

Artificial Intelligence in Healthcare: Applications, Evidence, Challenges, and Directions for Responsible Implementation

Prof. Verena Lengston

Faculty of Computer Engineering, Vienna University of Technology, Vienna, Austria.

Submitted: September 23, 2026

Accepted: September 28, 2026

Published: October 05, 2026

Citation: Verena Lengston (2026) Artificial Intelligence in Healthcare: Applications, Evidence, Challenges, and Directions for Responsible Implementation. Epistora J. Artif. Intell. & Intell. Syst. 1(2), 1-07. Article EJAIS-116

Abstract

Artificial intelligence (AI) has become an important part of the digital transformation of healthcare. It is applied to medical imaging, clinical decision support, patient monitoring, administrative documentation, drug discovery, public-health surveillance, and health-system management. The promise of AI is not simply faster computation; it is the ability to identify clinically useful patterns in large and heterogeneous data sources, including electronic health records, laboratory results, medical images, genomic data, and free-text notes. However, technical performance alone does not establish clinical value. Healthcare AI can create or worsen risks related to bias, privacy, cybersecurity, lack of transparency, workflow disruption, automation bias, and unclear accountability. This narrative review examines major applications of AI in healthcare, summarizes its potential benefits and limitations, and discusses requirements for safe and equitable implementation. The review argues that successful health AI must be evaluated across the full lifecycle—from problem selection and data collection to clinical validation, deployment, post-market monitoring, and governance. Human oversight, representative data, transparent communication, prospective evaluation, and clear responsibility are essential. AI is most likely to improve care when it augments rather than replaces professional judgment and when it is integrated into real clinical workflows in ways that protect patient autonomy and public trust.

Keywords: Artificial Intelligence, Machine Learning, Healthcare, Clinical Decision Support, Medical Imaging, Ethics, Governance, Health Equity

Introduction

Artificial intelligence refers to computational systems designed to perform tasks commonly associated with aspects of human intelligence, such as recognizing patterns, making predictions, processing language, or generating content. In healthcare, AI includes machine learning, deep learning, natural-language processing, computer vision, and, increasingly, generative AI systems. These methods can process large quantities of information at speeds that would be difficult for individual clinicians or care teams to match.

Interest in AI has grown because contemporary healthcare generates vast quantities of data. Hospitals collect medical images, laboratory values, physiological signals, prescriptions, clinical notes, claims data, genomic information, and patient-reported outcomes. Yet much of this information is fragmented, unstructured, or difficult to use at the point of care. AI may help transform these data into information that supports diagnosis, prognosis, triage, treatment planning, and operational decisions.

Despite this potential, AI should not be considered inherently beneficial merely because it is technically advanced. A highly accurate model may still be clinically unhelpful if it is evaluated on an unrepresentative population, used outside its intended setting, poorly integrated into a workflow, or trusted beyond its actual capability. The World Health Organization (WHO) emphasizes that health AI must be guided by protection of autonomy, promotion of human well-being and safety, transparency, accountability, inclusion and equity, and responsiveness and sustainability [1]. These principles frame AI as a socio-technical intervention rather than a purely technical product.

This narrative review discusses the principal uses of AI in healthcare; the benefits that may arise from implementation; major clinical, ethical, and organizational risks; and practical strategies for responsible adoption. It also considers how regulators, health systems, developers, clinicians, and patients can share responsibility for ensuring that AI serves health rather than simply introducing new forms of risk.

Major Applications of AI in Healthcare

Medical imaging and diagnostic support

Medical imaging is among the most visible uses of AI in healthcare. Computer-vision systems can analyze radiographs, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound images, digital pathology slides, retinal photographs, and dermatological images. Deep-learning models may identify image features associated with conditions such as fractures, pulmonary abnormalities, cancer, diabetic retinopathy, stroke, or cardiovascular disease.

The strongest practical role for many imaging systems is not autonomous diagnosis but clinician support. For example, an AI tool can flag a suspicious region on an image, prioritize potentially urgent examinations in a radiology worklist, quantify anatomical structures, or identify scans that may require rapid specialist review. Such tools can reduce repetitive workload and may improve consistency in narrow, well-defined tasks. However, their performance can vary across scanners, institutions, image quality, disease prevalence, and patient populations. Therefore, a system that performs well in retrospective testing may not deliver the same benefit in routine clinical practice.

The U.S. Food and Drug Administration (FDA) maintains information on AI-enabled medical devices that have been authorized for marketing in the United States, illustrating the breadth of AI applications in regulated medical products [2]. Regulatory authorization, however, should not be interpreted as proof that a product will improve every clinical workflow or every outcome. Local validation and post-deployment monitoring remain necessary.

Clinical decision support and risk prediction

AI-based clinical decision support systems use patient data to estimate risk, identify patterns, or generate recommendations. Potential applications include predicting deterioration in hospitalized patients, identifying risk of sepsis, estimating the likelihood of hospital readmission, detecting medication-related harm, prioritizing referrals, and supporting diagnosis.

These applications may be especially useful when clinicians must synthesize numerous variables under time pressure. A model may recognize a pattern distributed across vital signs, laboratory trends, medication history, and clinical notes that is not obvious from any single variable. However, risk prediction is not the same as causal explanation. A model may accurately predict an outcome while relying on correlations that do not generalize, or on signals that reflect unequal access to care rather than biological risk.

A useful clinical decision-support tool should answer a clearly defined question, provide information at the appropriate time, fit into the clinician's workflow, and lead to an action that can improve care. If an alert does not change management, or if it generates too many false alarms, it may create alert fatigue rather than benefit. Evaluation should therefore examine patient-centered outcomes and workflow effects—not only statistical measures such as area under the receiver-operating-characteristic curve.

Natural-language processing and generative AI

A large share of healthcare information exists as free text: progress notes, discharge summaries, referral letters, pathology reports, radiology reports, and messages between patients and care teams. Natural-language processing (NLP) can extract diagnoses, symptoms, medications, adverse events, or social determinants of health from these records. It can also support coding, research cohort identification, quality measurement, and clinical documentation.

Generative AI, especially large language models, extends these capabilities. It can draft summaries, simplify complex medical language, help structure clinical notes, translate patient information, support education, and assist with administrative communication. It may also help researchers search and synthesize large bodies of literature.

Nevertheless, generative systems have important limitations. They can produce fluent but false statements, omit key uncertainty, reflect biases in training data, or provide recommendations that are not suited to an individual patient. WHO has urged caution in the use of large language model tools for health because incorrect, biased, or fabricated content can compromise patient safety and public trust [3]. Clinical use therefore requires human review, clear limits on autonomous action, and safeguards for confidential data.

Personalized medicine and precision health

AI may help identify subgroups of patients who are likely to respond differently to treatment, develop particular complications, or require more intensive monitoring. In precision medicine, AI can integrate clinical data with genomic, proteomic, imaging, and lifestyle information. The goal is to move beyond a one-size-fits-all approach toward more individualized prevention and treatment.

Potential benefits include more accurate risk stratification, identification of therapeutic targets, and selection of patients for clinical trials. Yet this area also raises difficult questions. Genomic and multi-modal datasets are highly sensitive; unequal representation in research can cause models to work less well for historically underrepresented groups; and complex model outputs may be difficult to interpret in patient discussions. Precision should not become a justification for excluding people from care or for making opaque decisions without accountability.

Drug discovery and biomedical research

AI is increasingly used in biomedical research to analyze molecular structures, identify potential drug targets, predict properties of candidate compounds, and organize scientific literature. These applications may make early-stage research more efficient by narrowing large search spaces and prioritizing experiments.

However, AI-generated hypotheses still require laboratory validation, toxicology testing, and appropriately designed clinical trials. Faster hypothesis generation does not eliminate the need for evidence of safety, efficacy, and real-world effectiveness. In this sense, AI can accelerate parts of discovery, but it does not replace the scientific method or regulatory standards for therapeutic products.

Public health and healthcare operations

At the population level, AI can assist with disease surveillance, forecasting resource demand, identifying geographic patterns, analyzing social and environmental determinants of health, and targeting public-health interventions. Within hospitals and clinics, it can support scheduling, bed management, supply-chain planning, staffing forecasts, billing, and appointment triage.

Operational applications can improve efficiency, but they may also embed institutional priorities that are not aligned with patient needs. For example, a scheduling system optimized only for throughput could worsen access for patients with complex needs. Governance should therefore define the objective function explicitly: efficiency matters, but so do equity, safety, continuity of care, and patient experience.

Potential Benefits of AI

The potential value of AI in healthcare is substantial when systems are appropriately designed and used.

First, AI can increase the speed with which information is processed. In an environment with expanding data volume and limited clinician time, this may help teams identify urgent cases and reduce delays.

Second, AI may improve consistency in repetitive, narrowly defined tasks. In imaging, for example, a model can apply the same decision rule to every eligible image. This does not guarantee correctness, but it can reduce variation and provide a second layer of review.

Third, AI can reduce administrative burden. Documentation, coding, summarization, scheduling, and retrieval of information often consume significant staff time. If carefully implemented, automation may allow clinicians to devote more attention to communication, examination, and complex decision-making.

Fourth, AI may expand access to specialized expertise, particularly in settings with shortages of clinicians. A decision-support system could assist primary-care or frontline teams in recognizing cases that need specialist assessment. However, these benefits depend on infrastructure, training, connectivity, local validation, and referral capacity. Technology alone cannot solve structural shortages in healthcare.

Finally, AI may support learning health systems: organizations that continuously use data to assess performance and improve care. In such a system, implementation is not a one-time event. Models, workflows, clinical outcomes, and equity impacts are regularly monitored and improved.

Risks and Challenges

Bias, fairness, and health equity

Bias is one of the most important risks in health AI. A model learns from historical data, and historical data may reflect unequal diagnosis, treatment, access, measurement, or documentation. If a dataset underrepresents a population, the resulting model may perform less accurately for that group. If a target outcome reflects unequal access to services, the model may reproduce rather than correct inequity.

Fairness cannot be established through a single overall performance number. Developers and healthcare organizations should examine performance across clinically relevant subgroups, including age, sex, gender where relevant, race and ethnicity where appropriate and ethically collected, disability, language, geography, and socioeconomic conditions. They should also investigate whether different error types have unequal consequences for different groups.

Privacy, consent, and cybersecurity

Healthcare data are among the most sensitive forms of personal information. AI development often requires access to large datasets, creating risks of unauthorized use, re-identification, data leakage, and cyberattack. Generative AI creates additional concerns when users enter identifiable patient information into tools that do not have appropriate data-protection arrangements.

Organizations should use data minimization, access controls, encryption, audit trails, de-identification where suitable, and clear contractual and technical safeguards. Patients should receive understandable information about significant uses of their data. Where consent is not practical or legally required for each use, governance mechanisms should still protect patient interests and provide public accountability.

Transparency and explainability

Some machine-learning systems, especially deep neural networks, are difficult to interpret. Clinicians may know that a model predicts high risk without understanding which factors drove the prediction. A lack of transparency can undermine trust, limit error detection, and complicate informed clinical use.

Explainability should be matched to context. A clinician may need to know the model's intended population, inputs, performance limitations, and recommended response more than they need a complete technical account of its internal mathematics. Regulators and health systems, however, may require deeper documentation of training data, validation methods, uncertainty, and known failure modes. Transparency is therefore not a single feature; it is a set of disclosures tailored to different users.

Automation bias and professional responsibility

Automation bias occurs when people place excessive trust in a system's output, even when it conflicts with other evidence. In healthcare, this could lead to missed diagnoses, inappropriate treatment, or delayed escalation. Conversely, clinicians may also ignore a useful alert if systems are poorly designed or have previously generated too many false positives.

The solution is not simply to remind clinicians to "use judgment." Interfaces, training, escalation pathways, and organizational culture should support appropriate reliance. AI outputs should communicate uncertainty and should not be presented as unquestionable conclusions. Clear responsibility must remain with accountable humans and institutions.

Generalizability and model drift

A model trained in one hospital may perform differently in another because of differences in patient populations, equipment, clinical practice, disease prevalence, or data quality. Over time, performance can also change as workflows, treatments, and coding systems evolve. This phenomenon is often described as model drift.

For this reason, pre-deployment testing should include external validation when possible, and post-deployment monitoring should track safety, effectiveness, calibration, subgroup performance, and unintended consequences. The FDA's work on good machine-learning practice emphasizes lifecycle considerations for AI/ML-enabled medical devices [4]. AI should be treated as a product or service that requires ongoing oversight, not as a finished tool that can be installed and forgotten.

Ethical and Regulatory Frameworks

WHO's guidance provides six central principles for AI in health: protecting human autonomy; promoting human well-being, safety, and the public interest; ensuring transparency, explainability, and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting responsive and sustainable AI [1]. These principles are useful because they extend beyond technical accuracy and focus on the social conditions of trustworthy health AI.

Regulation is also essential, particularly when AI functions as a medical device or influences clinical decisions. The FDA explains that it regulates medical devices, including AI-enabled devices, rather than regulating AI as a general category [2]. This distinction matters because risk depends on intended use. An AI tool that drafts a nonclinical administrative message presents different risks from a tool that analyzes imaging for signs of stroke or provides a treatment recommendation.

Effective governance should include the following elements:

- Clear definition of intended use, patient population, users, and clinical setting.
- Data governance policies addressing provenance, quality, privacy, access, and retention.
- Independent clinical and technical validation before routine deployment.
- Assessment of performance across relevant patient groups and care settings.
- Human oversight procedures and escalation pathways for uncertain or harmful outputs.
- Ongoing monitoring for safety, drift, cybersecurity risks, and equity impacts.
- Transparent communication to clinicians and patients about what the system can and cannot do.
- Clear accountability for vendors, healthcare organizations, and clinicians.

WHO has also highlighted the need for stronger ethics oversight of AI-related health research across the research lifecycle, including study design, ethics review, publication, regulation, and implementation [5]. This reinforces the need to view governance as continuous rather than limited to approval at the start of a project.

Implementation Strategies for Health Systems

A responsible implementation process begins with a real clinical problem, not with the technology. Health systems should ask: What problem is being solved? Who will use the tool? What action will follow its output?

What benefit is expected for patients? What harms are plausible? If these questions do not have clear answers, deployment should be reconsidered.

Second, organizations should involve multiple stakeholders early. Clinicians, nurses, pharmacists, data scientists, informaticians, ethicists, legal experts, security teams, administrators, and patients may each identify different risks and practical needs. Co-design can prevent a tool from being technically impressive but unusable in routine care [4].

Third, performance must be evaluated in the local context. Retrospective accuracy is insufficient. Prospective pilot studies, workflow testing, usability assessment, and monitoring of patient outcomes can reveal whether the tool improves actual care. Important metrics may include time to treatment, diagnostic error, alert burden, clinician workload, patient understanding, disparities in outcomes, and financial effects.

Fourth, health systems should establish governance structures with authority to approve, monitor, pause, or retire AI tools. These structures should maintain an inventory of deployed systems, document intended uses and evidence, monitor incidents, and review updates. An AI tool should be withdrawn or modified if harms emerge or if it no longer performs adequately.

Finally, education is vital. Clinicians need practical understanding of AI limitations, uncertainty, bias, and appropriate reliance. Patients need clear information when AI is materially involved in their care. Literacy does not require every user to become a data scientist; it requires enough understanding to ask appropriate questions and recognize when a system may be unreliable [6].

Future Directions

The future of AI in healthcare will likely involve more multimodal systems that combine text, images, physiological signals, laboratory data, and genomic information. These systems may support a more complete view of a patient's condition than models trained on a single data source. At the same time, more complex systems may be harder to validate and govern.

Generative AI may become increasingly common in documentation, patient communication, education, and research support. Its safe use will depend on verification, data protection, version control, and clear boundaries between administrative assistance and clinical decision-making. Systems that can independently act on patient information or clinical workflows require particularly strong oversight.

Another important direction is the development of evidence standards that focus on clinical impact. The central question should move from "Does the model predict accurately?" to "Does using the model improve outcomes, reduce inequity, and justify its costs and risks?" Answering this question requires prospective studies, real-world monitoring, and transparent reporting.

Global equity should remain central. AI could widen disparities if only well-resourced institutions can build, validate, and maintain safe systems. Conversely, it could support underserved settings if technologies are affordable, locally validated, interoperable, and paired with investment in workforce and infrastructure. The ethical goal is not widespread adoption for its own sake, but meaningful improvement in health for diverse populations.

Conclusion

Artificial intelligence has the capacity to reshape healthcare by supporting image interpretation, clinical decision-making, documentation, research, public health, and operational management. Its value, however, depends on more than algorithmic accuracy. Safe and effective AI requires high-quality and representative data, rigorous validation, transparent communication, human oversight, protection of privacy, attention to equity, and continuous performance monitoring.

Healthcare AI should be designed to augment not displace the expertise, responsibility, and compassion of health professionals. The most credible path forward is one in which developers, regulators, healthcare organizations, clinicians, and patients share responsibility for ensuring that AI systems are useful, safe, fair, and accountable. When these conditions are met, AI can contribute to a more responsive and evidence-informed healthcare system.

References:

1. Panahi U (2026) Exploiting Delay for Stealth: A Delay-Aware False Data Injection Framework for Wireless Sensor Networks. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım Ve Teknoloji, Advanced Online Publication. <https://doi.org/10.29109/gujsc.1934738>
2. Panahi P (2010) The feedback-based mechanism for video streaming over multipath ad hoc networks. Journal of Sciences, Islamic Republic of Iran 21(2).
3. Omid Panahi, Shabnam Dadkhah (2025) Transforming Dental Care: A Comprehensive Review of AI Technologies. J Stoma Dent Res 3(1): 1-5.
4. Panahi O, Eslamlou SF, Jabbarzadeh M. Digitale Zahnmedizin und künstliche Intelligenz. ISBN: 978-620-8-73910-2.
5. Panahi O. (2024) Dental Implants & the Rise of AI. (1):2024.
6. Omid P, Evil Farrokh E (2024) Beyond the Scalpel: AI, Alternative Medicine, and the Future of Personalized Dental Care. J Complement Med Alt Healthcare. 13(2):555860.