on non-representative behavioral or social data.[33] In such settings, ethical governance requires cautious deployment, a clearly defined scope, and enhanced monitoring mechanisms.

A Lifecycle Governance Model for Responsible AI Implementation

Responsible AI implementation should be understood as a continuous lifecycle rather than a one-time technical intervention. This lifecycle includes ethics-by-design, rigorous validation, ongoing monitoring, and strong institutional governance. Trustworthy AI frameworks propose measurable dimensions – such as fairness, robustness, privacy, explainability, and accountability – to guide this process.[3]

In parallel, governance-oriented studies have proposed structured frameworks that operationalize these principles through lifecycle-based oversight, emphasizing bias mitigation, data governance, privacy protection, and accountability as core implementation domains.[34] Complementary analyses further highlight that trustworthiness in medical AI is shaped by interconnected factors, including data quality, algorithmic opacity, safety, and responsibility attribution, all of which require coordinated ethical and regulatory oversight.[35]

Figure 1 presents a synthesized governance model derived from the ethical principles, regulatory requirements, and implementation case analyses discussed above.

Practical implementation requires risk management strategies, thorough documentation, subgroup

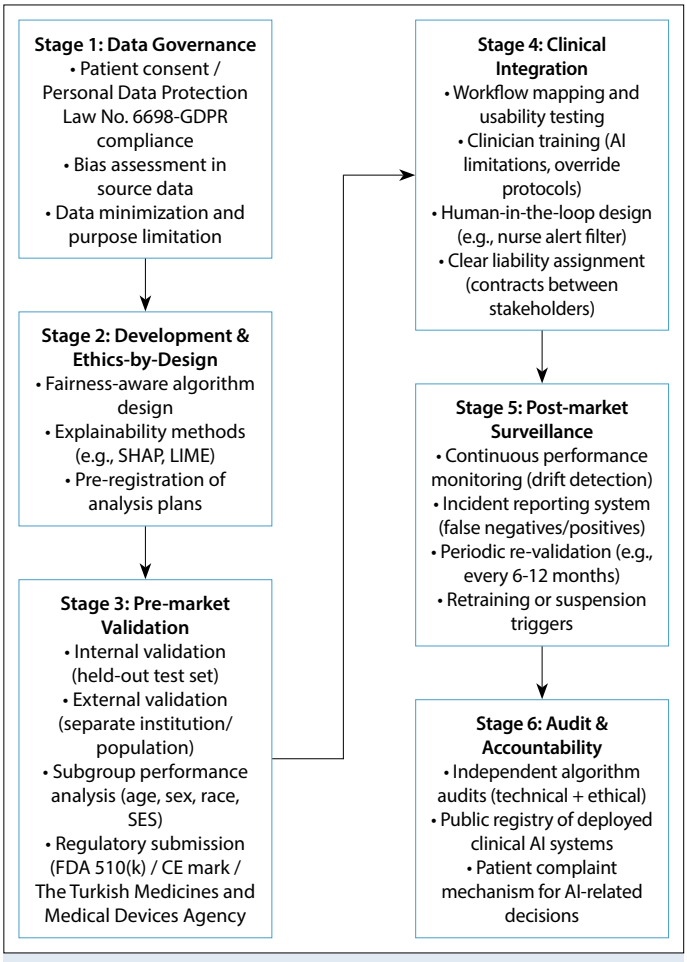


Figure 1. Lifecycle governance model for AI in healthcare. This figure presents a synthesized lifecycle-based governance model for artificial intelligence in healthcare. Derived from the ethical principles, regulatory frameworks, and real-world implementation challenges discussed in this review, the model conceptualizes AI oversight across six sequential yet interdependent stages: Data Governance, Development and Ethics-by-Design, Pre-market Validation, Clinical Integration, Post-market Surveillance, and Audit and Accountability. The framework emphasizes that safe and equitable clinical AI requires continuous feedback between technical validation, regulatory compliance, human oversight, and post-deployment accountability.

AI: Artificial intelligence; FDA: U.S. Food and Drug Administration; GDPR: General data protection regulation; LIME: Local interpretable model-agnostic explanations; SES: Socioeconomic status, SHAP: Shapley additive explanations; WHO: World Health Organization.

performance audits, user training, and structured incident-reporting mechanisms. Frameworks such as the NIST AI risk management framework provide practical guidance for institutional governance, while binding regulations such as the EU AI Act establish minimum compliance obligations.[10,14,16] Ultimately, AI should serve as a tool that enhances, rather than replaces, clinical judgment, while preserving patient autonomy, equity, and accountability in healthcare delivery.[1,2]

Clinical Implications for Healthcare Professionals and Institutions

Beyond theoretical ethical principles, AI-related risks materialize within everyday clinical workflows and institutional decision structures. Healthcare professionals increasingly interact with probabilistic algorithmic outputs that may influence diagnosis, treatment prioritization, or risk prediction. Ethical implementation, therefore, requires explicit documentation of whether AI-generated recommendations contributed to clinical decisions and how disagreements between clinician judgment and algorithmic output were resolved.

From a legal perspective, institutions deploying AI systems should establish clearly delineated accountability frameworks involving developers, vendors, healthcare organizations, and clinical end-users.[13] Based on the case analyses presented above, specific actionable recommendations include:

- For clinicians: Document when AI recommendations are overridden or followed; complete mandatory training on AI limitations and automation bias; report all AI-related adverse events or near-misses.[2,16]
- For institutions: Establish an AI governance committee with clinical, legal, and technical membership; mandate local validation before any clinical deployment; implement continuous performance dashboards with predefined suspension thresholds; secure liability agreements with AI vendors.[16,29]
- For developers: Provide model cards[36] detailing training data, performance subgroups, and intended use limitations; enable audit logging of all predictions; support explainability methods accessible to clinicians.[16]

In the Turkish context, institutions should additionally ensure compliance with Article 6 of the Personal Data Protection Law No. 6698 for health data processing, register AI-based medical devices with the Turkish Medicines and Medical Devices Agency, and consider obtaining legal opinions on liability allocation given the absence of AI-specific jurisprudence.[20,21]

Human oversight should not be interpreted as passive supervision but rather as an operational governance function supported by clinician training, institutional AI committees, and continuous performance evaluation.[2] Such measures reduce automation bias, maintain professional accountability, and align healthcare deployment practices with emerging international regulatory expectations.[14,16,17]

Limitations

This narrative review has several limitations that should be acknowledged. First, the selection of sources involves inherent subjectivity; despite a structured search strategy with predefined inclusion criteria, relevant studies – particularly those reporting null or negative findings – may have been omitted. Second, the analysis predominantly reflects regulatory frameworks and implementation cases from the EU and the US, with only initial exploration of the Turkish context and limited coverage of other regions (e.g., Asia, Latin America, Africa). Third, the inclusion of non-peer-reviewed institutional documents (e.g., WHO guidance, NIST frameworks, national regulations) – while essential for capturing current regulatory thinking and authoritative governance standards – means that some sources have not undergone traditional peer review. Where possible, priority was given to documents with transparent development processes and public consultation mechanisms. Fourth, the rapidly changing legal landscape (e.g., phased implementation of the EU AI Act, ongoing development of Turkish digital health governance, and amendments to the Personal Data Protection Law No. 6698 means that some regulatory details may require updating. Fifth, the narrative review format does not permit quantitative synthesis or meta-analysis of implementation outcomes. Sixth, because the review was not conducted as a systematic review, no formal study-quality appraisal or evidence-certainty assessment was undertaken. Finally, the exclusion of non-English publications (except for Turkish regulatory documents) may have introduced language bias. Despite these limitations, the synthesis provides a structured foundation for understanding the governance challenges of AI in healthcare and identifying priority areas for future empirical research.

Conclusion

AI is rapidly transforming healthcare by improving diagnostic accuracy, operational efficiency, and the personalization of care. Yet its integration into clinical practice also introduces complex ethical and legal challenges related to autonomy, fairness, transparency, privacy, and accountability. Addressing these challenges requires not only robust ethical frameworks but also clearly defined liability structures and comprehensive regulatory oversight.

International initiatives – including WHO guidance, the EU AI Act, and FDA regulatory approaches – reflect a growing global consensus that AI in healthcare must be governed responsibly. As this review demonstrates through a comparative analysis of EU, US, and Turkish frameworks, national regulatory contexts differ substantially, creating challenges for multinational deployment and underscoring the need for context-sensitive governance. The analysis of real-world implementation cases (Duke Sepsis Watch vs. Epic Sepsis Model) further reveals that responsible governance requires not only regulatory compliance but also local validation, transparency, human oversight, and continuous monitoring. Specifically, Turkish policymakers and healthcare institutions should prioritize: (a) Developing binding guidance on how Article 11 of the Personal Data Protection Law No. 6698 ("right to object to automated outcomes") applies in clinical settings; (b) clarifying the Turkish Medicines and Medical Devices Agency's oversight role for AI as a medical device beyond initial conformity assessment; and (c) establishing a national AI incident reporting system for healthcare to generate the empirical evidence needed for future liability frameworks. The effectiveness of these efforts will depend not solely on regulatory compliance but also on fostering a culture of ethical awareness and interdisciplinary collaboration among clinicians, regulators, developers, and patients. Ultimately, AI should function as a tool that augments, rather than replaces, human clinical judgment while preserving the fundamental ethical principles of medical practice.

Future research should therefore move beyond normative ethical discussions toward empirically grounded governance models capable of evaluating real-world AI performance, safety, and accountability within clinical environments. Priority areas include: (1) Prospective studies of AI implementation measuring both clinical outcomes and unintended harms; (2) legal analyses of liability allocation in actual malpractice cases involving AI; (3) comparative effectiveness research on different governance models (e.g., centralized vs. distributed oversight); and (4) development of validated audit tools for algorithmic fairness in healthcare contexts, including in middle-income countries like Türkiye.

Ethics Committee Approval: Ethical approval was not required for this study since this is a review article.

Conflict of Interest: None declared.

Financial Disclosure: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Use of AI for Writing Assistance: During the preparation of this work, the author used artificial intelligence for language editing and grammar checking. After using this, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

Peer-review: Double blind peer-reviewed.

References

1. World Health Organization. WHO outlines considerations for regulation of artificial intelligence for health. Available at: https://iris.who.int/server/api/core/bitstreams/ad62580f-540f-4e36-b957-e7f2946ae1fb/content Accessed 12 Feb, 2026.
2. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021. Available at: https://www.who.int/publications/i/item/9789240029200 Accessed 12 Feb, 2026.
3. Ahadian P, Xu W, Liu D, Guan Q. Ethics of trustworthy AI in healthcare: challenges, principles, and practical pathways. Neurocomputing 2026;661:131942. [CrossRef]
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [CrossRef]
5. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1(1):389-99. [CrossRef]
6. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harv Data Sci Rev 2019;1(1):1-14. [CrossRef]
7. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28(1):31-8. [CrossRef]
8. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-53. [CrossRef]
9. Goodman B, Flaxman S. European Union regulations on algorithmic decision making and a "right to explanation". AI Mag 2017;38(3):50-7. [CrossRef]
10. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg (MD): U.S. Department of Commerce; 2023. Available at: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf Accessed 12 Feb, 2026.
11. Pantanowitz L, Hanna M, Pantanowitz J, Lennerz J, Henricks WH, Shen P, et al. Regulatory aspects of artificial intelligence and machine learning. Mod Pathol 2024;37(12):100609. [CrossRef]
12. Organisation for Economic Co-operation and Development. Recommendation of the Council on Artificial Intelligence. 2019. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 Accessed 12 Feb, 2026.
13. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans A Math Phys Eng Sci 2018;376(2133):20180080. [CrossRef]
14. European Commission. AI Act enters into force. Available at: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en Accessed 12 Feb, 2026.
15. European Commission, Directorate-General for Health and Food Safety. Artificial intelligence in healthcare (EU eHealth and digital health & care). Available at: https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_en Accessed 12 Feb, 2026.
16. Medical Device Coordination Group. MDCG 2025-6: Interplay between the Medical Devices Regulations (MDR/IVDR) and the Artificial Intelligence Act (AIA) 2025. Available at: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en Accessed 12 Feb, 2026.
17. Kalodanis K, Feretzakis G, Rizomiliotis P, Verykios VS, Papapavlou C, Skrekas A, et al. Evaluating the impact of the EU AI act on medical device regulation. Stud Health Technol Inform 2025;323:40-4. [CrossRef]
18. U.S. Food and Drug Administration. Clinical decision support software: guidance for industry and Food and Drug Administration staff. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software Accessed 12 Feb, 2026.
19. International Medical Device Regulators Forum. Software as a medical device (SaMD): key definitions. Available at: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf Accessed 12 Feb, 2026.
20. Diri F. Yapay zeka teknolojisi ve beraberinde getirdiklerinin Türk sağlık hukuku kapsamında değerlendirilmesi. Bilişim Hukuku Dergisi 2024;6(1):270-320. [Article in Turkish] [CrossRef]
21. Yüzbaşıoğlu C. The use of artificial intelligence by the administration in health services and the resulting liability in Turkish law. J Acad Res Med 2025;15(3):109-11. [CrossRef]
22. Veale M, Zuiderveen Borgesius F. Demystifying the Draft EU Artificial Intelligence Act - analysing the good, the bad, and the unclear elements of the proposed approach. Comput Law Rev Int 2021;22(4):97-112. [CrossRef]
23. Günday HM, Albayrak Günday E. Artificial intelligence in mental health practices: legal liability analysis under Turkish, European, and common law frameworks. Psikiyatride Güncel Yaklaşımlar 2025;17:806-21. [CrossRef]
24. Republic of Türkiye. Personal Data Protection Law No. 6698. Official Gazette No. 29677. Available at: https://mevzuat.gov.tr/mevzuat?MevzuatNo=6698&MevzuatTur=1&MevzuatTertip=5 Accessed 12 Feb, 2026.
25. Republic of Türkiye. Medical Device Regulation. Official Gazette No. 31499 (Reprint). Available at: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=38657&MevzuatTur=7&MevzuatTertip=5 Accessed 12 Feb, 2026. [in Turkish]
26. Ministry of Health of the Republic of Türkiye. Regulation on the provision of remote healthcare services. Official Gazette No. 31746, 10 February 2022. [cited 2026 Apr 02]. Available at: https://www.resmigazete.gov.tr/eskiler/2022/02/20220210-2.htm [in Turkish]
27. Şişli Z. Yapay zeka alanında hukuki gelişmeler ve hastane öncesi sağlık hizmetleri. Medical Technologies Congress 2024:117-20. [Article in Turkish]
28. Republic of Türkiye. Turkish Code of Obligations No. 6098. Official Gazette No. 27836. Available at: https://resmigazete.gov.tr/eskiler/2011/02/20110204-1.htm [Article in Turkish]
29. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020;8(7):e15182. [CrossRef]

30. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021;181(8):1065-70. [CrossRef]
31. Hua SBZ, Heller N, He P, Towbin AJ, Chen IY, Lu AX, et al. Underrepresentation of children in public medical imaging datasets. Nature Health 2026. [CrossRef]
32. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med 2019;11(1):70. [CrossRef]
33. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019;21(11):116. [CrossRef]
34. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020;27(3):491-7. [CrossRef]
35. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023;23(1):7. [CrossRef]
36. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019; Atlanta, GA, USA. New York: ACM; 2019. p. 220-9. [CrossRef]