World Journal of Pharmaceutical Science and Research

www.wjpsronline.com

Review Article

ISSN: 2583-6579
SJIF Impact Factor: 6.916
Year - 2026
Volume: 5; Issue: 7
Page: 303-321

GENERATIVE ARTIFICIAL INTELLIGENCE IN DIGITAL HEALTH: CLINICAL TRANSLATION, SAFETY GOVERNANCE, REGULATORY SCIENCE, AND IMPLEMENTATION ROADMAP

Md Gulzar Ahmad*[1], Hasibullah Khairi[2], Ravi Dhakad[3], Prem Shankar[4], Md Faiyaz[5], Alisha Singh[6]

[1]Pharmacist Company - Apollo Pharmacies Limited, New Delhi, India
[2]Pharm D student, ISF College of Pharmacy, Moga, Punjab, India.
[3,4]PG student, Department of Pharmacy Practice, ISF College of Pharmacy, Moga, Punjab, India.
[5]PG student, Department of Pharmacology, Faculty of Pharmacy, Uttar Pradesh University of Medical Sciences, Saifai, Etawah, U.P, India.
[6]Assistant Professor, SNS College of Pharmacy, Motihari, Bihar, India.

Article Received: 19 May 2026 | | Article Revised: 09 June 2026 | | Article Accepted: 30 June 2026

***Corresponding Author: Md Gulzar Ahmad**
Pharmacist Company - Apollo Pharmacies Limited, New Delhi, India
DOI: https://doi.org/10.5281/zenodo.21166928

How to cite this Article: Md Gulzar Ahmad, Hasibullah Khairi, Ravi Dhakad, Prem Shankar, Md Faiyaz, Alisha Singh (2026) GENERATIVE ARTIFICIAL INTELLIGENCE IN DIGITAL HEALTH: CLINICAL TRANSLATION, SAFETY GOVERNANCE, REGULATORY SCIENCE, AND IMPLEMENTATION ROADMAP. World Journal of Pharmaceutical Science and Research, 5(7), 303-321.

ABSTRACT

Background: Generative artificial intelligence (AI), including large language models and multimodal systems, is increasingly evaluated in digital health for clinical documentation, information synthesis, decision support, diagnostic assistance, and workflow optimization. Despite rapid technical progress, safe clinical translation remains limited by unresolved challenges related to clinical validity, real-world performance, workflow integration, safety monitoring, regulatory readiness, ethical governance, and accountability. **Objective:** This narrative review proposes a translational regulatory-science framework for the responsible integration of generative AI into digital health systems by connecting clinical use cases, generative AI-specific safety risks, task-risk stratification, global regulatory expectations, ethical governance, multimodal accountability, lifecycle monitoring, and de-implementation readiness. **Review Approach:** A structured literature review was conducted of peer-reviewed studies, clinical AI reporting guidelines, implementation science literature, and official regulatory or governance documents published between 2000 and 2026. The review prioritized high-impact biomedical, digital health, regulatory science, and AI governance sources, including guidance from the World Health Organization, the U.S. Food and Drug Administration, European Union authorities, Health Canada, and the International Medical Device Regulators Forum. **Main Synthesis:** Generative AI can support digital health only when deployed through a lifecycle-based system that integrates intended-use definition, clinical validation, workflow assessment, human oversight, equity evaluation, privacy and cybersecurity safeguards, post-deployment monitoring, controlled model updating, and withdrawal pathways when safety or clinical value is not sustained. **Framework Contribution:** The proposed framework aligns clinical application categories with task-risk profiles, maps generative AI-specific risks to validation and governance requirements, and links regulatory expectations with accountability structures involving developers, healthcare organizations, clinicians, regulators, and patients. **Conclusion:** Trustworthy implementation of generative AI in digital health requires technical robustness, ethical oversight, regulatory compliance, real-world monitoring, and accountable governance, rather than relying on model capabilities alone.

KEYWORDS: Generative artificial intelligence; Large language models; Digital health; Clinical validation; Regulatory science; Trustworthy AI; Multimodal accountability.

Highlights

- Generative AI in digital health requires lifecycle validation beyond benchmark performance.
- Clinical deployment needs human oversight, workflow integration, and real-world monitoring.
- Hallucination, bias, privacy risk, and agentic error cascades remain key safety concerns.
- Regulatory readiness depends on intended use, benefit-risk evidence, and controlled model updating.
- Trustworthy implementation requires multimodal accountability across developers, clinicians, regulators, and patients.

1. INTRODUCTION

1.1 Generative AI and the changing digital-health landscape

Generative artificial intelligence (AI) is increasingly positioned as a translational technology in digital health because it can generate, summarize, classify, retrieve, and transform clinical information across multiple healthcare workflows.[1]

Large language models (LLMs) are transformer-based systems capable of producing human-like text and supporting information retrieval, summarization, documentation, and conversational decision support in medicine.[2] Recent healthcare literature has evaluated LLMs for mental health applications, medication-related harm mitigation, triage, referral support, diagnostic reasoning, and clinical decision-support workflows.[3–5] Multimodal large language models extend these capabilities by integrating clinical text with imaging, structured records, and contextual patient information, a capability that is especially relevant for radiology and other data-rich specialties.[6,7] Agentic AI further expands the field by enabling autonomous planning, tool use, and multistep reasoning toward complex goals, although this autonomy increases safety and governance concerns.[8]

1.2 The clinical translation problem

The promise of generative AI cannot be inferred from technical capabilities alone, as its deployment in healthcare requires clinical validity, workflow fit, safety monitoring, and governance in real clinical settings.[9,10] Benchmark performance may not predict clinical utility when models are exposed to incomplete patient information, local workflow variation, demographic heterogeneity, and changing clinical environments.[11,12] Digital-health translation therefore requires evaluation across the full lifecycle, including intended-use definition, pre-deployment validation, prospective testing, implementation monitoring, post-deployment surveillance, and controlled update management.[1]

Safety risks include hallucination, prompt sensitivity, bias, privacy leakage, cybersecurity vulnerability, automation bias, and unclear accountability when outputs influence patient care.[2,9] These risks become more complex in multimodal and agentic systems because generated outputs may combine text, images, voice, and structured data in ways that are difficult for clinicians to quickly verify.[6–8]

1.3 Aim and novelty of this review

This review synthesizes clinical, ethical, regulatory, and implementation-science evidence to examine how generative AI can be translated into digital health without compromising safety, equity, or public trust.[1,9] It differs from technology-centered reviews by framing generative AI as a socio-technical healthcare intervention requiring validation, human oversight, lifecycle governance, regulatory alignment, and de-implementation readiness.[10–12] The review proposes a translational regulatory science framework that links intended use, evidence generation, ethical governance, multimodal accountability, controlled model updating, post-deployment monitoring, and implementation feedback.[1]

Figure 1 provides the conceptual orientation of the review by showing how generative AI capabilities must be filtered through clinical validation, regulatory science, and health-system governance before routine deployment.[9,12]

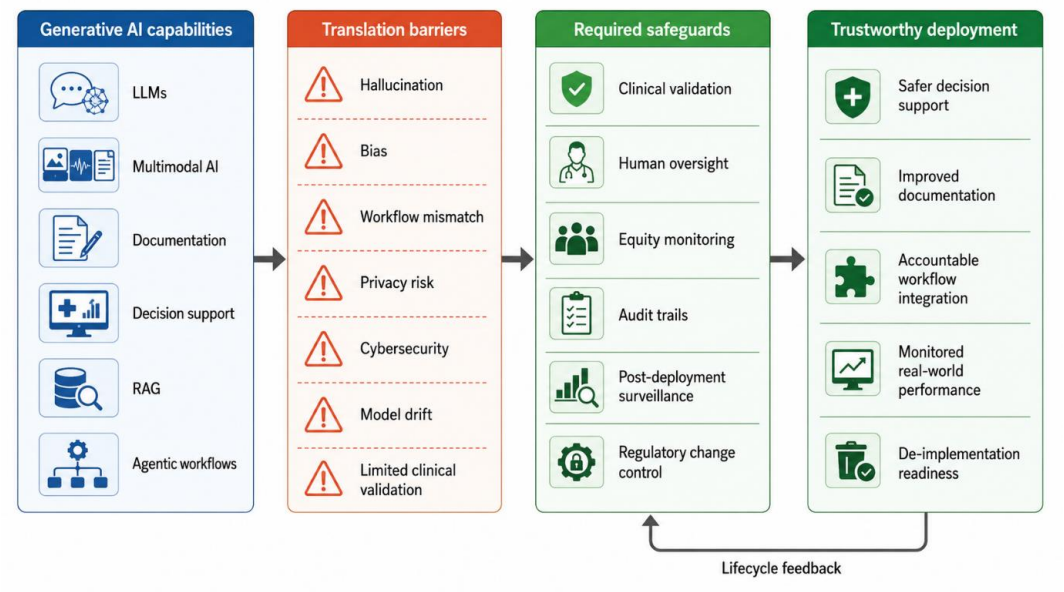

Figure 1: Conceptual overview of generative AI translation in digital health.

This figure summarizes the translational pathway from generative AI capabilities to clinically accountable digital-health deployment.[1,11] Generative AI functions, including LLMs, multimodal reasoning, documentation support, decision support, RAG, and agentic workflows, must pass through safeguards before clinical use.[2,6,8] Key barriers include hallucinations, bias, workflow mismatch, privacy risks, cybersecurity vulnerabilities, model drift, and insufficient clinical validation.[9,12] Trustworthy deployment requires clinical validation, human oversight, equity monitoring, post-deployment surveillance, regulatory change control, accountability, and de-implementation readiness.[1,12]

Abbreviations: AI, artificial intelligence; EHR, electronic health record; GenAI, generative artificial intelligence; LLM, large language model; MLLM, multimodal large language model; RAG, retrieval-augmented generation.

2. Review Approach and Evidence Mapping

This article uses a structured narrative and scoping-style review approach, consistent with PRISMA-ScR principles, to map heterogeneous evidence across clinical, ethical, implementation, and regulatory domains.[13] Methodological decisions were aligned with the updated JBI guidance for scoping reviews, as the included literature spans clinical studies, reporting guidelines, regulatory documents, implementation frameworks, and conceptual analyses.[14]

Searches were conducted in PubMed/MEDLINE, Scopus, Web of Science, Embase, Google Scholar, and official regulatory or institutional websites to identify peer-reviewed studies, reporting guidelines, regulatory documents, and implementation science literature relevant to generative AI in digital health.[13,14] Search terms included combinations of "generative artificial intelligence," "large language model," "multimodal AI," "digital health," "clinical validation," "clinical decision support," "hallucination," "bias," "privacy," "cybersecurity," "regulatory science," "implementation science," "post-deployment monitoring," and "trustworthy AI." Eligible sources included English-language peer-reviewed articles, reporting guidelines, regulatory guidance documents, and high-relevance conceptual or

implementation papers published between 2000 and 2026. Exclusion criteria included non-healthcare AI papers, opinion pieces without methodological or regulatory relevance, non-English records, duplicate records, and sources without clear relevance to clinical translation, safety, governance, or implementation. Evidence was mapped narratively across clinical use cases, safety-risk mechanisms, validation requirements, regulatory expectations, ethical-governance domains, multimodal accountability, lifecycle monitoring, and de-implementation readiness.[13,14]

Clinical trial reporting and protocol considerations were mapped to CONSORT-AI and SPIRIT-AI because AI interventions require transparent reporting of intervention details, user interactions, and input-output handling.[15,16]

Early-stage clinical evaluation was informed by DECIDE-AI, as many generative AI systems remain in the formative or early-deployment stages rather than at the level of mature randomized evidence.[17] Prediction-model reporting and appraisal were mapped to TRIPOD+AI and PROBAST+AI because diagnostic and prognostic AI claims require transparent reporting, risk-of-bias assessment, and applicability evaluation.[18,19] Earlier methodological work on TRIPOD-AI and PROBAST-AI was used only to contextualize the development of AI-specific reporting and bias-assessment standards.[20]

3. Generative AI in Digital Health: Definitions, Modalities, and Clinical Use Cases

3.1 Core technologies and model types

Generative AI refers to models that produce new text, images, code, structured outputs, summaries, or synthetic content in response to input data or prompts.[1] LLMs represent a major class of generative AI and are increasingly evaluated for healthcare communication, summarization, question answering, education, and clinical decision-support tasks.[2]

Mental health studies show that LLMs may support generative tasks such as screening support, psychoeducation, documentation assistance, and conversational tools, but evidence remains heterogeneous and requires careful clinical validation.[3,21] Voice-processing and LLM-based documentation systems suggest that generative AI may reduce documentation burden, although accuracy, privacy, clinician review, and workflow integration remain essential.[22,23]

Multimodal AI systems are particularly relevant for medical imaging because they may combine radiologic images with clinical context to support interpretation, reporting, and education.[6,7] Agentic AI systems extend beyond single-turn generation by using planning, tools, memory, and autonomous workflows, which creates new implementation and governance challenges.[8]

3.2 Clinical and operational use cases

Generative AI has been explored for administrative documentation, clinical summarization, medication-related harm mitigation, triage and referral support, diagnostic reasoning, education, imaging interpretation, and clinical simulation design.[4,5] EHR summarization and clinical-note generation may improve information management, but these applications require auditability, clinician review, and local workflow validation before routine use.[22,23]

Cardiovascular imaging work has also begun evaluating multimodal large language models for transthoracic echocardiography interpretation, illustrating specialty-specific expansion beyond general text-based use cases.[24]

Medical education and simulation may benefit from AI-supported scenario generation, but training outputs require expert review because plausible content can still be clinically inaccurate or incomplete.[25] Retrieval-augmented

generation may reduce unsupported outputs by grounding responses in curated evidence sources, but RAG systems still require validation of retrieval quality, generated reasoning, and clinical utility.[26] Generative methods have also entered biomedical discovery, including de novo molecular design, but drug-discovery applications require experimental validation beyond computational plausibility.[27]

3.3 Evidence maturity and clinical-risk level

Evidence maturity varies substantially across generative AI use cases, with administrative and educational tools generally carrying lower direct patient risk than diagnostic, prognostic, treatment-selection, or autonomous agentic applications.[1,11] High-risk clinical use cases require stronger evidence because errors may directly affect diagnosis, medication choice, triage priority, referral decisions, or patient communication.[4,5] Multimodal and agentic applications require additional safeguards because their outputs may emerge from complex interactions among heterogeneous data streams, tool calls, model reasoning, and user prompts.[6,8] **Table 1** summarizes representative use cases, expected benefits, risk level, evidence needs, and key safeguards for generative AI in digital health.[1,11]

Table 1: Representative use cases of generative AI in digital health.

Use case	Main approach	Potential benefit	Main risk	Evidence and safeguard requirement	Citation
Clinical documentation	LLM summarization, voice-to-note generation	Reduced documentation burden	Inaccurate or missing clinical details	Clinician review, audit trails, privacy control	[22,23]
Medication-related harm mitigation	LLM question answering and safety support	Medication education and error detection	Unsafe or incomplete drug information	Expert validation, evidence grounding, safety review	[4]
Triage and referral support	LLM-assisted workflow evaluation	Improved referral guidance	Misclassification or inappropriate triage	Prospective validation and workflow monitoring	[5]
Mental health generative tasks	LLM-based conversational or support tools	Scalable psychoeducation and documentation support	Safety, empathy, escalation, and reliability concerns	Clinical oversight and risk escalation pathways	[3,21]
Medical imaging	Multimodal image-text models	Report support and image-text reasoning	Domain shift and image misinterpretation	External validation across scanners and populations	[6,7]
Cardiovascular imaging	Multimodal echocardiography interpretation	Specialty-specific interpretation support	Misinterpretation across views or clinical contexts	External validation and clinician verification	[24]
Retrieval-augmented decision support	RAG models linked to knowledge sources	Grounded responses and literature-aware support	Retrieval error or misleading synthesis	Retrieval validation and clinician verification	[26]
Agentic workflows	Autonomous planning and tool use	Multistep task execution	Error cascades and unclear responsibility	Checkpoints, logs, human oversight, deactivation rules	[8]
Drug discovery	Generative molecular design	Candidate generation	Computational plausibility without biological proof	Experimental validation and reproducibility	[27]

Note: This table summarizes representative clinical and operational use cases of generative AI in digital health and maps each use case to its main benefit, potential risk, and required safeguard.

Abbreviations: LLM, large language model; RAG, retrieval-augmented generation.

4. Trustworthy Healthcare Delivery: Validation, Workflow Integration, and Real-World Performance

4.1 From benchmark performance to clinical utility

Trustworthy healthcare deployment requires more than high benchmark performance, as clinical utility depends on the patient population, intended use, workflow integration, user behaviour, and measurable outcomes.[11,12] Clinical trials involving AI interventions should report how the AI system is used, how outputs are presented, how clinicians interact with the system, and how patient outcomes are assessed.[15,16] Early-stage evaluation should document human-AI interaction in real clinical settings before claims of mature clinical effectiveness are made.[17] Prediction models should be reported and assessed with attention to external validity, risk of bias, calibration, applicability, and clinical consequences.[18,19] Medication-related generative AI evaluations show that outputs can appear useful while still requiring expert review for completeness, accuracy, and safety.[4]

4.2 Validation hierarchy

Validation should begin with technical performance testing but should not end there, as laboratory or retrospective datasets cannot fully capture the complexity of clinical workflow.[17] Internal validation estimates model performance under development conditions, whereas external validation tests generalization across institutions, populations, time periods, devices, and clinical contexts.[18] Prospective evaluation is especially important for decision-support systems because clinician interaction, workflow burden, alert fatigue, and implementation fidelity can influence effectiveness.[11] Post-deployment monitoring should track safety events, demographic performance, benefit-risk balance, and patient-centered outcomes after real-world implementation.[12] Controlled model updating should be governed through predefined change-control processes when AI-enabled device software is expected to evolve after deployment.[28]

4.3 Workflow integration and human oversight

Human oversight should be designed into the workflow rather than added informally after deployment because clinicians must understand when to accept, question, override, or escalate AI outputs.[15,17] Usability testing is essential because a technically accurate system can still fail if outputs are poorly timed, poorly explained, or poorly integrated into clinical routines.[11] Clinician review is particularly important for generative AI because fluent responses can mask hallucinations, omissions of context, or unsupported reasoning.[2,29] Health systems should define accountability for AI output review, documentation, patient communication, and incident reporting before patient-facing deployment.[1]

Implementation should also include de-implementation pathways to pause, restrict, update, or withdraw systems when safety or benefit-risk performance deteriorates.[12,28] The pathway by which computational errors may translate into downstream clinical harm is summarized in **Figure 2**.[11,17]

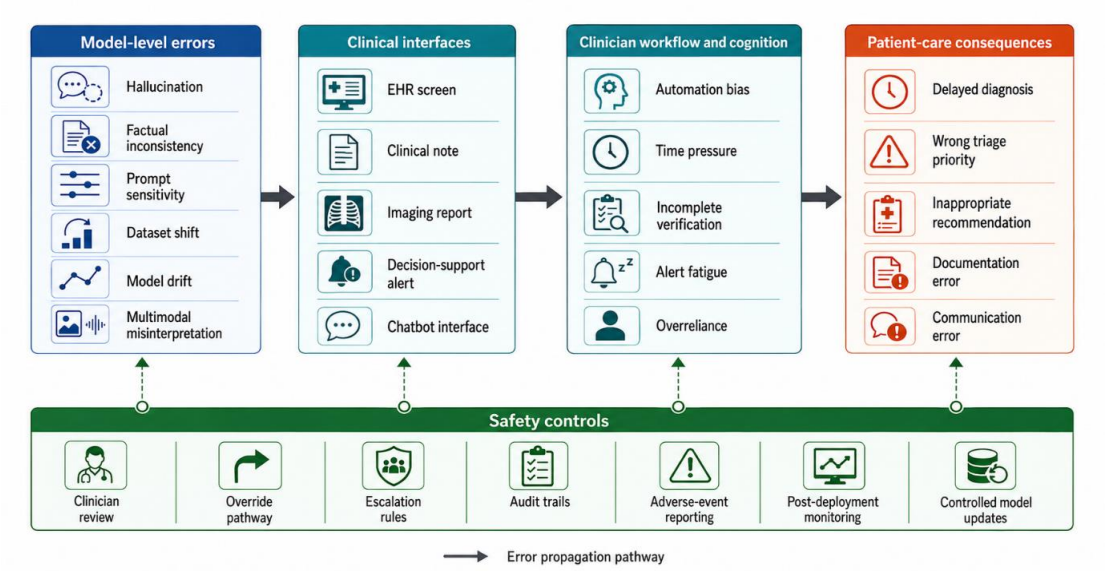

Figure 2: From computational error to clinical harm: safety cascade in generative AI-enabled healthcare delivery.

This figure illustrates how model-level errors, including hallucinations, factual inconsistencies, prompt sensitivity, dataset shift, model drift, and multimodal misinterpretations, may enter clinical workflows through EHR interfaces, notes, reports, or decision-support alerts.[2,6] Clinical harm may be amplified by automation bias, time pressure, incomplete verification, alert fatigue, and overreliance.[11,17] Safety controls include clinician review, override pathways, escalation rules, audit trails, adverse event reporting, post-deployment monitoring, and controlled model updates.[12,28]

Abbreviations: AI, artificial intelligence; AI/ML, artificial intelligence/machine learning; CDS, clinical decision support; EHR, electronic health record; FDA, U.S. Food and Drug Administration; GenAI, generative artificial intelligence; LLM, large language model; MLLM, multimodal large language model; PCCP, predetermined change control plan; SaMD, software as a medical device.

5. Generative AI-Specific Safety Risks: Hallucination, Bias, Explainability, Privacy, and Agentic Error Cascades

5.1 Hallucination and fabricated clinical reasoning

Hallucination refers to plausible yet incorrect, unsupported, or not well-grounded AI output, and it is a central safety concern for the use of LLMs in healthcare.[29] Clinical hallucinations may include unsupported diagnostic explanations, inaccurate medication information, fabricated citations, misleading summaries, or reasoning that appears coherent but lacks evidentiary support.[21,30] Multimodal hallucination creates additional risk because errors may arise from image interpretation, text generation, or the interaction between visual and clinical context.[6,7] Generated outputs should therefore be treated as draft decision support requiring clinician verification rather than as autonomous clinical reasoning.[2] Retrieval-augmented generation may reduce unsupported output when retrieval is accurate, but both the retrieval pipeline and the final generated answer must be validated.[26]

5.2 Bias and inequity

Generative AI may reproduce or amplify inequities when training data, evaluation datasets, deployment contexts, or user populations underrepresent demographic, linguistic, socioeconomic, or clinical subgroups.[31] Bias matters

clinically because model errors may affect diagnosis, access to treatment, triage priority, communication quality, patient trust, and the allocation of healthcare resources.[12] Mental health applications also raise concerns about cultural context, language coverage, safety escalation, and equitable access to high-quality support.[3,21] Imaging models may experience domain shift across scanner types, acquisition protocols, institutions, disease distributions, and patient populations.[6,7] Equity evaluation should therefore include subgroup performance, calibration, error analysis, and workflow impact before broad deployment.[18,19]

5.3 Explainability and clinician verification

Generated explanations from LLMs do not prove that the underlying reasoning is clinically valid, as fluent narratives may still be unsupported or incorrect.[2,29] Explainability methods can improve interpretability, but they cannot replace external validation, prospective testing, and outcome monitoring.[18,32] High-risk clinical systems should provide outputs that are traceable to reliable data, understandable for clinician review, and well documented for audit and accountability.[11,31] Explainability should therefore be combined with human oversight, transparent reporting, incident review, and post-deployment surveillance rather than treated as a standalone safety solution.[12,15]

5.4 Privacy and cybersecurity

Healthcare LLM deployment raises privacy risks when patient data are processed through third-party systems, cloud infrastructure, or models that are not governed by institutional data-control mechanisms.[33] Cybersecurity concerns include malicious prompting, data leakage, unauthorized access, adversarial manipulation, and unsafe integration of systems with clinical information systems.[34] These risks are especially relevant to documentation tools, patient-facing chatbots, imaging workflows, and EHR-connected systems, as they may handle sensitive patient information at scale.[22] Privacy-preserving deployment should include data minimization, access control, de-identification where appropriate, secure infrastructure, audit logging, and institutional governance.[31,33]

5.5 Agentic AI and error cascades

Agentic AI differs from single-response generative AI in that it can plan, retrieve, reason, use tools, and execute multistep actions with greater autonomy.[8] In healthcare, early errors in an agentic workflow may propagate through information retrieval, note drafting, alert generation, triage suggestions, or treatment recommendations.[1] Error cascades are important because small computational mistakes can worsen through user interface design, clinician workload, automation bias, and organizational routines.[11] Human oversight should therefore occur at predefined checkpoints before agentic outputs influence diagnosis, medication decisions, referral priority, or patient communication.[17] These risks reinforce the need for layered safeguards across technical, clinical, ethical, and governance levels.[28,31]

6. Ethical Governance, Equity, and Multimodal Accountability

6.1 Ethical principles in AI-mediated healthcare

Ethical governance for healthcare AI should protect autonomy, beneficence, non-maleficence, justice, transparency, accountability, and sustainability across the AI lifecycle.[35] Trustworthy AI also requires attention to clinical responsibility, informed use, patient communication, and professional judgment when AI contributes to care decisions.[31] A socio-technical design approach is needed because ethical performance depends not only on the model but also on workflows, users, data governance, institutional policy, and feedback mechanisms.[36] Ethical governance

must therefore be operationalized through procedures, committees, documentation, monitoring, and accountability structures rather than treated as abstract principles.[1]

6.2 Equity as a conditional outcome

Generative AI does not automatically improve equity because access, language coverage, dataset representation, digital literacy, and institutional capacity vary across populations and settings.[3] Clinical decision-support workflows should be tested across relevant subgroups to prevent hidden performance disparities from becoming embedded in routine care.[5,12] Imaging and multimodal systems require domain-specific equity evaluation because patient characteristics, acquisition protocols, equipment, and clinical context may influence performance.[6,7] Equity safeguards should include subgroup validation, fairness monitoring, patient-centered communication, and mechanisms for reporting harm or exclusion.[31,36]

6.3 Multimodal accountability

Multimodal accountability is needed when AI systems process or generate information across text, images, voice, wearables, laboratory data, genomic data, and patient-generated information.[6,35] Data provenance should be documented because model outputs may depend on upstream data quality, clinical context, and modality-specific preprocessing.[36] Model-version control is essential because changes to prompts, models, retrieval sources, or deployment infrastructure can alter output behavior.[28] Output traceability should allow clinicians and auditors to identify what information was used, what was generated, who reviewed it, and how it influenced patient care.[12,31]

6.4 Micro-, meso-, and macro-level governance

Governance should operate at the micro level of clinician-patient interaction, the meso level of institutional deployment, and the macro level of regulatory and professional oversight.[36] At the micro level, clinicians need authority to question, override, document, and explain AI-assisted recommendations.[15,17] At the meso level, healthcare organizations should maintain AI governance committees, local validation processes, staff training, incident-reporting systems, and equity monitoring.[1,31] At the macro level, regulators and professional societies should support lifecycle evidence requirements, postmarket surveillance, reporting standards, and public accountability.[12,28] **Table 2** summarizes responsibility allocation across clinical, organizational, technical, and regulatory levels for accountable generative AI implementation.[1,31,36]

Table 2: Multimodal accountability and implementation responsibility matrix for generative AI in digital health.

Level	Responsible actors	Core responsibility	Practical governance mechanism	Citation
Micro	Clinicians, patients, caregivers	Safe interpretation, disclosure, review, and override	Clinician review, documentation, escalation, patient communication	[15,17,31]
Meso	Hospitals and AI governance committees	Local validation, implementation, training, equity monitoring	Governance board, audit trail, incident review, subgroup monitoring	[1,31,36]
Meso	Informatics, cybersecurity, vendor teams	Data security, version control, privacy, update governance	Access control, logging, cybersecurity review, change management	[28,33,34]
Macro	Regulators and policy bodies	Benefit-risk oversight, lifecycle surveillance, public safety	Premarket evidence, postmarket reporting, PCCP, corrective action	[12,28]
Macro	Professional societies and journals	Evidence standards, reporting norms, transparency, public trust	Guidelines, publication standards, registries, professional education	[13,15,18]

Note: This table summarizes accountability responsibilities across micro-, meso-, and macro-level implementation of generative AI in digital health.

Abbreviations: AI, artificial intelligence; PCCP, predetermined change control plan.

7. Global Regulatory-Science Landscape and Compliance Alignment

7.1 Risk-based regulation of generative AI in health

Regulatory classification of generative AI should be based on intended use, clinical risk, patient-facing impact, and whether the system functions as medical-device software.[9,10] The EU AI Act establishes a risk-based legal framework for AI and includes obligations for high-risk systems.[37] The European Commission identifies AI-based software intended for medical purposes as a high-risk area requiring attention to safety, transparency, human oversight, and regulatory compliance.[38] The FDA regulates AI-enabled device software functions through medical-device pathways when the software meets device definitions and intended-use criteria.[39] Health Canada guidance similarly applies risk-based expectations to machine learning-enabled medical devices submitted for regulatory review.[40]

7.2 Lifecycle control and model updating

Generative AI and AI-enabled device software may require lifecycle governance because model behavior can change due to updates, new data, modified prompts, retrieval changes, or shifts in deployment context.[1] FDA predetermined change control guidance recommends describing planned modifications, validation methods, implementation procedures, and impact assessments before specified changes are made.[28] Good Machine Learning Practice principles support safe and effective AI/ML medical device development throughout the entire product lifecycle.[41] IMDRF clinical-evaluation guidance emphasizes scientific validity, analytical validity, and clinical performance for software as a medical device.[42] These sources collectively support a regulatory-science approach in which evidence, monitoring, and change control are maintained after deployment.[28,39]

7.3 Compliance alignment for multijurisdictional deployment

International alignment is important because developers may face different expectations across the FDA, the EU, Health Canada, the IMDRF, the WHO, and institutional governance systems.[35,37] Compliance files should therefore describe intended use, population, data sources, validation evidence, safety monitoring, cybersecurity controls, update policy, and benefit-risk rationale.[39,40] Regulatory harmonization should not eliminate local accountability, as health-system capacity, patient populations, legal requirements, and clinical workflows vary across jurisdictions.[10,36]

Table 3 summarizes the major regulatory science domains and their practical implications for generative AI in digital health.[28,42]

Table 3: Comparative regulatory-science alignment matrix for generative AI in digital health.

Regulatory-science domain	Major framework or body	Core expectation	Practical implication	Citation
Risk-based AI governance	EU AI Act and European Commission	Obligations depend on intended use, risk, and medical purpose	Patient-facing AI should be classified according to clinical risk	[37,38]
AI-enabled device software	FDA	Device oversight depends on intended use, risk, and evidence	Clinical AI software may require regulatory review	[39]
ML-enabled medical devices	Health Canada	Risk-based evidence expectations for MLMDs	Planned learning or updates should remain within evaluated boundaries	[40]
Predetermined change control	FDA	Planned changes should be documented and validated	Model updates should occur under controlled governance	[28]
GMLP	FDA, Health Canada, MHRA	High-quality AI/ML device development across lifecycle	Development and monitoring should be lifecycle-based	[41]

SaMD clinical evaluation	IMDRF	Scientific validity, analytical validity, and clinical performance	Evidence should match intended use and claimed benefit	[42]
LMM governance	WHO	Ethical governance, transparency, accountability, safety, and equity	Multimodal systems require responsible oversight	[35]

Note: This table summarizes key regulatory-science domains relevant to generative AI-enabled digital health systems and maps them to major regulatory or governance expectations.

Abbreviations: AI, artificial intelligence; EU, European Union; FDA, U.S. Food and Drug Administration; GMLP, Good Machine Learning Practice; IMDRF, International Medical Device Regulators Forum; LMM, large multimodal model; MHRA, Medicines and Healthcare products Regulatory Agency; MLMD, machine learning-enabled medical device; PCCP, predetermined change control plan; SaMD, software as a medical device; WHO, World Health Organization.

8. Proposed Translational Regulatory-Science Framework

8.1 Framework overview

This review proposes a conceptual, lifecycle-based translational regulatory science framework for generative AI in digital health that links clinical need, evidence generation, ethical governance, regulatory readiness, implementation monitoring, and de-implementation.[1,17] The framework should be interpreted as an author-proposed synthesis of existing principles from clinical evaluation, implementation science, ethical governance, and regulatory science, rather than an externally validated tool.[31,36] Trustworthy deployment requires lifecycle governance because the safety of healthcare AI depends on transparency, accountability, fairness, usability, robustness, human oversight, and post-deployment learning.[41] **Figure 3** illustrates the proposed framework and shows how feedback loops connect intended use, validation, governance, regulatory control, implementation, and de-implementation.[1,36]

8.2 Clinical need and intended use

The first domain requires explicit definition of the clinical problem, target users, intended population, intended workflow, and expected decision-support role before generative AI is introduced into care.[17] AI trial reporting guidance supports clear description of the clinical setting, user role, intervention function, input-output handling, and human-AI interaction.[15,16] Overly broad intended-use claims create safety and regulatory risks because evidence from one task, population, or workflow cannot automatically justify deployment in another setting.[28,42] Clinician and patient engagement during intended-use specification helps ensure that the system addresses real workflow needs, preserves judgment, and reflects patient-centered priorities.[1,31]

8.3 Evidence generation and validation

Clinical evidence generation should include transparent reporting of AI intervention details, trial context, user interaction, and protocol-level safety considerations.[15,16] Prediction-model evidence should address reporting quality, risk of bias, applicability, and external validity when AI supports diagnostic or prognostic decisions.[18,19] Early-stage evaluation should document how an AI decision-support system is used in live clinical settings and how clinicians interact with its outputs.[17] Real-world evidence should monitor benefit-risk balance, demographic representation, safety outcomes, patient-centered outcomes, and post-deployment performance over time.[12] This evidence should

inform model refinement, regulatory submissions, local governance, and de-implementation decisions when performance is not sustained.[1,28]

8.4 Ethical governance and multimodal accountability

Ethical governance embeds autonomy, beneficence, non-maleficence, justice, transparency, and accountability across the AI lifecycle.[31] WHO guidance for large multimodal models emphasizes governance needs across health care, public health, research, and drug development.[35] Multimodal medical-imaging models require additional accountability because images, clinical text, and contextual data may jointly influence model output and clinician interpretation.[6] Data provenance, model-version tracking, output traceability, and stakeholder responsibility should therefore be documented across the complete data-to-decision pathway.[36] These safeguards are especially important when generative AI outputs enter clinical documentation, referral pathways, medication counselling, or patient-facing communication.[4,22]

8.5 Regulatory readiness and lifecycle control

Regulatory readiness should begin with quality management and lifecycle documentation for AI/ML-enabled medical device development.[41] FDA PCCP guidance recommends that sponsors describe planned AI-enabled device modifications, validation methods, implementation procedures, and impact assessments.[28] Benefit-risk reporting should include study design, demographic representation, safety evidence, patient outcomes, adverse events, and post-deployment performance where available.[12] De-implementation criteria should be predefined around validated safety signals, loss of clinical utility, unacceptable workflow burden, equity concerns, or failure to sustain benefit-risk performance rather than unsupported universal numerical thresholds.[1,31]

8.6 Implementation, feedback, and de-implementation

Implementation requires clinician training, usability testing, workflow integration, local evaluation, and ongoing monitoring, as generative AI performance depends on both socio-technical context and model capability.[17] Feedback mechanisms should capture clinicians' experiences, patients' understanding, incident reports, workflow burden, unintended consequences, and implementation fidelity.[36] De-implementation pathways should define how organizations pause, restrict, update, or withdraw systems when post-deployment evidence shows unsafe performance, poor clinical value, or unacceptable governance risk.[28] In this framework, real-world feedback is not a final step but a continuous loop that can refine intended use, trigger new validation, modify governance, or support withdrawal from practice.[1] The complete lifecycle-based translational regulatory-science framework is summarized in Figure 3.[17,36]

Figure 3: Proposed lifecycle-based translational regulatory-science framework for trustworthy generative AI in digital health.

This figure presents a lifecycle framework linking intended use, evidence generation, ethical governance, multimodal accountability, regulatory readiness, implementation feedback, and de-implementation.[17,36] Evidence generation includes trial reporting, prediction-model reporting, risk-of-bias appraisal, applicability assessment, and early clinical evaluation.[15,18,19] Regulatory readiness integrates benefit-risk reporting, predetermined change-control planning, lifecycle monitoring, and controlled model modification.[12,28] Feedback loops indicate that real-world evidence, safety signals, equity findings, and user experience should inform refinement, restriction, or withdrawal from clinical use.[1,36]

Abbreviations: AI, artificial intelligence; AI/ML, artificial intelligence/machine learning; CONSORT-AI, Consolidated Standards of Reporting Trials–Artificial Intelligence; FDA, U.S. Food and Drug Administration; GenAI, generative artificial intelligence; GMLP, Good Machine Learning Practice; PROBAST+AI, Prediction model Risk Of Bias Assessment Tool–Artificial Intelligence; SPIRIT-AI, Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence; TRIPOD+AI, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis–Artificial Intelligence.

9. Future Perspectives and Implementation Roadmap

9.1 Future technical directions

Future generative AI in digital health should progress from general-purpose text generation toward clinically grounded systems that combine reasoning, verification, and task-specific oversight.[8] Large medical imaging models and multimodal LLMs may support broader clinical reasoning by integrating radiologic images, clinical text, and contextual patient information, but their safe deployment requires validation across modalities, institutions, and patient groups.[6,7]

Retrieval-augmented generation offers one pathway to ground outputs in curated biomedical literature or institutional knowledge bases, rather than relying solely on static training data.[26] Voice-enabled and documentation-focused systems may reduce administrative burden, but clinical value depends on accuracy, clinician review, privacy protection, and workflow fit.[22] Future progress should therefore be judged by clinical usefulness, safety, fairness, and implementation readiness rather than model scale alone.[3,5]

9.2 Future regulatory and policy directions

Regulatory science is likely to shift toward lifecycle-based oversight because AI-enabled medical devices may change after deployment and require ongoing benefit-risk monitoring.[12] Health-system policies should strengthen governance of study design, demographic representation, transparency, safety outcomes, patient-centered evidence, and post-deployment performance.[31,36] Predetermined change-control planning can support controlled model updates when planned modifications, validation methods, implementation procedures, and impact assessments are prospectively documented.[28] Regulatory and institutional alignment remains important because validation, transparency, postmarket monitoring, and accountability expectations differ across implementation settings.[10,36]

9.3 Health-system implementation readiness

Sustained adoption of generative AI requires institutional readiness rather than simple software procurement.[1] Health systems should establish governance structures for tool selection, local validation, clinician training, incident review, equity monitoring, and periodic performance reassessment.[31,36] Clinician education should focus on appropriate use, recognition of model limitations, interpretation of uncertain outputs, and preservation of professional judgment.[5]

Documentation and ambient-assistance use cases should be implemented with audit trails, human review, data-security controls, and feedback mechanisms before broad deployment.[22]

9.4 Practical roadmap

Short-term priorities should include prospective validation of high-priority clinical use cases, standardized evaluation of generative tasks, and establishment of institutional AI governance processes.[3,11] Medium-term priorities should include postmarket evidence networks, structured benefit-risk reporting, and controlled update pathways for AI-enabled medical-device software.[12,28] Long-term priorities should focus on clinically grounded agentic systems, mature governance, transparent accountability, and routine withdrawal or restriction of systems that fail to sustain safety or clinical value.[8,31] A staged implementation roadmap for clinically accountable generative AI is shown in **Figure 4.**[28]

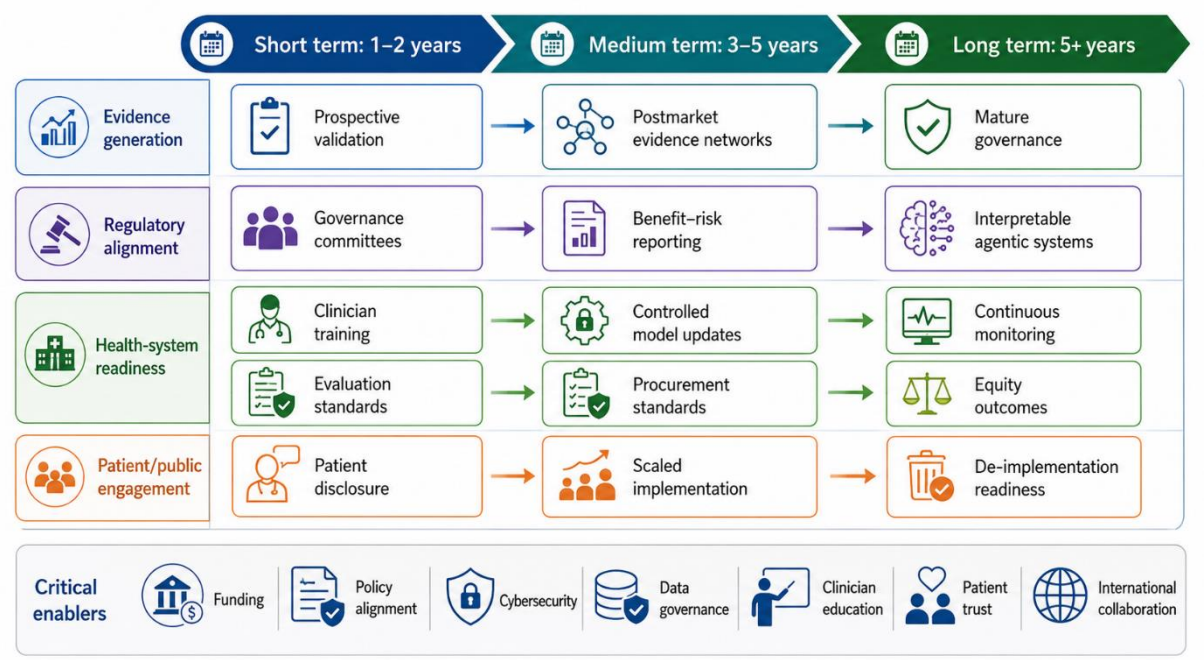

Figure 4: Staged implementation roadmap for clinically accountable generative AI in digital health.

This figure outlines short-, medium-, and long-term priorities for responsible generative AI implementation across evidence generation, regulatory alignment, health-system readiness, and patient or public engagement.[1,12] Short-term priorities include prospective validation, governance structures, clinician training, and standardized evaluation.[3,11]

Medium-term priorities include postmarket evidence networks, benefit-risk reporting, controlled update pathways, procurement standards, and scaled implementation.[12,28] Long-term priorities include clinically grounded agentic systems, mature governance, continuous monitoring, equity outcomes, and de-implementation readiness.[8,31]

Abbreviations: AI, artificial intelligence; FDA, U.S. Food and Drug Administration; GenAI, generative artificial intelligence; RAG, retrieval-augmented generation.

10. CONCLUSION

Generative AI can strengthen health delivery only when deployment is guided by implementation planning, clinical validation, ethical oversight, lifecycle monitoring, and accountability.[1] This review positions generative AI as a socio-technical healthcare system shaped by stakeholders, workflows, data governance, and trust requirements.[36]

Benchmark performance and market clearance are insufficient surrogates for clinical utility because benefit-risk reporting and evidence of patient outcomes for AI/ML devices remain incomplete.[12] Trustworthy use also requires multimodal accountability, transparent oversight, fairness safeguards, and clinician authority where AI outputs influence patient care.[6,31] Model changes should be managed through lifecycle controls and, where relevant, predetermined change-control planning rather than uncontrolled post-deployment modification.[28] The proposed framework synthesizes these requirements into an evidence-bound pathway linking intended use, validation, governance, monitoring, feedback, and de-implementation.[1,36] Future progress will require collaboration among clinicians, developers, regulators, patients, and health systems to ensure generative AI supports healthcare quality, safety, equity, and trust.[12,31]

Declarations

Ethics approval and consent to participate

Not applicable. This article is a narrative review and did not involve human participants, animal experiments, identifiable patient data, or primary data collection.

Consent for publication

Not applicable.

Availability of data and materials

No new datasets were generated or analyzed during the preparation of this review. All information discussed in this article was derived from previously published peer-reviewed literature, reporting guidelines, and publicly available regulatory or governance documents cited in the reference list.

Competing interests

The author(s) declare no competing interests.

Funding

No specific funding was received for the preparation of this review.

Authors' contributions

All author(s) contributed to the conceptualization, literature review, evidence synthesis, manuscript drafting, critical revision, and approval of the final version of the manuscript.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

1. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci, 2024; 19: 27. doi: 10.1186/s13012-024-01357-9.
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med, 2023; 29(8): 1930-1940. doi: 10.1038/s41591-023-02448-8.
3. Hua Y, Na H, Li Z, Liu F, Fang X, Clifton DA, et al. A scoping review of large language models for generative tasks in mental health care. npj Digit Med, 2025; 8: 230. doi: 10.1038/s41746-025-01611-4.
4. Ong JCL, Chen MH, Ng N, Elangovan K, Tan NYT, Jin L, et al. A scoping review on generative AI and large language models in mitigating medication related harm. npj Digit Med, 2025; 8: 182. doi: 10.1038/s41746-025-01565-7.
5. Gaber F, Shaik M, Allega F, Bilecz AJ, Busch F, Goon K, et al. Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. npj Digit Med, 2025; 8: 263. doi: 10.1038/s41746-025-01684-1.
6. Nam Y, Kim DY, Kyung S, Seo JY, Song JM, Kwon J, et al. Multimodal large language models in medical imaging: current state and future directions. Korean J Radiol, 2025; 26(10): 900-923. doi: 10.3348/kjr.2025.0599.
7. Fang M, Wang Z, Pan S, Feng X, Zhao Y, Hou D, et al. Large models in medical imaging: advances and prospects. Chin Med J., 2025; 138(14): 1647-1664. doi: 10.1097/CM9.0000000000003699.

8. Acharya DB, Kuppan K, Divya B. Agentic AI: autonomous intelligence for complex goals—a comprehensive survey. IEEE Access, 2025; 13: 18912-18936. doi: 10.1109/ACCESS.2025.3532853.
9. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models or generative AI in healthcare. npj Digit Med, 2023; 6: 120. doi: 10.1038/s41746-023-00873-0.
10. Ong JCL, Chang SYH, William W, Butte AJ, Shah NH, Chew LST, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health, 2024; 6(6): e428-e432. doi: 10.1016/S2589-7500(24)00061-X.
11. You JG, Hernandez-Boussard T, Pfeffer MA, Landman A, Mishuris RG. Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. npj Digit Med, 2025; 8: 107. doi: 10.1038/s41746-025-01506-4.
12. Lin JC, Jain B, Iyer JM, Rola I, Srinivasan AR, Kang C, et al. Benefit-risk reporting for FDA-cleared artificial intelligence-enabled medical devices. JAMA Health Forum, 2025; 6(9): e253351. doi: 10.1001/jamahealthforum.2025.3351.
13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews: checklist and explanation. Ann Intern Med, 2018; 169(7): 467-473. doi: 10.7326/M18-0850.
14. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth, 2020; 18(10): 2119-2126. doi: 10.11124/JBIES-20-00167.
15. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med, 2020; 26(9): 1364-1374. doi: 10.1038/s41591-020-1034-x.
16. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med, 2020; 26(9): 1351-1363. doi: 10.1038/s41591-020-1037-7.
17. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med, 2022; 28(5): 924-933. doi: 10.1038/s41591-022-01772-9.
18. Collins GS, Dhiman P, Ma J, Andaur Navarro CL, Reitsma JB, Logullo P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 2024; 385: e078378. doi: 10.1136/bmj-2023-078378.
19. Moons KGM, Damen JAAG, Kaul T, Dhiman P, Andaur Navarro CL, Reitsma JB, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 2025; 388: e082505. doi: 10.1136/bmj-2024-082505.
20. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline and risk of bias tool for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open, 2021; 11(7): e048008. doi: 10.1136/bmjopen-2020-048008.
21. Guo Z, Lai A, Thygesen JH, Farrington J, Keen T, Li K. Large language models for mental health applications: systematic review. JMIR Ment Health, 2024; 11: e57400. doi: 10.2196/57400.
22. Xu Y, Jia H, Wang M, Feng J, Xu X, Wang H, et al. Enhancing clinical documentation with voice processing and large language models: a study on the LAOS system. npj Digit Med, 2025; 8: 798. doi: 10.1038/s41746-025-02170-4.

23. Croxford E, Gao Y, Pellegrino N, First E, Sun Q, Chen G, et al. Evaluating clinical AI summaries with large language models as judges. npj Digit Med, 2025; 8: 640. doi: 10.1038/s41746-025-02005-2.
24. Li J, He J, et al. Exploring multimodal large language models on transthoracic echocardiography interpretation. J Biomed Inform, 2025; 171: 104930. doi: 10.1016/j.jbi.2025.104930.
25. Barra FL, Rodella G, Costa A, Scalogna A, Carenzo L, Monzani A, et al. From prompt to platform: an agentic AI workflow for healthcare simulation scenario design. Adv Simul (Lond), 2025; 10: 29. doi: 10.1186/s41077-025-00357-z.
26. Ozmen BB, Sozener CB, Schnier S, Chiarini S, Oliver JD, Kwong JW, et al. MicroRAG: development of a novel artificial intelligence retrieval-augmented generation model for microsurgery clinical decision support. Microsurgery, 2025; 45(8): e70138. doi: 10.1002/micr.70138.
27. Alakhdar A, Poczos B, Washburn NR. Diffusion models in de novo drug design. J Chem Inf Model, 2024; 64(19): 7238-7256. doi: 10.1021/acs.jcim.4c01107.
28. US Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: guidance for industry and Food and Drug Administration staff [Internet]. Silver Spring (MD): US Food and Drug Administration; 2025 [cited 2026 Jun 22]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence
29. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on large language models. npj Digit Med, 2024; 7: 183. doi: 10.1038/s41746-024-01157-x.
30. Giorgi S, Isman K, Liu T, Fried Z, Sedoc J, Curtis B. Evaluating generative AI responses to real-world drug-related questions. Psychiatry Res, 2024; 339: 116058. doi: 10.1016/j.psychres.2024.116058.
31. Goktas P, Grzybowski AE. Shaping the future of healthcare: ethical clinical challenges and pathways to trustworthy AI. J Clin Med, 2025; 14(5): 1605. doi: 10.3390/jcm14051605.
32. Zhu L, Chen Z, Zhang H, Chen H, Liu L, Yu W, et al. Explainable AI unravels sepsis heterogeneity via coagulation-inflammation profiles for prognosis and stratification. Nat Commun, 2025; 16: 10396. doi: 10.1038/s41467-025-65365-z.
33. Zhong X, Li S, Chen Z, Ge L, Yu D, Wang S, et al. Considerations for patient privacy of large language models in health care: scoping review. J Med Internet Res, 2025; 27: e76571. doi: 10.2196/76571.
34. Akinci D'Antonoli T, Tejani AS, Khosravi B, Bluethgen C, Busch F, Bressem KK, et al. Cybersecurity threats and mitigation strategies for large language models in health care. Radiol Artif Intell, 2025; 7(4): e240739. doi: 10.1148/ryai.240739.
35. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models [Internet]. Geneva: World Health Organization; 2025 [cited 2026 Jun 22]. Available from: https://www.who.int/publications/i/item/9789240084759
36. Moreno-Sánchez PA, Del Ser J, van Gils M, Hernesniemi J. A design framework for operationalizing trustworthy artificial intelligence in healthcare: requirements, tradeoffs and challenges for its clinical adoption. Inf Fusion, 2026; 127: 103812. doi: 10.1016/j.inffus.2025.103812.
37. European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [Internet]. Official

Journal of the European Union; 2024 [cited 2026 Jun 22]. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

38. European Commission. Artificial intelligence in healthcare [Internet]. Brussels: European Commission; [date unknown; cited 2026 Jun 22]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_en
39. US Food and Drug Administration. Artificial intelligence in software as a medical device [Internet]. Silver Spring (MD): US Food and Drug Administration, 2025 [cited 2026 Jun 22]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device
40. Health Canada. Pre-market guidance for machine learning-enabled medical devices [Internet]. Ottawa: Health Canada; 2026 [cited 2026 Jun 22]. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/pre-market-guidance-machine-learning-enabled-medical-devices.html
41. US Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Good machine learning practice for medical device development: guiding principles [Internet]. Silver Spring (MD): US Food and Drug Administration, 2021 [cited 2026 Jun 22]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
42. International Medical Device Regulators Forum. Software as a Medical Device: clinical evaluation. IMDRF/SaMD WG/N41FINAL:2017 [Internet]. Geneva: International Medical Device Regulators Forum, 2017 [cited 2026 Jun 22]. Available from: https://www.imdrf.org/documents/software-medical-device-samd-clinical-evaluation