Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Governance Embedded Agentic Artificial Intelligence Framework for Healthcare Cybersecurity in Resource Constrained Settings

Ibrahim Adabara

adabara.ibrahim@studwc.kiu.ac.ug

Kampala International University

Bashir Olaniyi Sadiq

Kampala International University

Aliyu Nuhu Shuaibu

Kampala International University

Yale Ibrahim Danjuma

Kampala International University

Venkateswarlu Maninti

Kampala International University

Mutebi Joe

Kampala International University

comment

Keywords: Agentic artificial intelligence, healthcare cybersecurity, AI governance, intrusion detection, resource constrained settings, ethical AI, governance by design

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9231871/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Healthcare systems are increasingly exposed to cyber threats that compromise patient safety, data integrity, and service continuity, particularly in resource-constrained environments where governance and cybersecurity capacity remain limited. Although artificial intelligence is widely adopted to enhance intrusion detection and automated response, most existing systems remain reactive, opaque, and insufficiently aligned with ethical and regulatory requirements, creating a critical governance gap. This study addresses this challenge by developing the Agentic Artificial Intelligence Framework, a governance-embedded cybersecurity architecture that integrates ethical reasoning, regulatory compliance, and adaptive intelligence within autonomous decision-making processes. The framework was developed using a design science research approach and evaluated through large-scale simulation involving more than 280,000 network events representing diverse healthcare cyber threat scenarios. The results demonstrate that the framework reduced response time by 34.8 percent, improved system recovery rates by 41.6 percent, and achieved an Ethical Compliance Rate of 0.99 without false escalations. These findings show that governance-by-design can enhance both cybersecurity performance and accountability. The study provides a scalable and policy-relevant approach for strengthening resilient, transparent, and ethically aligned healthcare cybersecurity systems in resource-constrained settings.

1. Introduction

Healthcare systems are increasingly exposed to sophisticated cyber threats that compromise patient safety, data integrity, and continuity of care. The rapid expansion of digital health technologies, including electronic health records, telemedicine platforms, and Internet of Medical Things devices, has significantly increased the attack surface of healthcare infrastructures. These developments have improved service delivery but have also introduced critical vulnerabilities that expose healthcare systems to evolving cyber risks (Algarni & Thayananthan, 2025; Ewoh & Vartiainen, 2024).

Artificial intelligence is increasingly deployed to enhance intrusion detection and automated threat response in healthcare cybersecurity. Machine learning and deep learning approaches have demonstrated improved capability in identifying complex and evolving attack patterns compared to traditional rule-based systems (Ejeofobiri et al., 2024; Huang et al., 2025). However, most existing AI-driven cybersecurity systems remain reactive, relying on historical data rather than anticipating emerging threats. In addition, these systems often operate as opaque models with limited transparency and weak integration of ethical reasoning and regulatory compliance (Adeniran et al., 2024; Emmanuel et al., 2024).

This limitation is particularly significant in resource-constrained settings, where healthcare systems face infrastructural constraints, limited cybersecurity capacity, and dependence on externally developed technologies that may not align with local governance frameworks. These conditions increase systemic

vulnerability and reduce the ability of institutions to ensure accountability, transparency, and policy compliance in automated cybersecurity operations (Cuadros et al., 2025; Folorunso et al., 2024).

To address this gap, this study introduces the Agentic Artificial Intelligence Framework, a governance-embedded cybersecurity architecture that integrates ethical reasoning, regulatory compliance, and adaptive intelligence within autonomous decision-making processes. The framework is grounded in agent-based systems theory, reinforcement learning, and machine ethics, enabling intelligent systems to operate adaptively while remaining aligned with governance constraints (Fard et al., 2023; Wooldridge & Jennings, 1995). Developed using a design science research approach and evaluated through large-scale simulation, the framework demonstrates that governance and technical performance can be jointly optimized. This study makes three contributions. First, it advances a governance-by-design paradigm for artificial intelligence in cybersecurity. Second, it presents a practical and scalable architecture for healthcare cybersecurity in resource-constrained environments. Third, it provides policy-relevant insights for strengthening institutional resilience and accountability in digital health systems.

2. Governance Gap in AI-Driven Cybersecurity

Artificial intelligence is increasingly deployed to enhance intrusion detection and automated threat response within healthcare systems. These systems improve the speed and accuracy of identifying known attack patterns using data-driven techniques. However, most current implementations remain fundamentally reactive, relying on predefined datasets and historical attack signatures rather than anticipating emerging or context-specific threats. This limitation reduces their effectiveness in dynamic healthcare environments, where cyber risks continuously evolve and require adaptive, real-time responses (Ali et al., 2025; Hassan et al., 2025).

A second critical limitation is the opacity of many AI models. Advanced machine learning approaches, particularly deep learning systems, often operate as black-box models that provide limited insight into how decisions are made. This lack of transparency constrains the ability of healthcare institutions to interpret, validate, and audit automated cybersecurity actions. In safety-critical environments, where decisions may involve restricting access to patient data or isolating clinical systems, the absence of explainability introduces significant ethical and operational risks (Marey et al., 2024; Mohale & Obagbuwa, 2025).

More critically, governance mechanisms in most AI-driven cybersecurity systems remain external to the system architecture. Ethical principles, regulatory requirements, and institutional policies are typically enforced through post-deployment audits or human oversight rather than being embedded within the decision-making process. This separation allows systems to achieve high technical performance while operating outside acceptable ethical and regulatory boundaries, limiting accountability and trust (Emmanuel et al., 2024; Sunkara, 2024).

These challenges are particularly pronounced in resource-constrained settings. Healthcare institutions in these environments often depend on externally developed cybersecurity technologies that are not

aligned with local regulatory frameworks or institutional capacities. This dependency limits contextual adaptability and increases exposure to governance and operational risks, especially in systems where cybersecurity failures can directly affect patient safety and service continuity (Folorunso et al., 2024; Njoroge, 2024).

Taken together, the reactive nature of existing AI systems, their limited transparency, and the externalization of governance mechanisms reveal a fundamental weakness in current cybersecurity approaches. This creates a structural accountability gap in which autonomous systems are capable of acting but are not inherently aligned with the ethical, regulatory, and institutional frameworks within which they operate.

3. The AAIF Framework

The Agentic Artificial Intelligence Framework (AAIF) is a governance-embedded cybersecurity architecture designed to integrate ethical reasoning, regulatory compliance, and adaptive intelligence directly within autonomous decision-making processes. Unlike conventional artificial intelligence systems that treat governance as an external constraint, the AAIF operationalizes a governance-by-design paradigm in which ethical and policy considerations are embedded within the core logic of system behavior. This enables cybersecurity systems to not only detect and respond to threats but also ensure that all actions remain aligned with institutional policies, regulatory requirements, and patient safety considerations.

At the architectural level, the AAIF is structured around an agentic decision loop that continuously monitors network activity, evaluates potential threats, and determines appropriate responses in real time. This loop integrates perception, reasoning, decision, and action, allowing the system to adapt dynamically to evolving threat conditions. The decision-making process is supported by adaptive learning mechanisms, including reinforcement learning, which enable continuous improvement while maintaining alignment with predefined governance rules (Fard et al., 2023; Feltus, 2020). The overall architecture of the framework is illustrated in Fig. 1.

A central contribution of the framework is the Governance Gate, which functions as a dual-validation mechanism embedded within the decision pipeline. Each proposed action is evaluated against ethical principles and regulatory constraints before execution. Actions that satisfy governance requirements are implemented autonomously, while those associated with uncertainty, elevated risk, or potential ethical conflict are escalated for human oversight. This mechanism ensures that autonomy operates within clearly defined boundaries, preserving accountability while maintaining operational efficiency (Frenette, 2023; Joseph et al., 2024).

The AAIF is theoretically grounded in the integration of agent-based systems, reinforcement learning, and machine ethics. Agent theory provides the foundation for autonomous and adaptive system behavior, reinforcement learning supports dynamic optimization in adversarial environments, and machine ethics enables the incorporation of moral reasoning into decision-making processes (Nath &

Sahu, 2020; Vishwanath et al., 2024; Wooldridge & Jennings, 1995). This interdisciplinary foundation ensures that the framework balances technical effectiveness with ethical and governance requirements.

To ensure practical relevance, the AAIF embeds policy and regulatory logic directly within its architecture through alignment with established governance frameworks, including the NIST Cybersecurity Framework 2.0 and the HIPAA Security Rule. This integration transforms governance from a retrospective compliance activity into a continuous and enforceable system capability. As a result, the framework provides a scalable and context-aware approach for deploying ethically aligned cybersecurity systems in healthcare environments, particularly in resource-constrained settings.

4. Empirical Validation

The performance of the Agentic Artificial Intelligence Framework was evaluated using a large-scale simulation comprising more than 280,000 network events representing diverse healthcare cybersecurity scenarios, including ransomware, phishing, insider threats, and Internet of Medical Things attacks. The simulation environment was designed to reflect the operational conditions of a Ugandan teaching hospital, incorporating the infrastructural constraints and resource limitations characteristic of healthcare systems in low- and middle-income countries.

The results demonstrate that governance-embedded artificial intelligence can achieve substantial performance improvements without compromising accountability. The framework reduced response time by 34.8 percent and improved system recovery rates by 41.6 percent compared to conventional AI-based cybersecurity approaches. In addition, the framework achieved an Ethical Compliance Rate of 0.99, with no false escalations, indicating consistent alignment with predefined ethical and regulatory constraints. Comparative performance across different cybersecurity approaches is summarized in Table 1.

Table 1
Comparative performance of the proposed AAIF against existing AI-based cybersecurity frameworks across accuracy, ethical compliance, resilience, and decision latency.

Framework	Accuracy	F1 Score	Ethical Compliance Rate	Resilience Index	Decision Latency (ms)
Traditional IDS	0.92	0.90	0.67	0.43	28
Explainable IDS	0.94	0.91	0.81	0.47	32
Trust-Aware IDS	0.95	0.93	0.89	0.49	34
AAIF (Proposed)	**0.97**	**0.95**	**0.99**	**0.51**	35

These findings provide clear evidence that embedding governance mechanisms within AI architectures enhances both cybersecurity effectiveness and institutional accountability in resource-constrained

healthcare environments.

5. Policy Implications and Recommendations

The empirical findings demonstrate that governance and cybersecurity performance can be jointly optimized within a unified artificial intelligence architecture. To translate these findings into practice, four strategic actions are recommended.

5.1 Governance by Design as a Requirement

Governance by design should be established as a mandatory requirement in national cybersecurity strategies and digital health policies. Artificial intelligence systems deployed in healthcare must embed ethical oversight, regulatory compliance, and transparent decision-making within their core architecture rather than relying on post-deployment controls.

5.2 Institutionalization of Regulatory Sandboxes

Regulatory authorities should establish controlled sandbox environments to enable the testing and validation of governance-embedded AI systems under real-world conditions. These environments support iterative evaluation, risk management, and safe deployment before large-scale implementation.

5.3 Investment in Local Capacity Development

Governments and development partners should prioritize investment in local research, technical training, and innovation ecosystems focused on AI governance and healthcare cybersecurity. Strengthening local expertise reduces dependency on external systems and ensures alignment with national regulatory and institutional requirements.

5.4 Strengthening International and Regional Alignment

National policies should align with global and regional governance frameworks to support interoperability, knowledge sharing, and coordinated cybersecurity responses. Harmonized standards enhance cross-border collaboration and reduce fragmentation in interconnected healthcare systems.

6. Conclusion

Healthcare cybersecurity is increasingly challenged by the rapid expansion of digital health systems and the growing sophistication of cyber threats. Existing artificial intelligence approaches, while effective in detection, remain limited by reactive design, lack of transparency, and weak integration of governance mechanisms. These limitations are particularly critical in resource-constrained environments, where institutional capacity, regulatory alignment, and accountability are essential for safeguarding patient safety and maintaining trust. This study introduced the Agentic Artificial Intelligence Framework as a governance-embedded cybersecurity architecture that integrates ethical reasoning, regulatory compliance, and adaptive intelligence within autonomous decision-making processes. The empirical

results demonstrate that governance-by-design enhances both system performance and accountability, confirming that governance is not an optional layer but a foundational requirement for effective AI-driven cybersecurity. The findings provide both a scalable technical solution and actionable policy direction, supporting the development of resilient, transparent, and ethically aligned healthcare cybersecurity systems in resource-constrained settings.

Declarations

Ethics Approval and Consent to Participate

This study did not involve human participants, human subjects, or identifiable personal data. The research was conducted as part of an approved doctoral study. Ethical clearance was obtained from the Kampala International University Research Ethics Committee (KIU-REC) and the Uganda National Council for Science and Technology (UNCST). As no human participants were involved, informed consent was not applicable.

Consent for Publication

Not applicable. This manuscript does not contain any person’s data in any form.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was conducted as part of doctoral research at Kampala International University. No specific external funding was received for this study.

Author Contribution

I.A. conceived the study, developed the governance protocol and Agentic Artificial Intelligence Framework, designed the system architecture, and drafted the original manuscript. B.O.S. supervised the research and contributed to experimental validation and manuscript revision. A.N.S. contributed to methodological development and technical validation. Y.I.D. conducted model training, performed formal analysis, and contributed to results interpretation. V.M. managed data curation and supported experimental implementation and governance integration. J.M. contributed to data curation, experimental configuration, and governance operationalization. All authors reviewed and approved the final manuscript.

Data Availability

The AAIF source code and associated validation datasets, including CICIDS2017, CSE-CIC-IDS2018, and DARPA r6.2, are publicly available at [https://github.com/ibrahimadabara01/aaif-cybersecurity-framework](https:/github.com/ibrahimadabara01/aaif-cybersecurity-framework) . Simulation configurations and governance parameter files are provided within the repository documentation.

References

1. Adeniran AA, Peace A, William P. Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making. World J Adv Res Reviews. 2024;23(3):2447–658. https://doi.org/10.30574/WJARR.2024.23.3.2936.
2. Algarni AM, Thayananthan V. Cybersecurity for Analyzing Artificial Intelligence (AI)-Based Assistive Technology and Systems in Digital Health. Syst 2025. 2025;13(6):439. https://doi.org/10.3390/SYSTEMS13060439. *13*.
3. Ali ML, Thakur K, Schmeelk S, Debello J, Dragos D. Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study. Appl Sci. 2025;15(4). https://doi.org/10.3390/APP15041903.
4. Cuadros DF, Kiragga A, Tu L, Awad S, Bwanika JM, Musuka G. (2025). Unpacking social and digital determinants of health in Africa: a narrative review on challenges and opportunities. *MHealth*, *11*. https://doi.org/10.21037/MHEALTH-24-73
5. Ejeofobiri CK, Victor-Igun OO, Okoye C. AI-Driven Secure Intrusion Detection for Internet of Things (IOT) Networks. Asian J Math Comput Res. 2024;31(4):40–55. https://doi.org/10.56557/AJOMCOR/2024/V31I48971.
6. Emmanuel C, Edima DE, Iboro AE, Joshua OA, Eseoghene DE, Cadet E, Etim ED, Essien IA, Ajayi JO, Erigha ED. Ethical Challenges in AI-Driven Cybersecurity Decision-Making. Int J Sci Res Comput Sci Eng Inform Technol. 2024;10(3):1031–64. https://doi.org/10.32628/CSEIT25113577.
7. Emmanuel Cadet ED, Etim IA, Essien, Eseoghene Daniel Erigha. Joshua Oluwagbenga Ajayi, &. (2024). Ethical Challenges in AI-Driven Cybersecurity Decision-Making. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *10*(3), 1031–1064. https://doi.org/10.32628/CSEIT25113577
8. Ewoh P, Vartiainen T. Vulnerability to Cyberattacks and Sociotechnical Solutions for Health Care Systems: Systematic Review. J Med Internet Res. 2024;26. https://doi.org/10.2196/46904.
9. Fard NE, Selmic RR, Khorasani K. A Review of Techniques and Policies on Cybersecurity Using Artificial Intelligence and Reinforcement Learning Algorithms. IEEE Technol Soc Mag. 2023;42(3):57–68. https://doi.org/10.1109/MTS.2023.3306540.
10. Feltus C. Reinforcement Learning's Contribution to the Cyber Security of Distributed Systems. Int J Distrib Artif Intell. 2020;12(2):35–55. https://doi.org/10.4018/IJDAI.2020070103.

11. Folorunso A, Babalola O, Nwatu CE, Ukonne U. Compliance and Governance issues in Cloud Computing and AI: USA and Africa. Global J Eng Technol Adv. 2024;21(2):127–38. https://doi.org/10.30574/GJETA.2024.21.2.0213.
12. Frenette J. Ensuring human oversight in high-performance AI systems: A framework for control and accountability. World J Adv Res Reviews. 2023;20(2):1507–16. https://doi.org/10.30574/WJARR.2023.20.2.2194.
13. Hassan SADH, Rasheed AA, Mousa AA, Hussein ZA, Ambudkar B. The Rising Cost of Cyberattacks: Trends and Impacts across Industries. HighTech Innov J. 2025;6(2):524–36. https://doi.org/10.28991/HIJ-2025-06-02-011.
14. Huang K, Huang J, Mehmood Y, Atta H, Baig M, Haq MAU. (2025). AAGATE: A NIST AI RMF-Aligned Governance Platform for Agentic AI. *ArXiv*, *abs/2510.2*. https://doi.org/10.48550/ARXIV.2510.25863
15. Joseph SA, Kolade TM, Obioha-Val O, Adebiyi OO, Ogungbemi OS, Olaniyi OO. AI-Powered Information Governance: Balancing Automation and Human Oversight for Optimal Organization Productivity. Asian J Res Comput Sci. 2024;17(10):110–31. https://doi.org/10.9734/AJRCOS/2024/V17I10513.
16. Marey A, Arjmand P, Alerab ADS, Eslami MJ, Saad AM, Sanchez N, Umair M. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nuclear Med. 2024;55(1). https://doi.org/10.1186/S43055-024-01356-2.
17. Mohale VZ, Obagbuwa IC. (2025). A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/FRAI.2025.1526221
18. Nath R, Sahu V. The problem of machine ethics in artificial intelligence. AI Soc. 2020;35(1):103–11. https://doi.org/10.1007/S00146-017-0768-6.
19. Njoroge JW. Comparative Analysis Of Ai Governance In Africa Relative To Global Standards And Practices. IOSR J Comput Eng. 2024;26(5):19–25. https://doi.org/10.9790/0661-2605031925.
20. Sunkara G. Ethical and regulatory implications of AI in cybersecurity surveillance. World J Adv Res Reviews. 2024;24(2):2895–905. https://doi.org/10.30574/WJARR.2024.24.2.3571.
21. Vishwanath A, Dennis LA, Slavkovik M. (2024). Reinforcement Learning and Machine ethics:a systematic review. *ArXiv*, *abs/2407.0*. https://doi.org/10.48550/ARXIV.2407.02425
22. Wooldridge M, Jennings NR. Intelligent agents: theory and practice. Knowl Eng Rev. 1995;10(2):115–52. https://doi.org/10.1017/S0269888900008122.

Figures

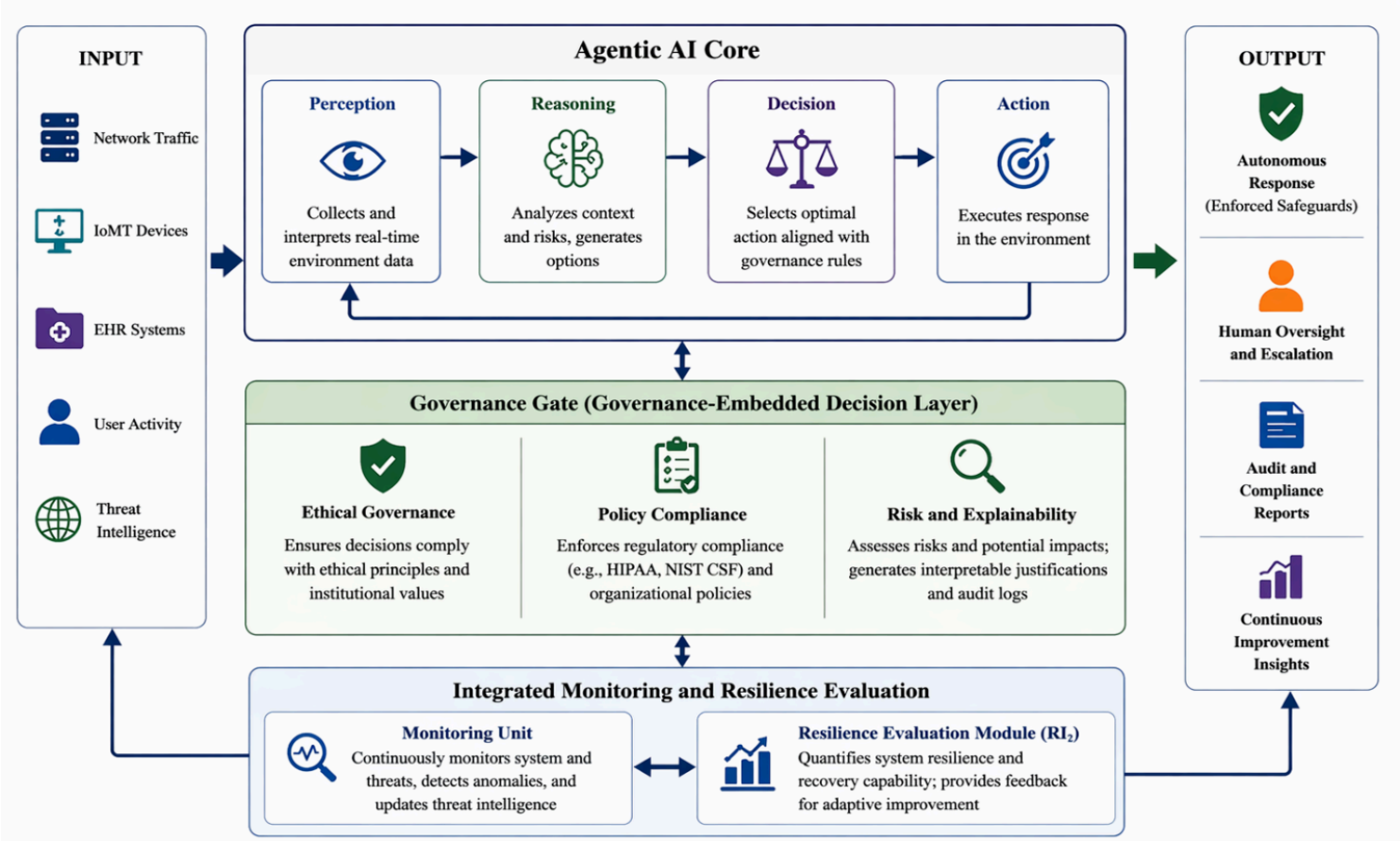


Figure 1. Architecture of the Agentic Artificial Intelligence Framework (AAIF) showing governance-embedded decision-making, adaptive response, and integrated monitoring and resilience evaluation.

Figure 1

Architecture of the Agentic Artificial Intelligence Framework (AAIF) showing governance-embedded decision-making, adaptive response, and integrated monitoring and resilience evaluation.