

Agentic Artificial Intelligence and ethical sovereignty: A framework for global AI governance

Ibrahim Adabara¹, Bashir Olaniyi Sadiq², Aliyu Nuhu Shuaibu³, Yale Ibrahim Danjuma⁴, Venkateswarlu Maninti⁵,
Mutebi Joe⁶

^{1,4,5,6} Department of Computing, Faculty of Science and Technology, Kampala International University.

^{2,3} Department of Electrical, Telecommunication, and Computer Engineering, Kampala International University.

adabara.ibrahim@studwc.kiu.ac.ug¹, bosadiq@kiu.ac.ug² nuhua@kiu.ac.ug³ yaleibrahim@kiu.ac.ug⁴ venkatm@kiu.ac.ug⁵
mutebi.joe@kiu.ac.ug⁶

Corresponding Author: Ibrahim Adabara, adabara.ibrahim@studwc.kiu.ac.ug

Paper history:

Received 29 October
2025

Accepted in revised form
14 December 2025

Keywords

**agentic AI; ethical
sovereignty; AI
governance; Ubuntu
ethics, moral agency;
design science research;
adaptive governance**

Abstract

The growing adoption of artificial intelligence (AI) in critical domains such as healthcare, cybersecurity, and governance expose the limits of current regulatory models, which treat AI as a passive tool rather than a moral participant. Existing frameworks, including the EU AI Act, OECD AI Principles, and AU AI Strategy, emphasize compliance and risk control but overlook AI's emerging autonomy and ethical agency. This paper proposes the Agentic Artificial Intelligence Framework (AAIF), a model that repositions AI as a moral co-governor within socio-technical systems. Grounded in Ubuntu ethics and global governance theory, the AAIF integrates three moral dimensions: autonomy, accountability, and adaptability to support ethically resilient and culturally responsive governance. Developed through the Design Science Research methodology, the framework embeds ethics within institutional and policy design. By harmonizing legal, cultural, and ethical imperatives, the AAIF advances a pluralistic approach to AI governance that strengthens ethical sovereignty, particularly in Global South contexts. Future research will focus on simulation-based and policy-level validation of the framework in applied domains such as healthcare cybersecurity.

Nomenclature and units

AAIF	Agentic Artificial Intelligence Framework
AI	Artificial Intelligence
DSR	Design Science Research
Au	Autonomy
Ac	Accountability
Ad	Adaptability
M	Moral Governance Integrity, expressed as ($M = f(Au, Ac, Ad)$)
EAL	Ethical Anchoring Layer
GCC	Governance Control Core
APL	Accountability Protocol Layer
ARM	Adaptive Resilience Module
CII	Contextual Integration Interface
Ubuntu	African relational philosophy emphasizing interdependence and communal responsibility

1. Introduction

Artificial Intelligence (AI) is increasingly deployed in high-stakes domains such as healthcare, education, finance, and cybersecurity. These systems now make decisions that affect patient care, national security, social equity, and institutional governance. As AI becomes more embedded in critical infrastructures, its capabilities for learning, adaptation, and decision-making have advanced far beyond simple automation. Yet, existing governance frameworks continue to rely on regulatory models that conceptualize AI as a passive object, an entity to be constrained, certified, or risk-classified. Frameworks such as the European Union's AI Act and the OECD AI Principles focus primarily on compliance, technical transparency, and risk mitigation, but offer limited mechanisms for managing the emerging moral agency of AI systems (Chaffer et al., 2025; Ronanki et al., 2023; Sillberg et al., 2025).

This gap reflects a deeper conceptual limitation: traditional AI governance assumes that only humans possess ethical agency, while AI systems are merely instruments. However, contemporary AI increasingly demonstrates features such as context sensitivity, value alignment, and autonomous decision-making. These characteristics challenge static, human-centered ethics protocols and raise the question of whether AI systems themselves should be considered participants in moral reasoning processes (Dobbe, 2025; Gilbert et al., 2023). Without new models that account for AI's evolving autonomy and embeddedness in society, governance strategies risk becoming obsolete or ethically inadequate.

This paper addresses this critical shortcoming by asking the following research questions:

1. How can AI systems be governed as context-sensitive moral agents rather than passive tools?
2. What ethical and institutional principles are required to support the integration of agentic AI in real-world domains?
3. How can relational philosophies, such as Ubuntu, contribute to global AI governance in ways that support ethical sovereignty?

To address these questions, the paper proposes the Agentic Artificial Intelligence Framework (AAIF), a five-layered theoretical model developed using Design Science Research (DSR) methodology. The AAIF repositions AI as a co-governor within socio-technical systems by embedding ethical agency across three interdependent dimensions: autonomy (relational moral reasoning), accountability (distributed ethical traceability), and adaptability (ongoing moral learning and resilience). Grounded in both global governance theory and African relational ethics, the framework contributes a pluralistic and culturally inclusive approach to AI governance.

The remainder of this paper is organized as follows. Section 2 related work. Section 3 outlines the research methodology, grounded in the Design Science Research paradigm. Section 4 reviews the theoretical foundations of the framework, drawing from machine ethics, socio-technical systems theory, and Ubuntu philosophy. Section 5 presents the structure, dimensions, and architecture of the AAIF. Section 6 illustrates the framework's practical application through a use case in healthcare cybersecurity governance. Section 7 discusses the theoretical implications, policy insights, and limitations of the proposed approach. Finally, Section 8 concludes the paper with recommendations for future research and directions for ethical AI integration.

2. Related Study

Recent advancements in AI governance have emphasized the growing need to integrate ethical reasoning and socio-technical accountability within global regulatory systems. Frameworks such as the EU AI Act, OECD AI Principles, and the African Union (AU) AI Strategy have established baselines for responsible AI use. However, these models largely conceptualize AI as a passive object of control rather than an agent capable of moral reasoning. Scholars have noted that such compliance-focused paradigms lack provisions for ethical adaptability and agentic autonomy.

2.1 Global AI Governance Frameworks

The EU AI Act represents a leading attempt to regulate AI based on risk classification, trust, and safety. Nevertheless, researchers argue that this framework, while legally robust, does not sufficiently address moral accountability or AI's emerging cognitive autonomy (Palmieri & Goffin, 2023). Comparative analyses highlight that the EU's model, though centralized and enforceable, struggles to incorporate context-sensitive moral participation and culturally diverse ethics (Cancela-Outeda, 2024). Likewise, both the OECD Principles and the AU AI Strategy emphasize transparency and inclusion but remain limited in operationalizing adaptive moral reasoning and distributed accountability (Chaffer et al., 2025).

2.2 Machine Ethics and Moral Agency

Research in machine ethics has evolved from static, rule-based systems toward relational and adaptive ethical frameworks. Contemporary studies emphasize that AI systems can exhibit context-sensitive reasoning and limited moral co-agency, thereby necessitating governance structures that accommodate moral participation (Adabara et al., 2025b; Hogan & Lasek-Markey, 2024). However, these frameworks continue to be grounded primarily in Western moral philosophies, overlooking collective and non-Western traditions that emphasize interdependence and communal responsibility.

2.3 Socio-Technical and Relational Governance Approaches

Socio-technical governance research argues that moral responsibility must be distributed across both human and artificial agents within dynamic systems. Decentralized oversight models such as ETHOS propose mechanisms for participatory and reflexive AI governance, enabling ethical traceability and shared accountability (Chaffer et al., 2025). Yet, most existing studies lack integration of cultural ethics frameworks, for instance, Ubuntu philosophy, which stresses moral co-production and community-based accountability (Metz, 2020).

2.4 Research Gap and Justification

While these global frameworks mark important progress toward responsible AI, they remain insufficient for addressing AI’s evolving moral agency and the necessity for context-sensitive ethical adaptation. Current scholarship calls for reflexive governance models capable of ethical learning, distributed accountability, and moral sovereignty (Adabara et al., 2025a; Kim et al., 2025). This study responds to these gaps by introducing the AAIF, a design-science-based, multi-layered governance model that integrates autonomy, accountability, and adaptability as co-dependent moral dimensions. Grounded in Ubuntu ethics, the AAIF advances a pluralistic and culturally responsive approach to global AI governance, particularly for Global South contexts where ethical sovereignty and inclusivity are essential.

3. Methodology: Design Science Research Approach

This study adopts the DSR methodology to develop and present the AAIF as a normative, conceptual governance artifact. DSR is particularly suited for research that aims to solve practical, real-world problems through the development of theoretically grounded artifacts such as frameworks, models, or methods (Hevner et al., 2004). Since the AAIF is intended to address a conceptual and ethical gap in global AI governance rather than test hypotheses through empirical data, DSR provides a rigorous foundation for both the construction and validation of this framework.

3.1 Justification for Using DSR

DSR is well-established in information systems and design-based disciplines where the objective is to create artifacts that are both innovative and problem-solving in nature. It supports the design of new constructs or models grounded in both theory and contextual relevance. In the case of this study, the artifact is the AAIF’s response to the lack of governance models that account for AI systems as moral agents within complex socio-technical environments. This makes DSR a suitable methodology for guiding a structured, theory-informed development process (Peppers et al., 2007).

3.2 DSR Stages and Implementation

This study follows the canonical six-stage DSR process to guide the development of the AAIF:

Table 1: DSR Stages and AAIF Development

3.3 Nature of the Study: Conceptual and Normative

This research is a conceptual framework study, not an empirical investigation. It does not include quantitative data collection or statistical analysis. Instead, it develops a normative artifact, the

N/S	DSR Stage	Application to This Study
1	Problem Identification	Fragmented, compliance-focused governance models fail to account for AI’s moral agency.
2	Objective Definition	Develop a governance framework enabling ethical autonomy, accountability, and adaptability.
3	Design of Artifact	Construction of the five-layer AAIF based on ethical theory, Ubuntu philosophy, and policy gaps
4	Demonstration	Application of AAIF in a healthcare cybersecurity use case
5	Evaluation	Conceptual comparison with global frameworks (e.g., EU AI Act, OECD Principles, AU Strategy)
6	Communication	Scholarly dissemination through this journal publication and potential policy forums

AAIF, through philosophical reasoning, design logic, and interdisciplinary synthesis. The framework is proposed as a theoretical contribution intended to guide policy discussions, stimulate ethical reflection, and inform future empirical testing.

3.4 Epistemological and Ethical Grounding

The epistemological foundation of this research is rooted in constructivist realism: while moral concepts and institutional values are socially constructed, the challenges posed by autonomous AI in governance systems are real, systemic, and require structured solutions. The ethical foundation is further shaped by African relational philosophy, especially Ubuntu, which emphasizes interdependence, communal accountability, and moral co-production (Metz, 2020). This orientation supports the pluralistic and decolonial dimensions of the framework, particularly in addressing Global South governance contexts.

3.5 Moral Governance Function: A Conceptual Equation

To express the interdependence of the framework’s ethical dimensions, the AAIF is guided by a moral governance function, expressed as:

M = f(Au, Ac, Ad) (1)

where M denotes moral governance integrity, Au refers to autonomy (relational and bounded decision-making), Ac represents accountability (distributed and explainable moral traceability), and Ad indicates adaptability (context-aware ethical learning and resilience). The function articulates the central claim of the AAIF: that sustainable and ethically sovereign governance of AI systems depends on the dynamic interaction among these three moral components.

Although the function is not quantitatively operationalized at this stage, it serves as a conceptual compass for both policy modeling and architectural design. For example, in a clinical triage scenario, an intelligent system applying the moral governance function may interpret autonomy through patient-centered decision-making that respects individual consent and cultural sensitivity. Accountability would be assessed through the presence of traceable diagnostic logs and system explainability that supports oversight by medical professionals and ethics boards. Adaptability would involve the system's ability to integrate feedback from emerging treatment protocols, shifting epidemiological data, and stakeholder concerns. The moral governance integrity (M) in this case reflects the degree to which these dimensions collectively contribute to ethically robust AI behavior in high-stakes, real-world conditions.

4. Theoretical Foundations

The AAIF is informed by a synthesis of insights from machine ethics, socio-technical governance theory, and African relational philosophy. These perspectives collectively challenge the dominant view of AI as a tool governed solely through external compliance mechanisms and instead promote a shift toward understanding AI systems as evolving participants in moral ecosystems. This section elaborates on the theoretical basis of this shift and provides the conceptual foundation for the AAIF.

4.1 From Machine Ethics to Agentic AI

Machine ethics has traditionally focused on embedding predefined moral rules into AI systems, often grounded in utilitarian or deontological frameworks. While effective for constrained decision environments, these approaches tend to treat AI systems as instruments of human intention rather than active agents within ethical contexts. As AI capabilities expand, enabling systems to learn, adapt, and operate semi-independently, new ethical demands arise. Scholars have called for a move toward relational, agentic models of machine ethics that recognize AI systems as co-constructors of moral meaning within social environments (Gilbert et al., 2023). This reframing positions AI not simply as an object of regulation but as an actor whose actions have normative implications, especially in settings where autonomy and context-sensitive reasoning are central.

4.2 Socio-Technical Governance

The governance of AI must be situated within the broader dynamics of socio-technical systems, complex networks in which human agents, institutions, cultural values, and technologies interact. In these systems, moral outcomes are not the result of isolated choices but of interdependent processes distributed across multiple actors. As AI technologies become more entangled in decision-making infrastructures, ethical responsibility must also become distributed. Socio-technical governance models call for frameworks that are capable of tracing accountability across these networks and ensuring that both human and artificial agents are held to shared moral standards (Chaffer et al., 2025). The dynamic nature of these systems necessitates adaptive ethical approaches that move beyond static compliance toward ongoing moral negotiation and contextual responsiveness.

4.3 Ubuntu Ethics

Ubuntu, a foundational concept in African relational philosophy, provides a culturally grounded alternative to individualistic models of ethical agency. It holds that personhood is constituted through relationships, encapsulated in the phrase “a person is a person through other persons.” In the context of AI, Ubuntu challenges the view of intelligence as merely computational or individual, instead framing it as a capacity that emerges within communities and is accountable to collective well-being. This relational ontology emphasizes mutual responsibility, communal autonomy, and ethical interdependence. By applying Ubuntu to AI governance, researchers have argued for the design of intelligent systems that align with shared social values and actively support human flourishing within diverse cultural settings (Ferlito et al., 2024). This perspective strengthens the normative claim that ethical AI must be embedded within the moral traditions and sovereignty of the communities it serves.

4.4 The Agentic Moral Paradigm

These theoretical lenses of agentic machine ethics, socio-technical governance, and Ubuntu philosophy converge in a new paradigm for AI governance grounded in relational moral agency. Within this paradigm, autonomy is not defined as individual independence but as the capacity for bounded, ethically aware decision-making within a social context. Accountability is understood as a distributed process, where responsibility is co-constructed and traceable across both human and artificial agents. Adaptability refers to the capacity of systems to evolve ethically over time, learning from feedback and adjusting to new cultural, social, or institutional conditions. Together, these three dimensions, autonomy, accountability, and adaptability, form the normative triad that underpins the AAIF. They establish a foundation for governance models that treat AI as a co-governor in moral ecosystems, capable of participating in ethically meaningful ways within diverse institutional and cultural environments.

5. The Agentic Artificial Intelligence Framework

The AAIF is a normative governance model designed to embed ethical agency, accountability, and adaptability into the institutional design of AI systems. Developed using Design Science Research, the framework offers both conceptual clarity and structural depth. It is structured around three core moral dimensions and operationalized through a five-layer architectural model, which responds to limitations in current regulatory paradigms and advances a pluralistic, culturally grounded approach to global AI governance.

5.1 Conceptual Model: Three Core Moral Dimensions

At the core of the AAIF are three interdependent dimensions that define agentic moral capacity in AI systems: autonomy, accountability, and adaptability. These elements are not treated as independent traits but as mutually reinforcing ethical conditions required for resilient and trustworthy governance.

Autonomy refers to the capacity for bounded, relational decision-making. Unlike traditional notions of autonomy grounded in individualism or technical independence, the AAIF emphasizes autonomy as context-sensitive moral reasoning embedded within communal and institutional structures. This form of autonomy allows AI systems to act ethically while remaining aligned with human values and policy constraints.

Accountability represents the capacity to trace, justify, and distribute moral responsibility across both human and artificial agents within a socio-technical system. In the AAIF, accountability is not merely a matter of auditing outputs but involves structured feedback loops that enable ethical traceability across decision-making layers.

Adaptability denotes the ability of AI systems and governance mechanisms to learn from ethical failures, stakeholder feedback, and evolving contextual norms. Ethical resilience, defined here as the capacity for reflective adaptation, ensures that governance systems do not become static or obsolete in the face of moral complexity or cultural variation.

These dimensions are visualized in Figure 1, which illustrates the AAIF's ethical triad. The diagram represents the overlapping nature of autonomy, accountability, and adaptability, with moral governance integrity emerging at their intersection. The figure emphasizes that removing or weakening any of these components undermines the ethical foundation of agentic AI governance.

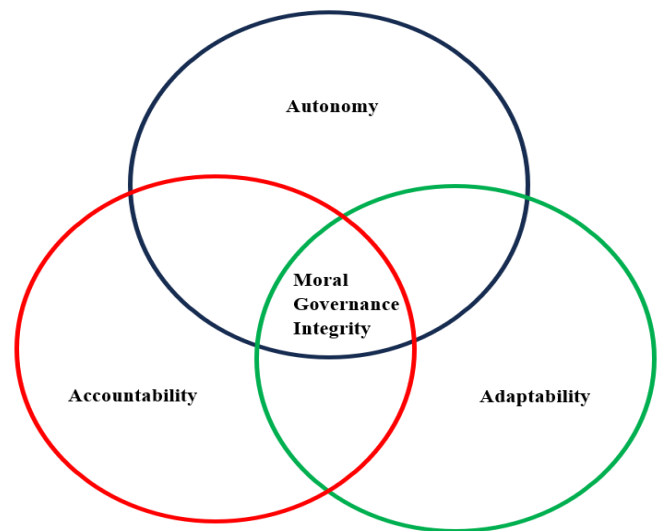


Figure 1. Autonomy–accountability–adaptability triad illustrating the interdependent moral dimensions of the AAIF. Moral governance integrity emerges at their intersection.

5.2 Framework Architecture: Five-Layer Model

To operationalize its ethical foundation, the AAIF is structured into a five-layer architecture, each representing a critical functional zone in the moral governance of AI. These layers are designed to integrate ethical reasoning, institutional responsibility, and contextual responsiveness throughout the lifecycle of intelligent systems.

The **Ethical Anchoring Layer (EAL)** serves as the foundational stratum, encoding shared moral values, cultural principles, and philosophical commitments such as Ubuntu into the system's design parameters and governance ethos. This layer ensures that all technical operations remain grounded in a coherent ethical orientation.

The **Governance Control Core (GCC)** functions as the system's policy enforcement engine. It translates institutional mandates and regulatory requirements into programmable constraints and behavioral expectations, ensuring alignment between governance objectives and system behavior.

The **Accountability Protocol Layer (APL)** facilitates traceability and decision auditing across system actions. This includes metadata tracking, decision logs, and explainability interfaces designed to support both internal review and external accountability to human stakeholders.

The **Adaptive Resilience Module (ARM)** enables ethical learning and reconfiguration in response to new data, feedback, or contextual shifts. This module supports reflexive governance by allowing systems to adjust their ethical parameters over time while maintaining structural coherence.

The **Contextual Integration Interface (CII)** connects the AI system to its operating environment, including cultural norms, domain-specific values, and stakeholder feedback. This layer ensures that moral reasoning remains context-aware and culturally situated, particularly in pluralistic or Global South settings.

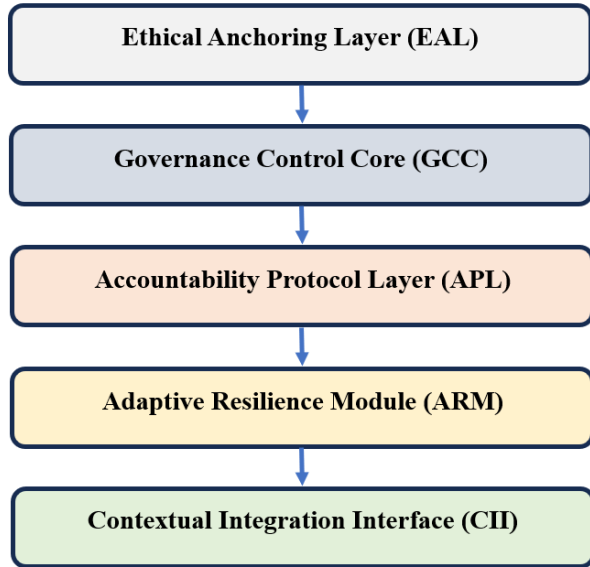


Figure 2. Five-layer architecture of the AAIF, showing vertical integration from ethical anchoring to contextual integration. Each layer supports embedded ethics, accountability, and adaptive governance.

5.3 Comparative Analysis with Existing Frameworks

While several prominent AI governance models have emerged globally, including the European Union AI Act, OECD AI Principles, and the African Union's AI Strategy, they often share common limitations. Chief among these is their tendency to treat AI as a passive entity subject to regulation rather than as an agent embedded in dynamic socio-technical systems.

In comparison, the AAIF emphasizes agency, cultural inclusivity, and adaptive governance. Table 1 summarizes key features of each framework, highlighting how the AAIF differs in terms of ethical anchoring, distributed accountability, and reflexive adaptability.

Table 1: Comparative Features of AI Governance Frameworks

Feature	EU AI Act	OECD AI Principles	AU AI Strategy	AAIF
AI as a moral agent	X	X	X	✓
Ethical/cultural grounding	Limited	Moderate	Moderate	Strong (Ubuntu-based)

Feature	EU AI Act	OECD AI Principles	AU AI Strategy	AAIF
Adaptive feedback mechanisms	X	Limited	X	✓
Distributed accountability	X	Moderate	X	✓
Support for Global South contexts	X	X	✓	✓
Reflexive ethical learning	X	X	X	✓
Embedded ethical architecture	X	X	X	✓

Note: "✓" indicates the feature is present and substantively addressed; "X" indicates absent or underdeveloped.

Further, Table 2 identifies specific governance gaps in these existing models and explains how the AAIF addresses each through its five-layer architecture and ethical triad.

Table 2: Governance Gaps vs. AAIF Responses

Governance Gap	Existing Models	AAIF Response (This Study)
Lack of recognition of AI moral agency	✓	Ethical triad: autonomy, accountability, adaptability
Weak cultural and ethical contextualization	✓	Ethical Anchoring Layer grounded in Ubuntu
Inflexible/static governance mechanisms	✓	Adaptive Resilience Module (ARM) for ongoing ethical learning
Centralized/opaque responsibility structures	✓	Accountability Protocol Layer for distributed moral traceability
Limited Global South integration	✓	Contextual Integration Interface (CII) with regional policy linkage
Absence of reflexive governance loops	✓	Feedback-enabled multi-layered governance architecture

✓ indicates that the governance gap exists in most current global frameworks.

This comparative analysis positions the AAIF not as a replacement but as a complementary and future-oriented model capable of addressing limitations in existing frameworks.

5.4 Limitations of the AAIF

While the AAIF advances a novel and ethically grounded framework for agentic AI governance, it is not without limitations. First, its development is conceptual rather than empirical. Although grounded in philosophical reasoning and comparative analysis, future work is needed to validate the framework through simulations, policy testing, and real-world deployments. Second, institutional adoption of such a multi-layered, ethics-embedded system may face challenges in terms of regulatory inertia, political feasibility, and technical integration. Many existing institutions may lack the capacity or willingness to operationalize relational moral frameworks within their current infrastructures. Third, there is a risk of over-generalization. While the AAIF is designed to be globally relevant and culturally inclusive, its application must remain sensitive to context-specific variables. Local adaptation, stakeholder consultation, and domain-specific tailoring will be essential to avoid ethical or cultural misalignment. Despite these challenges, the AAIF offers a promising step toward reconceptualizing AI governance as a shared moral project, one that values ethical sovereignty, pluralistic inclusion, and adaptive integrity.

6. Use Case Illustration: Healthcare Cybersecurity Governance

The healthcare sector presents one of the most ethically charged and technically complex domains for artificial intelligence deployment. Clinical diagnostics, medical imaging, patient triage, and hospital logistics increasingly depend on AI-driven systems. At the same time, the sector faces persistent cybersecurity threats, data breaches, and algorithmic opacity, especially as health data becomes a valuable target for malicious actors. The ethical stakes in healthcare, ranging from patient autonomy and consent to equity of access and systemic bias, make it a compelling context for testing the viability of agentic governance frameworks like the AAIF. Unlike conventional governance models, which emphasize static compliance or narrow technical oversight, the AAIF enables AI systems to function as moral co-governors, embedded within institutional structures that demand responsiveness, accountability, and contextual sensitivity.

In this use case, the AAIF operates through its core ethical dimensions to address key challenges in healthcare cybersecurity governance. Autonomy within the AAIF allows AI systems, such as diagnostic or triage agents, to make bounded, context-aware decisions without relying solely on hard-coded rules. For example, a machine learning model that identifies anomalous login patterns in an electronic health record (EHR) system can respond dynamically to potential threats while respecting patient data sovereignty and clinical workflows. Rather than simply triggering alerts, the system is capable of justifying its decisions based on ethical priorities embedded in its Ethical Anchoring

Layer. Accountability is supported through structured audit trails and traceability mechanisms within the Accountability Protocol Layer. In a real-time cyber threat scenario, decision paths and data flows can be monitored and reconstructed, enabling oversight by healthcare administrators, ethics boards, and cybersecurity teams. This distributed traceability facilitates legal compliance, stakeholder trust, and institutional resilience. Adaptability is achieved via the Adaptive Resilience Module, which allows the system to incorporate new information, such as novel attack vectors or changing patient privacy norms, into its ethical parameters. This capacity is crucial in environments like hospitals, where threat landscapes evolve rapidly and where fixed governance rules may quickly become obsolete.

These interactions are visualized in Figure 3, which presents a flowchart illustrating the operation of AAIF's five layers within a real-world healthcare cybersecurity context. The diagram demonstrates how ethical values are anchored at the base layer and flow upward through the governance stack, dynamically influencing real-time decision-making, accountability structures, and adaptive responses to emerging threats.

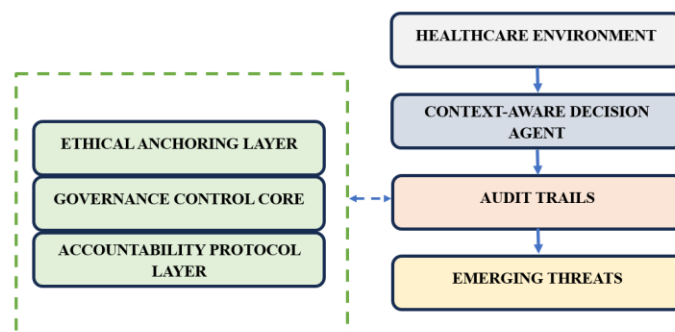


Figure 3. Application of the AAIF in healthcare cybersecurity governance, demonstrating how ethical values flow through the five layers to enable context-sensitive autonomy and real-time accountability.

This case illustrates what distinguishes the AAIF from existing governance frameworks. While models like the EU AI Act or OECD Principles mandate risk classification and data transparency, they do not provide mechanisms for embedding ethical sovereignty or culturally responsive reasoning into AI behavior. Nor do they accommodate moral learning or distributed accountability across non-human agents. By contrast, the AAIF offers an architecture that internalizes moral reasoning, supports cross-cultural value integration, and facilitates adaptive governance. It enables intelligent systems not only to act but to act ethically while remaining traceable, responsive, and culturally grounded. This capacity is particularly vital in healthcare, where decisions directly affect human lives and where institutional legitimacy depends on ethical trust.

7. Implications and Discussion

The AAIF marks a foundational shift in the way AI governance is conceived and practiced. Rather than treating intelligent systems as passive tools of regulation, the framework positions AI as a moral co-governor, embedded within ethical, institutional, and cultural structures. This reconceptualization carries significant implications for theory, policy, and research.

Theoretically, the AAIF challenges traditional compliance-based models of AI ethics that impose moral oversight externally. By embedding ethical reasoning and reflexive feedback into the design architecture itself, the framework promotes a more integrated and relational understanding of moral agency. Drawing on constructivist realism, socio-technical governance theory, and Ubuntu philosophy, it extends the notion of accountability beyond humans to include intelligent systems as co-participants in ethical decision-making. This perspective contributes to the growing discourse on agentic ethics and machine responsibility, offering a basis for more adaptive and distributed models of moral governance.

At the policy level, the AAIF has particular relevance for the Global South, where technological governance must balance global standards with local moral traditions. Its grounding in Ubuntu and adaptive architecture aligns with emerging initiatives such as the African Union's Continental AI Strategy, providing a framework for ethical sovereignty governance that recognizes the moral diversity of communities while ensuring legitimacy and trust. In contrast to the centralized and prescriptive approaches common in Global North policies, the AAIF offers a culturally pluralistic and context-sensitive model that promotes inclusivity and resilience in AI governance systems.

The framework also has practical implications for policy implementation. It can be accompanied by ethical assessment templates, stakeholder engagement guidelines, and adaptive feedback mechanisms designed to integrate moral reasoning into existing institutional infrastructures. Such tools can bridge the gap between theoretical design and operational practice, helping decision-makers navigate complex ethical and regulatory landscapes.

From a research standpoint, the AAIF opens multiple pathways for empirical validation and interdisciplinary collaboration. Future studies may employ agent-based simulations, pilot governance testbeds, or cross-sectoral partnerships to evaluate how the framework's moral dimensions of autonomy, accountability, and adaptability perform in dynamic contexts such as healthcare or cybersecurity. These explorations will refine the framework's operational layers and assess its capacity to sustain ethical integrity in complex socio-technical systems. Overall, the AAIF contributes a pluralistic and ethically integrative model of AI governance. By framing moral agency as a shared enterprise between human and artificial actors, it promotes the co-evolution

of ethical responsibility and institutional legitimacy in an increasingly intelligent world.

8. Conclusion

As artificial intelligence becomes integral to sectors such as healthcare, cybersecurity, and governance, the limits of current compliance-oriented frameworks are increasingly exposed. These models often overlook AI's growing autonomy, contextual awareness, and ethical entanglement within human systems. This paper proposed the Agentic Artificial Intelligence Framework AAIF, a multi-layered model that redefines AI as a moral co-governor rather than a passive tool. Grounded in machine ethics, socio-technical governance, and Ubuntu philosophy, the AAIF operationalizes three interdependent dimensions: autonomy, accountability, and adaptability as the foundation for ethically resilient governance. By embedding these dimensions in a five-layer architecture, the AAIF advances the concept of ethical sovereignty in global AI governance. It offers a culturally inclusive, reflexive framework capable of adapting to diverse values and institutional contexts, particularly within the Global South.

Future research should focus on empirical validation through simulation-based modeling, stakeholder testing, and pilot policy implementation. These efforts will help evaluate how the AAIF performs under dynamic ethical and institutional conditions. Ultimately, the framework seeks to bridge the gap between ethical theory and policy practice, fostering a more inclusive and morally resilient future for AI governance.

Funding statement: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no competing interests.

Ethics Statement: No ethical approval was required for this study.

References

- Adabara, I., Sadiq, B. O., Shuaibu, A. N., Danjuma, Y. I., & Maninti, V. (2025a). A Review of Agentic AI in Cybersecurity: Cognitive Autonomy, Ethical Governance, and Quantum-Resilient Defense. *F1000Research*. <https://doi.org/10.12688/f1000research.169337.1>
- Adabara, I., Sadiq, B. O., Shuaibu, A. N., Danjuma, Y. I., & Maninti, V. (2025b). Trustworthy agentic AI systems : a cross-layer review of architectures , threat models , and governance strategies for real-world deployment. *F1000Research*. <https://doi.org/10.12688/f1000research.169927.1>
- Cancela-Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet Things*, 27.

- <https://doi.org/10.1016/J.IOT.2024.101291>
- Chaffer, T. J., Von Goins Ii, C., Cotlage, D., Okusanya, B., & Goldston, J. (2025). *Decentralized Governance of AI Agents*.
- Dobbe, R. (2025). AI Safety is Stuck in Technical Terms - A System Safety Response to the International AI Safety Report. *ArXiv, abs/2503.0*.
<https://doi.org/10.48550/ARXIV.2503.04743>
- Ferlito, B., Segers, S., De Proost, M., & Mertes, H. (2024). Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. *Science and Engineering Ethics*, 30(4). <https://doi.org/10.1007/S11948-024-00501-4>
- Gilbert, T., Brozek, M. W., & Brozek, A. (2023). Beyond Bias and Compliance: Towards Individual Agency and Plurality of Ethics in AI. *ArXiv, abs/2302.1*.
<https://doi.org/10.48550/ARXIV.2302.12149>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly: Management Information Systems*, 28(1), 75–105.
<https://doi.org/10.2307/25148625>
- Hogan, L., & Lasek-Markey, M. (2024). Towards a Human Rights-Based Approach to Ethical AI Governance in Europe. *Philosophies*, 9(6).
<https://doi.org/10.3390/PHILOSOPHIES9060181>
- Kim, B., Jeong, S., Cho, B.-K., & Chung, J.-B. (2025). AI Governance in the Context of the EU AI Act: A Bibliometric and Literature Review Approach. *ArXiv, abs/2502.0*. <https://doi.org/10.48550/ARXIV.2502.03468>
- Metz, T. (2020). Humility and the African philosophy of ubuntu. *The Routledge Handbook of Philosophy of Humility*, 257–268. <https://doi.org/10.4324/9781351107532-27>
- Palmieri, S., & Goffin, T. (2023). A Blanket That Leaves the Feet Cold: Exploring the AI Act Safety Framework for Medical AI. *European Journal of Health Law*, 30 4, 406–427. <https://doi.org/10.1163/15718093-BJA10104>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77.
<https://doi.org/10.2753/MIS0742-1222240302>
- Ronanki, K., Cabrero-Daniel, B., Horkoff, J., & Berger, C. (2023). RE-centric Recommendations for the Development of Trustworthy(er) Autonomous Systems. *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 23.
<https://doi.org/10.1145/3597512.3599697>
- Sillberg, C. V., De Cerqueira, J. S., Sillberg, P., Kemell, K. K., & Abrahamsson, P. (2025). The EU AI Act is a good start but falls short. *Lecture Notes in Business Information Processing*, 539 LNBIP, 114–130.
https://doi.org/10.1007/978-3-031-85849-9_10