Artificial Intelligence Innovates in Healthcare and Balances Risks

Yanyi Sun *

Business School, University of Birmingham, Birmingham, United Kingdom

* Corresponding Author Email: Sunyanyi24@outlook.com

Abstract. Artificial intelligence is profoundly reshaping the global healthcare landscape. From cancer screening to clinical documentation, from drug discovery to public health forecasting, artificial intelligence demonstrates remarkable potential to enhance efficiency, improve diagnostic accuracy, and expand accessibility. It also significantly reduces physicians' administrative burden through automated documentation while advancing telemedicine and equitable healthcare delivery. However, issues such as data privacy breaches, algorithmic bias, limited interpretability of models, erosion of clinical skills, and unclear accountability have raised serious concerns. This paper adopts the framework of "value–risk–governance" to review the applications, challenges, and governance strategies of artificial intelligence in global healthcare. It first summarises the evidence of artificial intelligence's value in medical imaging, personalised treatment, drug discovery, public health prediction, and telemedicine. It then analyses challenges including privacy and security, algorithmic bias, responsibility allocation, ethical dilemmas, and cross-border governance. Finally, it highlights the importance of interdisciplinary integration, international cooperation, and sustainability. It argues that the future of healthcare artificial intelligence must be built on principles of human-centricity, trustworthiness, and equity in order to achieve global health for all.

Keywords: Artificial intelligence; healthcare; algorithmic bias; ethical governance; international standards

1. Introduction

Global healthcare systems are facing mounting challenges. Population ageing has exacerbated the burden of chronic diseases, with cardiovascular disease, cancer, and diabetes accounting for over 70% of global deaths, according to the World Health Organization (WHO). At the same time, healthcare workforce shortages are severe. Organisation for Economic Co-operation and Development (OECD) data show that while high-income countries have an average of 3.5 doctors per 1,000 people, Sub-Saharan Africa has only 0.3, a tenfold disparity. Shortages are particularly acute in radiology and pathology, further widening global health inequalities [1-3].

Artificial intelligence (AI) is seen as a promising solution to these contradictions. With deep learning, natural language processing, and computer vision, AI shows advantages in disease prediction, clinical decision support, medical imaging, drug discovery, and public health management. For example, AI has significantly improved early breast cancer detection rates, achieved dermatologist-level accuracy in skin disease diagnosis, and reduced physician overtime by automating clinical documentation, thereby improving patient experience [4].

Yet, AI is not panacea. Endoscopy studies reveal that excessive reliance on AI may erode clinical skills. Black-box models undermine patient trust due to a lack of interpretability. Training datasets with poor representation exacerbate algorithmic bias. Cross-border data flows without unified governance to create privacy and compliance risks. Thus, striking a balance between value and risk is imperative [5].

In recent years, the international community has actively explored governance strategies. The WHO has published guidelines on AI ethics and governance, emphasizing human rights and fairness. The European Union (EU) has enacted the Artificial Intelligence Act, classifying healthcare AI as a high-risk domain. The Medical Device Action Plan of the United States (US) Food and Drug Administration emphasizes adaptive regulation, while the United Kingdom (UK) has promoted a "Software and Artificial Intelligence as a Medical Device Transformation Plan" with a focus on lifecycle management. China and Japan, meanwhile, have developed policies based on risk control and clinical validation. These developments mark the institutionalization of global AI governance in healthcare [6].

2. Historical Development of AI in Healthcare

The exploration of artificial intelligence in the field of healthcare can be traced back several decades. In the 1960s, Stanford University attempted rule-based reasoning for chemical analysis and bacterial infection diagnosis. However, due to limited data and computing power, these are still laboratory experiments without clinical applications. By the early 2000s, machine learning techniques gained traction in imaging and genomics, but limited datasets and algorithmic performance constrained clinical utility. These early systems were often criticized as "laboratory achievements with little real-world application".

The turning point came around 2010, when deep learning enabled a step-change in performance. The breakthrough of convolutional neural networks in the 2012 ImageNet competition ushered in a golden era of AI-driven medical imaging. The proliferation of electronic health records and wearable devices further expanded data availability, fueling AI development [7].

More recently, generative AI has broadened the horizons of healthcare applications—from automatic clinical summaries to multimodal data integration. Overall, healthcare AI has evolved from rule-based systems to machine learning, deep learning, and now generative models. Reviewing this trajectory helps us appreciate AI's current strengths and limitations while reminding us not to repeat past mistakes of over-optimism and hype [8].

3. The Value and Risks of AI in Healthcare

3.1. Application Value

Medical Imaging: In a randomized controlled trial covering 80000 women, AI supported cancer screening increased early detection rates by 17%. In the field of ophthalmology, the screening of diabetes retinopathy based on artificial intelligence has been approved by the US Food and Drug Administration. This technology enables primary health care institutions and pharmacies to also implement diagnosis, thus significantly improving medical accessibility in rural areas. In addition, artificial intelligence can achieve faster and more accurate stroke detection in CT and MRI scans, effectively reducing treatment delays [9].

In terms of reducing the burden on doctors: Environmental artificial intelligence copywriters reduce the workload of clinical doctors by more than two hours per week, reduce overtime by 40%, and increase doctor-patient communication time. In the emergency department, artificial intelligence can identify critically ill patients 15 minutes earlier than manual triage by continuously monitoring vital signs [10].

In the field of precision medicine, artificial intelligence can integrate genomic, imaging, and clinical data to support personalized treatment. For example, it outperforms traditional methods in predicting the response of lung cancer patients to immunotherapy and achieves early heart failure risk detection by combining electrocardiogram and biomarkers.

In terms of drug discovery, artificial intelligence significantly shortens the drug development cycle. Insilico Medicine has demonstrated that artificial intelligence can complete the process from target recognition to candidate molecules within 18 months, while traditional methods typically take 4 to 6 years. Artificial intelligence has also optimized the recruitment process for clinical trials, reducing bias and improving trial success rates.

In the field of public health: AI models used to accurately predict the transmission trend several weeks in advance during the Coronavirus Disease 2019 (COVID-19) pandemic, providing support for policy intervention. Technology has also been used to predict the trend of tuberculosis and influenza.

In terms of telemedicine and equity: Mobile artificial intelligence ultrasound tools deployed in Africa assist frontline doctors in maternal care, thereby reducing maternal mortality rates. In China, AI assisted Internet hospitals provide medical services to hundreds of millions of people, effectively bridging the medical gap between urban and rural areas.

In terms of mental health and elderly care: Artificial intelligence tools based on natural language processing (NLP) can detect early signs of depression in patients' language patterns, providing support for timely intervention. In aging societies such as Japan, artificial intelligence robots are being explored to assist in elderly care and improve their quality of life.

3.2. Potential Risks

With the widespread popularity of electronic health records and wearable devices, sensitive medical data is becoming increasingly vulnerable to attacks. Currently, the lack of unified global standards seriously hinders cross-border collaboration and data sharing.

In terms of algorithm bias, due to the lack of representativeness in the training data, the diagnostic accuracy of artificial intelligence in the field of dermatology for patients with dark skin is relatively low. Similar biases also affect disease prediction for low-income populations, leading to unfair allocation of medical resources.

The lack of interpretability in black box models seriously undermines users' trust in artificial intelligence. In response to this issue, the EU Artificial Intelligence Act clearly stipulates that high-risk healthcare applications must meet interpretability and traceability requirements.

Research has shown that when AI assisted support is removed, the detection rate of adenomas by doctors significantly decreases, highlighting the risks associated with excessive reliance on AI. At present, the issue of medical liability attribution - whether the responsibility should be borne by doctors, developers, or institutions - has not been clearly resolved. At the same time, ethical issues still exist, mainly reflected in informed consent and transparency of data use. If patients cannot fully understand the use of artificial intelligence, their trust in AI healthcare will weaken.

3.3. Case Studies

Moorfields Eye Hospital and Google DeepMind: Their AI system can diagnose over 50 retinal diseases within seconds, performing on par with specialists, demonstrating the feasibility and the value of cross-sector collaboration. IBM Watson: Despite initial hype in oncology, Watson struggled with clinical relevance due to data mismatches, illustrating the necessity of rigorous clinical validation and clinician involvement.

Babylon Health (UK): AI chatbot services launched with NHS support, aimed to address GP shortages but faced criticism for diagnostic inaccuracies, raising concerns over safety and transparency. China's Internet Hospitals: AI-powered triage and consultations scaled rapidly, especially during the COVID-19 pandemic, showcasing potential in populous countries with significant regional disparities.

Together, these cases highlight both successes and setbacks, underscoring the complexity of real-world healthcare AI and the need to balance technology, ethics, and policy.

4. Governance Strategies for Healthcare AI

Data Governance: Adhere to data minimization principles, employing differential privacy, homomorphic encryption, and federated learning to safeguard patient data. Contextual Adaptation: AI systems should be tailored to local contexts. Lightweight models are essential for low-resource settings. Physicians must retain ultimate decision-making authority, supported by "human-in-the-loop" mechanisms. Some hospitals have instituted "AI-free days" to maintain clinician independence. Moral and regulatory standards: The Future AI Alliance emphasizes fairness, transparency, and stability. The EU requires strict supervision of medical artificial intelligence. The US Food and Drug Administration adopted a flexible and adaptable regulatory model, while the UK emphasizes cross agency coordination. China emphasizes' risk controllability ', while Japan prioritizes clinical validation and trust. Cooperation mode: The "triple helix" cooperation between government, industry, and medical institutions is crucial. Regulatory sandboxes can promote compliance innovation. The evaluation should go beyond short-term efficiency indicators, including long-term clinical outcomes, patient experience, and environmental sustainability.

The WHO advocates for human rights, privacy, and fairness as guiding principles. Governance models diverge: the EU emphasizes high-risk regulation, the US adopts adaptive iteration, the UK stresses cross-sector collaboration, China balances risk control with industrial development, and Japan prioritizes validation and trust-building.

Such diversity complicates international cooperation but also spurs movement toward global standards. BMJ suggests advancing harmonized international standards under WHO and frameworks to enable cross-border data flows and collaborative research. Moving forward, cross-national governance and interdisciplinary integration will be pivotal for sustainable healthcare AI.

5. Conclusion

In conclusion, the rapid advancement of artificial intelligence in healthcare presents both unprecedented opportunities and complex challenges. AI has already demonstrated measurable value in breast cancer screening, clinical documentation, drug discovery, and public health forecasting. Yet, its deployment is shadowed by concerns over data privacy, algorithmic bias, skill degradation, unclear accountability, and regulatory divergence. This dialectical relationship between value and risk suggests that the future of AI in healthcare should not be driven solely by technological breakthroughs but must also be underpinned by social responsibility and robust institutional safeguards.

The path forward requires continuous optimization across four dimensions: human-centricity, trustworthiness, safety, and equity. First, evidence-based medicine and explainability must remain central to ensure transparency and reliability in clinical practice. Second, a stratified governance model is essential, allowing AI applications of varying risk levels to undergo appropriate oversight. Third, interdisciplinary collaboration and international cooperation are indispensable for sustainable development, particularly in establishing global consensus on data standards, ethical norms, and environmentally responsible (“green”) AI. Finally, AI should not replace physicians but instead act as an “augmentative tool,” fostering a model of human–AI symbiosis that enhances care delivery.

Only by navigating the dynamic balance between value and risk can healthcare AI truly achieve a future that is inclusive, equitable, and sustainable, ultimately contributing to the shared goal of global health for all.

References

[1] Eisemann N, Bunk S, Mukama T, Baltus H, Elsner S A, Gomille T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. Nature Medicine, 2025, 31: 917-924.

[2] Duggan M J, Gervase J, Schoenbaum A, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Network Open, 2025, 8(2): e2460637.

[3] Mount Sinai Health System. Bias in generative AI clinical recommendations. Nature Medicine, 2025.

[4] Tursi A, Mocci G, Satta P U, et al. Impact of a symbiotic mixture on moderate-to-severe diverticular disease of the colon. Reviews on Recent Clinical Trials, 2025, 20(1): 9.

[5] World Health Organization. Ethics and governance of AI for health: guidance on large multi-modal models. Geneva: WHO, 2024.

[6] Chapelin J, Voisin A, Rose B, et al. Data-driven drift detection and diagnosis framework for predictive maintenance of heterogeneous production processes: Application to a multiple tapping process. Engineering Applications of Artificial Intelligence, 2025, 139.

[7] Khan U S, Khan S U R. Ethics by design: a lifecycle framework for trustworthy AI in medical imaging from transparent data governance to clinically validated deployment. 2025.

[8] Tsalidis A, Bharadwaj L, Shen F X. Standardization and accuracy of race and ethnicity data: Equity implications for medical AI. PLoS Digital Health, 2025, 4(5).

[9] Mhra. Software and AI as a medical device: change programme. London: MHRA, 2025.

[10] M. J P. Collaborative AI: Artificial Intelligence, Expertise, and Pedagogy in English Language Teaching. University of Toronto Press, 2025.