A MULTIDIMENSIONAL FRAMEWORK FOR EVALUATING ARTIFICIAL INTELLIGENCE IN HEALTHCARE GOVERNANCE

Yurii Cherleniuk
Postgraduate Student at the Department of Public Administration, Governance and Social Work, Shupyk National Healthcare University of Ukraine, Ukraine
e-mail: yuriy1995cherlenyuk@gmail.com, orcid.org/0009-0000-4729-514X

Summary

Artificial intelligence (AI) is increasingly used to support healthcare governance by improving planning, resource allocation, policy development, and organizational decision-making. However, existing approaches to AI evaluation remain largely focused on technical performance and clinical effectiveness, providing limited guidance for assessing AI systems used in healthcare management and public administration. This study aims to develop a multidimensional framework for evaluating the effectiveness of AI in healthcare governance. The study is based on a narrative review and synthesis of contemporary scientific literature, international guidelines, health technology assessment frameworks, and AI governance recommendations. The proposed framework integrates six complementary evaluation dimensions: technical, clinical, organizational, economic, governance, and ethical. For each dimension, key evaluation objectives and recommended indicators are identified based on evidence from international methodological frameworks and recent research. The framework is intended to support healthcare managers, policymakers, researchers, and healthcare organizations in conducting comprehensive assessments of AI systems throughout their lifecycle. Unlike existing approaches that typically evaluate individual aspects of AI implementation, the proposed framework combines multiple evaluation perspectives into a unified model specifically tailored to healthcare governance. The findings contribute to the development of standardized approaches for evaluating AI-enabled decision support systems and may facilitate more evidence-informed adoption of AI technologies in healthcare management.

Key words: artificial intelligence; healthcare governance; healthcare management; evaluation framework; digital health; decision-making; public administration.

DOI https://doi.org/10.23856/7641

1. Introduction

The rapid digital transformation of healthcare has significantly changed decision-making processes and public governance. Among emerging digital technologies, artificial intelligence (AI) has become an important tool for improving planning, resource allocation, epidemiological forecasting, and strategic healthcare management *(Hirani et al., 2024; Habehh & Gohel, 2021)*.

Although AI was initially developed to support clinical practice, its application has expanded to healthcare governance and public administration. AI systems are increasingly used for hospital capacity planning, epidemic intelligence, emergency medical service coordination, healthcare resource allocation, and policy development *(Qian et al., 2021; Quigley et al., 2025; Williams et al., 2025)*.

The growing use of AI in healthcare governance has increased the need for comprehensive evaluation methods. Existing approaches mainly assess technical and clinical performance

using indicators such as accuracy, sensitivity, specificity, predictive value, and patient outcomes. While these measures are essential for evaluating diagnostic and therapeutic AI systems, they do not fully capture the broader organizational and governance-related effects of AI implementation *(Morley et al., 2021)*.

Recent research has emphasized that AI evaluation should extend beyond technical performance to include organizational impact, implementation processes, and healthcare system performance *(Wells et al., 2025)*. Likewise, the World Health Organization and the Organisation for Economic Co-operation and Development recommend evaluating AI with consideration of transparency, accountability, fairness, human oversight, and societal impact *(OECD, 2021; WHO, 2021)*.

The novelty of this study lies in the development of a multidimensional framework that integrates technical, clinical, organizational, economic, governance, and ethical dimensions into a single evaluation model tailored to healthcare governance.

The aim of this study is to develop a multidimensional framework for evaluating the effectiveness of artificial intelligence in healthcare governance. To achieve this aim, the study analyzes existing evaluation approaches, identifies their limitations, determines the key evaluation dimensions reported in the literature, and synthesizes them into a comprehensive framework.

The study is based on a narrative literature review and conceptual synthesis of contemporary scientific publications, international guidelines, Health Technology Assessment frameworks, and AI governance documents. The paper first examines the specific features of AI in healthcare governance and the limitations of existing evaluation approaches, then analyzes the proposed evaluation dimensions, and finally presents the multidimensional evaluation framework.

2. Specific Features of Artificial Intelligence in Healthcare Governance

Artificial intelligence is increasingly used not only in clinical practice but also in healthcare governance. While many AI applications are designed to support diagnosis, treatment, or disease prediction, healthcare organizations and public authorities are increasingly using AI systems to support planning, resource management, and policy development *(Hirani et al., 2024)*.

Healthcare governance differs from clinical care in several important ways. Clinical decisions usually concern individual patients, whereas governance decisions affect healthcare organizations, population groups, and healthcare systems as a whole. As a result, AI systems used in governance are expected to support decisions related to resource allocation, service planning, public health management, and healthcare system performance rather than direct patient care.

One of the defining characteristics of AI in healthcare governance is its system-level orientation. Rather than supporting diagnosis or treatment for individual patients, governance-oriented AI systems are designed to improve planning, coordination, and allocation of healthcare resources. Such systems may support forecasting of healthcare demand, hospital capacity planning, epidemic surveillance, emergency medical service coordination, workforce management, and policy development *(Qian et al., 2021)*. Their primary objective is not direct clinical intervention but the optimization of healthcare system performance.

A second distinguishing feature is the variety of outcomes that governance-oriented AI systems may influence. In clinical settings, AI is often assessed through measures such as diagnostic accuracy or patient outcomes. In contrast, AI systems used in healthcare governance

may affect organizational efficiency, waiting times, resource utilization, service accessibility, and the quality of management decisions. Some of these effects may become visible only after prolonged use, making evaluation more complex *(Cresswell et al., 2025).*

Another important characteristic is the complexity of stakeholder involvement. Clinical AI systems are typically used by healthcare professionals and affect interactions between clinicians and patients. In contrast, governance-oriented AI systems involve a broader range of stakeholders, including policymakers, healthcare administrators, public health authorities, healthcare providers, and the general public. Consequently, the effectiveness of such systems depends not only on technical performance but also on organizational adoption, institutional capacity, regulatory compliance, and stakeholder acceptance. The social impact of governance decisions represents another important distinction. Decisions supported by AI may influence the distribution of healthcare resources, access to services, and the prioritization of healthcare programs. Because such decisions can affect entire populations, international organizations emphasize the importance of transparency, accountability, fairness, and human oversight when AI is used in healthcare governance *(WHO, 2021; OECD, 2021).*

The distinctive characteristics of governance-oriented AI have important implications for its evaluation. Recent studies increasingly recognize that conventional approaches developed for clinical technologies may not be sufficient to assess the broader impact of AI on healthcare systems. *Alami et al. (2020)* argue that artificial intelligence should be viewed not merely as a technological innovation but as a transformative force capable of reshaping healthcare organizations, workflows, and decision-making structures. Similarly, *Morley et al. (2021)* emphasize that the evaluation of AI systems should extend beyond measures of safety and technical performance to include their acceptability, real-world effectiveness, and broader organizational consequences.

This perspective is further reflected in emerging evaluation frameworks for healthcare AI. For example, the FAIR-AI framework highlights the importance of assessing implementation processes, organizational readiness, governance mechanisms, and ongoing monitoring in addition to technical performance *(Wells et al., 2025)*. Likewise, recent Health Technology Assessment (HTA) approaches for AI-based technologies advocate a multidimensional perspective that incorporates clinical, organizational, economic, ethical, and societal considerations *(Di Bidino et al., 2024)*. These developments suggest that evaluating AI in healthcare governance requires a broader conceptual lens than traditional assessments of clinical technologies.

3. Limitations of Existing AI Evaluation Approaches

The rapid adoption of artificial intelligence in healthcare has led to the development of numerous evaluation approaches, most of which were originally designed for clinical applications. These approaches primarily assess technical performance using measures such as accuracy, sensitivity, specificity, and predictive value. Although essential for evaluating diagnostic and predictive models, these indicators provide only a partial understanding of AI's impact on healthcare systems. As noted by *Morley et al. (2021),* existing evaluation frameworks emphasize safety, efficacy, and accuracy while paying less attention to real-world implementation and organizational consequences. Consequently, AI systems with strong technical performance may still fail to deliver meaningful improvements in healthcare organizations.

Another limitation is the tendency to evaluate AI as an isolated technology rather than as part of a broader healthcare system. Recent reviews have highlighted that traditional evaluation approaches often fail to adequately capture organizational, social, and system-level effects

associated with AI implementation in healthcare *(Boverhof et al., 2024)*. From this perspective, frameworks focused exclusively on technical outcomes may overlook important organizational and managerial consequences of AI adoption.

To address some of these challenges, researchers have proposed broader evaluation frameworks. The FAIR-AI framework, for example, emphasizes the importance of implementation readiness, governance processes, monitoring mechanisms, and organizational factors alongside technical performance *(Wells et al., 2025)*. Similarly, recent Health Technology Assessment (HTA) approaches have expanded evaluation criteria to include clinical effectiveness, economic value, ethical considerations, and organizational impact *(Di Bidino et al., 2024)*.

Despite these advances, important limitations remain. Most existing evaluation frameworks have been developed for clinical AI systems or AI-enabled medical technologies. Their primary focus is therefore placed on clinical outcomes, patient safety, implementation, and technology adoption rather than healthcare governance *(Morley et al., 2021; Di Bidino et al., 2024)*. Healthcare governance presents different evaluation challenges. AI systems supporting resource allocation, healthcare planning, public health surveillance, or policy development influence organizational performance, decision quality, equity, and public accountability. These outcomes cannot be adequately assessed using technical or clinical performance indicators alone *(Alami et al., 2020)*.

Another limitation is the lack of a standardized framework specifically designed for healthcare governance. Although recent approaches incorporate organizational, economic, and ethical aspects, they rarely integrate technical, clinical, organizational, economic, governance, and ethical dimensions into a single evaluation model *(Wells et al., 2025; Boverhof et al., 2024)*. As a result, decision-makers still lack a comprehensive framework for assessing whether AI systems generate meaningful value for healthcare organizations and healthcare systems beyond clinical performance.

4. Dimensions of AI Evaluation in Healthcare Governance

The growing complexity of AI applications in healthcare governance requires a broader evaluation perspective than that used for clinical AI systems. Because governance-oriented AI influences organizational performance, resource allocation, decision-making, and public accountability, its evaluation should integrate multiple complementary dimensions rather than rely on technical or clinical indicators alone.

The literature increasingly supports multidimensional evaluation of healthcare AI. Recent frameworks extend assessment beyond technical performance to include implementation processes, organizational readiness, economic value, and ethical considerations *(Wells et al., 2025; Di Bidino et al., 2024)*. However, these dimensions are usually presented within implementation or Health Technology Assessment frameworks rather than from the perspective of healthcare governance. Therefore, a governance-oriented evaluation framework should integrate evidence from existing approaches while reflecting the specific objectives of healthcare management and public administration.

Based on the analysis of the literature, six major evaluation dimensions can be identified: technical, clinical, organizational, economic, governance, and ethical. Together, these dimensions provide a comprehensive basis for assessing the effectiveness of AI systems used to support healthcare governance.

4.1 Technical Dimension. The technical dimension represents the foundation of AI evaluation, as it determines whether an AI system performs reliably, accurately, and consistently

under different operating conditions. Before considering organizational, economic, or governance outcomes, it is essential to establish that the system produces valid and reproducible results. Technical evaluation therefore focuses on the quality of the algorithm itself rather than on its impact after implementation. Traditionally, technical performance has been assessed using measures of predictive accuracy. Depending on the intended application, commonly reported indicators include accuracy, sensitivity, specificity, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUROC). These metrics evaluate the ability of an AI model to correctly classify cases or generate accurate predictions and remain the standard approach for assessing diagnostic and predictive models in healthcare *(Morley et al., 2021).*

However, recent studies emphasize that technical evaluation should extend beyond conventional performance metrics. Modern AI systems are expected to demonstrate robustness, reproducibility, and generalizability across different clinical settings and patient populations. This requires external validation using independent datasets, assessment of performance under varying clinical conditions, and continuous monitoring after deployment to detect performance degradation or model drift *(Wells et al., 2025).* Similarly, the FUTURE-AI international consensus guideline identifies robustness, traceability, and explainability as core technical principles that should be considered throughout the entire AI lifecycle, from development and validation to real-world implementation. The framework also recommends systematic documentation of model development, validation procedures, data provenance, and technical limitations to improve transparency and facilitate independent verification of AI systems *(Lekadir et al., 2025).*

Another important component of technical evaluation is model calibration. While discrimination metrics such as AUROC measure the ability to distinguish between positive and negative cases, calibration assesses whether predicted probabilities correspond to actual clinical outcomes. Poorly calibrated models may provide accurate rankings while systematically overestimating or underestimating risk, potentially leading to inappropriate management decisions *(Wells et al., 2025).*

Recent methodological guidance has also shifted attention toward the quality of evaluation and reporting. The STARD-AI guideline recommends standardized reporting of diagnostic accuracy studies involving AI, including detailed descriptions of datasets, reference standards, validation strategies, handling of missing data, and sources of potential bias. Such reporting improves reproducibility and enables independent assessment of technical performance across different healthcare settings *(Sounderajah et al., 2025).*

4.2. Clinical Dimension. The clinical dimension evaluates whether an AI system contributes to improvements in healthcare delivery and patient outcomes. Although many AI applications used in healthcare governance do not interact directly with patients, they frequently influence clinical services through resource allocation, capacity planning, prioritization of care, and operational decision-making. Consequently, governance decisions supported by AI may indirectly affect healthcare quality, patient safety, and access to medical services *(Di Bidino et al., 2024).*

Clinical evaluation extends beyond measuring the predictive performance of an algorithm. It aims to determine whether AI implementation leads to measurable improvements in healthcare delivery. Depending on the intended application, commonly assessed outcomes include diagnostic accuracy, patient safety, mortality, complication rates, hospital readmissions, length of hospital stay, waiting times, and access to healthcare services *(Wells et al., 2025; Alami et al., 2020).*

Recent methodological guidance recommends evaluating AI systems under real-world clinical conditions rather than relying solely on retrospective validation studies. The DECIDE-AI guideline emphasizes prospective clinical evaluation during the early stages of implementation, including assessment of how AI influences clinical workflows, user interaction, decision-making, and patient care. It also recommends transparent reporting of implementation context, human-AI interaction, unintended consequences, and system modifications throughout the evaluation process *(Vasey et al., 2022).*

Similarly, the FAIR-AI framework recommends continuous monitoring after implementation to ensure that clinical performance remains stable across different healthcare settings and patient populations *(Wells et al., 2025)*. Health Technology Assessment (HTA) approaches further suggest evaluating clinical effectiveness throughout the entire lifecycle of AI technologies, recognizing that clinical performance may change as healthcare systems, patient populations, and clinical practices evolve *(Di Bidino et al., 2024).*

Clinical evaluation should also identify unintended consequences of AI implementation. Even technically accurate systems may fail to improve patient outcomes if they are poorly integrated into clinical workflows, increase clinicians' workload, or introduce new safety risks *(Alami et al., 2020).* Therefore, clinical effectiveness should always be interpreted together with organizational and implementation factors.

Although governance-oriented AI primarily targets system performance, its long-term value should also be reflected in better clinical outcomes and improved quality of care. Therefore, evaluation of the clinical dimension should incorporate indicators related to clinical effectiveness (e.g., diagnostic accuracy and treatment effectiveness), patient outcomes (mortality, complication rates, hospital readmissions), patient safety (adverse events and medication errors), healthcare accessibility (waiting times and access to services), healthcare utilization (length of hospital stay and emergency department visits), and real-world clinical performance assessed through prospective validation and post-deployment monitoring *(Wells et al., 2025; Vasey et al., 2022)*.

4.3. Organizational Dimension. The organizational dimension evaluates how AI affects the functioning of healthcare organizations beyond its technical or clinical performance. Even highly accurate AI systems may fail to deliver value if they cannot be effectively integrated into existing clinical workflows or organizational processes. Therefore, organizational evaluation focuses on the implementation of AI within healthcare settings and its impact on the performance of healthcare organizations *(Alami et al., 2020).*

This dimension assesses whether AI improves the efficiency and quality of healthcare delivery at the organizational level. Important aspects include workflow optimization, staff workload, communication between healthcare professionals, integration with existing health information systems, implementation readiness, and user acceptance *(Wells et al., 2025)*. Successful implementation also depends on organizational leadership, staff training, availability of technical support, and alignment of AI systems with existing clinical and administrative processes.

Recent studies emphasize that organizational success cannot be explained solely by algorithm performance. For example, qualitative evaluation of ambient AI documentation systems has shown that successful implementation depends on how well AI supports clinicians' daily work, reduces administrative burden, and fits naturally into existing workflows. At the same time, challenges related to workflow disruption, trust, usability, and changes in professional responsibilities remain important barriers to adoption *(Duggan et al., 2025).*

Health Technology Assessment frameworks similarly recommend evaluating organizational consequences alongside clinical and economic outcomes. Organizational assessment

should consider implementation processes, resource requirements, workforce adaptation, interoperability, and the long-term sustainability of AI deployment within healthcare organizations *(Di Bidino et al., 2024).*

Based on the reviewed literature, the following indicators are recommended for evaluating the organizational dimension: workflow efficiency, staff workload, user acceptance, implementation readiness, interoperability, integration into clinical workflows, training requirements, organizational adaptability, technical support, and long-term sustainability of AI implementation.

4.4. Economic Dimension. The economic dimension evaluates whether the implementation of an AI system generates sufficient value to justify the resources required for its development, deployment, and maintenance. While technical evaluation determines whether an AI system functions correctly and clinical evaluation examines its impact on patient care, economic evaluation focuses on whether AI represents an efficient use of healthcare resources *(Di Bidino et al., 2024).*

Economic evaluation considers both the costs and the benefits associated with AI implementation. Relevant costs include software acquisition, infrastructure, system integration, maintenance, cybersecurity, staff training, and technical support. Expected benefits may include reduced healthcare expenditures, improved workforce productivity, more efficient use of healthcare resources, shorter hospital stays, and avoidance of unnecessary diagnostic or therapeutic procedures *(Alami et al., 2020).*

Several complementary methods are used to evaluate the economic performance of AI technologies, including cost-effectiveness analysis (CEA), cost-utility analysis (CUA), cost-benefit analysis (CBA), budget impact analysis (BIA), and return on investment (ROI). A systematic review by *El Arab and Al Moosa (2025)* found that although most studies reported favorable economic outcomes, they often underestimated implementation costs, infrastructure investments, and the long-term consequences of AI systems after deployment. The authors therefore recommend lifecycle-oriented economic evaluation that combines conventional health economic methods with continuous assessment of implementation costs and organizational impact.

This perspective is consistent with contemporary Health Technology Assessment frameworks, which recommend integrating economic evidence with clinical, organizational, and ethical evaluation throughout the lifecycle of AI technologies *(Di Bidino et al., 2024).* Similarly, the FAIR-AI framework highlights the importance of continuous monitoring to determine whether expected economic benefits are maintained after implementation and remain proportional to the resources invested *(Wells et al., 2025).*

For healthcare governance, economic evaluation provides evidence to support investment decisions and priority setting under conditions of limited resources. Based on the reviewed literature, the following indicators are recommended for evaluating the economic dimension: implementation costs, operating and maintenance costs, cost-effectiveness, cost-utility, budget impact, return on investment, healthcare resource utilization, productivity gains, and long-term financial sustainability.

4.5. Governance Dimension. The governance dimension evaluates how AI supports transparent, accountable, and evidence-informed decision-making within healthcare systems. Unlike technical or clinical evaluation, governance assessment focuses on the contribution of AI to healthcare management, policy implementation, resource allocation, and institutional performance. This dimension reflects the broader role of AI as a decision-support tool for healthcare organizations and public authorities rather than as a stand-alone clinical technology. Governance evaluation examines whether AI systems improve the quality of management decisions

while maintaining appropriate levels of transparency, accountability, and human oversight. Important aspects include decision traceability, clearly defined responsibilities, compliance with regulatory requirements, stakeholder engagement, and mechanisms for monitoring AI performance throughout its lifecycle *(WHO, 2021; WHO, 2023).*

Recent international frameworks emphasize that governance should be considered a continuous process rather than a one-time assessment. The OECD AI Principles recommend that AI systems operate under clear governance structures that promote transparency, robustness, accountability, and responsible innovation throughout the entire lifecycle of the technology *(OECD, 2021)*. Similarly, the National Institute of Standards and Technology Artificial Intelligence Risk Management Framework proposes continuous identification, assessment, mitigation, and monitoring of AI-related risks to support trustworthy AI implementation *(NIST, 2023).*

For healthcare governance, decision quality represents one of the most important evaluation outcomes. AI systems should support timely, evidence-based, and equitable decisions while preserving appropriate human oversight. Governance evaluation should therefore assess not only whether AI improves operational efficiency but also whether decision-making processes remain transparent, explainable, auditable, and aligned with healthcare policies and regulatory requirements.

Based on the reviewed literature, the following indicators are recommended for evaluating the governance dimension: decision quality, transparency, explainability of management decisions, accountability, human oversight, regulatory compliance, stakeholder engagement, auditability, risk management, policy alignment, and lifecycle monitoring.

4.6. Ethical Dimension. The ethical dimension evaluates whether AI systems are developed, implemented, and used in ways that respect fundamental ethical principles while maintaining public trust in healthcare systems. Unlike technical evaluation, which focuses on algorithm performance, ethical evaluation examines the broader consequences of AI for individuals, healthcare professionals, and society. This dimension is particularly important in healthcare governance, where AI-supported decisions may influence access to healthcare services, distribution of resources, and public health policies *(WHO, 2021).*

Contemporary ethical frameworks consistently emphasize that trustworthy AI should respect the principles of fairness, transparency, accountability, privacy, safety, human autonomy, and human oversight. These principles are intended to ensure that AI systems support rather than replace human decision-making and that their use does not create or reinforce existing inequalities *(WHO, 2021; European Union, 2024).*

Recent regulatory developments have strengthened the role of ethical evaluation throughout the AI lifecycle. The European Union Artificial Intelligence Act introduces a risk-based approach that requires high-risk AI systems, including many healthcare applications, to demonstrate appropriate risk management, transparency, human oversight, technical documentation, data governance, and post-market monitoring before and after deployment *(European Union, 2024).* These requirements illustrate that ethical evaluation is no longer limited to abstract principles but increasingly relies on measurable governance and compliance mechanisms.

For healthcare governance, ethical evaluation also requires consideration of the societal consequences of AI-supported decisions. AI systems should promote equitable access to healthcare services, avoid discriminatory outcomes, protect personal health information, and maintain public confidence in healthcare institutions. International organizations therefore recommend continuous ethical oversight throughout the lifecycle of AI systems rather than limiting ethical assessment to the development stage *(OECD, 2021).*

Based on the reviewed literature, the following indicators are recommended for evaluating the ethical dimension: fairness, transparency, explainability, privacy protection, accountability, human oversight, bias mitigation, data governance, patient autonomy, regulatory compliance, and public trust.

5. A Multidimensional Framework for Evaluating Artificial Intelligence in Healthcare Governance

Existing evaluation approaches assess individual aspects of AI performance, such as technical accuracy, clinical effectiveness, implementation, or health technology assessment, but do not provide a comprehensive framework for healthcare governance. AI systems supporting healthcare planning, resource allocation, policy development, and organizational management require evaluation across multiple complementary dimensions.

Table 1

Proposed Multidimensional Framework for Evaluating Artificial Intelligence in Healthcare Governance

Evaluation dimension	Purpose of evaluation	Recommended indicators
Technical	Technical performance and reliability.	Accuracy, sensitivity, specificity, AUROC, calibration, robustness, reproducibility, explainability, external validation, model drift.
Clinical	Healthcare quality and patient outcomes.	Diagnostic accuracy, mortality, complication rate, patient safety, waiting time, length of hospital stay, healthcare accessibility, real-world clinical performance.
Organizational	Organizational implementation and workflow integration.	Workflow efficiency, staff workload, user acceptance, interoperability, implementation readiness, training requirements, organizational adaptability, long-term sustainability.
Economic	Economic value and efficient resource use.	Cost-effectiveness, cost-utility, budget impact, return on investment, implementation costs, operating costs, productivity gains, healthcare resource utilization, financial sustainability.
Governance	Evidence-informed healthcare management.	Decision quality, transparency, accountability, regulatory compliance, stakeholder engagement, auditability, policy alignment, risk management, lifecycle monitoring.
Ethical	Responsible and trustworthy AI implementation.	Fairness, transparency, explainability, privacy protection, bias mitigation, human oversight, data governance, patient autonomy, regulatory compliance, public trust.

Based on the synthesis of the reviewed literature, a multidimensional framework for evaluating AI in healthcare governance is proposed (see Table 1). The framework integrates six evaluation dimensions identified in previous sections: technical, clinical, organizational, economic, governance, and ethical. It combines evidence from international guidelines, Health Technology Assessment frameworks, implementation models, and AI

governance recommendations into a single evaluation model tailored to healthcare governance.

The proposed framework is intended to support healthcare managers, policymakers, researchers, and healthcare organizations in selecting appropriate evaluation criteria throughout the AI implementation lifecycle. The relative importance of individual dimensions may vary according to the purpose of a specific AI system; however, consideration of all six dimensions enables a more comprehensive assessment than evaluation based solely on technical or clinical performance.

6. Conclusions

This study proposes a multidimensional framework for evaluating artificial intelligence in healthcare governance. The framework integrates six complementary dimensions including technical, clinical, organizational, economic, governance, and ethical aspects of AI implementation. Together, these dimensions provide a comprehensive basis for evaluating AI systems used to support healthcare management and public administration.

The proposed framework may assist healthcare managers, policymakers, researchers, and healthcare organizations in selecting appropriate evaluation criteria throughout the AI implementation lifecycle. It also contributes to the development of a more comprehensive approach to AI evaluation by combining multiple perspectives within a single model. Further research should focus on validating the proposed framework in real healthcare settings and refining evaluation indicators for different types of AI applications.

References

1. Alami, H., Lehoux, P., Auclair, Y., de Guise, M., Gagnon, M. P., Shaw, J., ... & Fortin, J. P. (2020). Artificial intelligence and health technology assessment: anticipating a new level of complexity. Journal of medical Internet research, 22(7), e17707. https://doi.org/10.2196/17707

2. Boverhof, B. J., Redekop, W. K., Visser, J. J., Uyl-de Groot, C. A., & Rutten-van Mölken, M. P. (2024). Broadening the HTA of medical AI: a review of the literature to inform a tailored approach. Health Policy and Technology, 13(2), 100868. https://doi.org/10.1016/j.techsoc.2024.102481

3. Cresswell, K., Williams, R., Dungey, S., Anderson, S., Bernabeu, M. O., Mozaffar, H., ... & Eason, S. (2025). A mixed methods formative evaluation of the United Kingdom National Health Service Artificial Intelligence Lab. npj Digital Medicine, 8(1), 448. https://doi.org/10.1038/s41746-025-01805-w

4. Di Bidino, R., Daugbjerg, S., Papavero, S. C., Haraldsen, I. H., Cicchetti, A., & Sacchini, D. (2024). Health technology assessment framework for artificial intelligence-based technologies. International journal of technology assessment in health care, 40(1), e61. https://doi.org/10.1017/S0266462324000308

5. Duggan, M. J., Gervase, J., Schoenbaum, A., Hanson, W., Howell III, J. T., Sheinberg, M., & Johnson, K. B. (2025). Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Network Open, 8(2), e2460637. https://doi.org/10.1001/jamanetworkopen.2024.60637

6. El Arab, R. A., & Al Moosa, O. A. (2025). Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. NPJ Digital Medicine, 8(1), 548. https://doi.org/10.1038/s41746-025-01722-y

7. European Union. Artificial Intelligence Act (Regulation (EU) 2024/1689). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689
8. Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. Current genomics, 22(4), 291-300. https://doi.org/10.2174/1389202922666210705124359
9. Hirani, R., Noruzi, K., Khuram, H., Hussaini, A. S., Aifuwa, E. I., Ely, K. E., Lewis, J. M., Gabr, A. E., Smiley, A., Tiwari, R. K., & Etienne, M. (2024). Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities. Life, 14(5), 557. https://doi.org/10.3390/life14050557
10. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., ... & Starmans, M. P. (2025). FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. bmj, 388. https://doi.org/10.1136/bmj-2024-081554
11. Morley, J., Morton, C., Karpathakis, K., Taddeo, M., & Floridi, L. (2021). Towards a framework for evaluating the safety, acceptability and efficacy of AI systems for health: an initial synthesis. arXiv preprint arXiv:2104.06910. https://doi.org/10.48550/arXiv.2104.06910
12. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
13. OECD. (2021). State of implementation of the OECD AI Principles: Insights from national AI policies (OECD Digital Economy Papers No. 311). OECD Publishing. https://doi.org/10.1787/1cd40c44-en
14. Qian, Z., Alaa, A. M., & van der Schaar, M. (2021). CPAS: the UK's national machine learning-based hospital capacity planning system for COVID-19. Machine Learning, 110(1), 15-35.
15. Quigley, A., Honeyman, M. D., Stone, M. H., Dawson, R., & MacIntyre, C. R. (2025). EPIWATCH, an artificial intelligence early-warning system as a valuable tool in outbreak surveillance. International Journal of Infectious Diseases, 152, 107579.
16. Sounderajah, V., Guni, A., Liu, X., Collins, G. S., Karthikesalingam, A., Markar, S. R., ... & Ashrafian, H. (2025). The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nature medicine, 31(10), 3283-3289. https://doi.org/10.1038/s41591-025-03953-8
17. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., ... & McCulloch, P. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. bmj, 377. https://doi.org/10.1136/bmj-2022-070904
18. Wells, B. J., Nguyen, H. M., McWilliams, A., Pallini, M., Bovi, A., Kuzma, A., ... & Isreal, M. (2025). A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. NPJ Digital Medicine, 8(1), 514. https://doi.org/10.1038/s41746-025-01900-y
19. Williams, G. S., Koua, E. L., Abdelmalik, P., Kambale, F., Kibangou, E., Nguna, J., ... & Gueye, A. S. (2025). Evaluation of the epidemic intelligence from open sources (EIOS) system for the early detection of outbreaks and health emergencies in the African region. BMC Public Health, 25(1), 857.
20. World Health Organization. (2021). Ethics and governance of artificial intelligence for health. World Health Organization. https://www.who.int/publications/i/item/9789240029200
21. World Health Organization. (2023). Regulatory considerations on artificial intelligence for health. World Health Organization. https://www.who.int/publications/i/item/9789240078871