Building Trust in AI: How Ethics, Standards, and Blockchain are Redefining Global Governance

Rachael Njeri Ndung'u
Department of Information Technology
Murang'a University of Technology
Kenya
Email: rndungu [AT] mut.ac.ke

Abstract---- **The rapid advancement of Artificial Intelligence (AI) has amplified concerns around trust, transparency, and accountability in automated decision-making systems. This paper explores the foundational concept of AI with Trust, examining how ethical governance, data integrity, and algorithmic transparency can be strengthened through technological and policy interventions. Drawing from emerging frameworks such as the OECD AI Principles and scholarly insights on trustworthy AI, the study highlights blockchain as a key enabler of verifiable and tamper-proof AI processes. By linking blockchain's decentralized auditability with AI's need for explainability and fairness, the paper argues for an integrated approach that enhances public confidence in AI systems. The discussion positions transparency and accountability as cornerstones of responsible AI adoption and offers a roadmap for aligning innovation with ethical and societal values.**

Keywords--- AI with trust, transparency, explainability, blockchain, ethical governance

I. INTRODUCTION: THE GLOBAL TRUST IMPERATIVE

Artificial intelligence (AI) continues to reshape how societies make decisions, produce knowledge, and deliver services. However, the rapid deployment of AI raises critical questions about trust, fairness, and governance. "AI with trust" is not simply a catchphrase, it represents a global call to design, regulate, and deploy AI in ways that enhance human well-being, respect ethics, and ensure accountability.

This paper explores the emerging discipline of AI governance through the lens of trust. It examines how trust functions as both a policy objective and a governance mechanism. The paper analyses how trust is operationalised through standards, ethics frameworks, and accountability mechanisms such as blockchain. The discussion emphasises that trust in AI must be built across technical, institutional, and societal levels.

It first outlines the conceptual foundations of AI governance, describing its evolution from data governance and digital ethics. It highlights the roles of national governments, industry consortia, and international organisations such as the Organisation for Economic Co-operation and Development (OECD), United Nations Educational, Scientific and Cultural Organisation (UNESCO), and the European Union (EU). Secondly, introduces the emerging models of AI regulation, including risk-based, principles-based, and adaptive governance models, showing how they contribute to trusted AI ecosystems. Then, it explores and analyses how trust-based AI governance can be implemented across sectors trust-building technologies, such as explainable AI (XAI) and blockchain-based audit trails, which make AI decisions transparent and verifiable. Also, it looks into what future research and policy priorities are necessary for sustainable, inclusive, and ethical AI adoption.

In 2024 and 2025, AI governance entered a new phase marked by coordinated international efforts. The [4] Act, enacted in 2024, became the first comprehensive legal framework regulating AI. Simultaneously, countries like Kenya, Singapore, and Canada advanced national AI strategies embedding ethics and trust principles. Recent studies [14], [18] reveal that AI governance now combines traditional regulatory oversight with emerging forms of algorithmic accountability driven by transparency technologies and civic participation.

Trust, therefore, must be treated as a measurable governance outcome, not just a rhetorical ideal. Mechanisms such as algorithmic audits, data lineage verification, and decentralised ledgers ensure that AI systems can be trusted across borders. Blockchain, for instance, can serve as a neutral infrastructure for recording algorithmic decisions, thereby reducing opacity and corruption [23].

At the institutional level, trusted AI demands robust ethics boards, participatory design processes, and inclusive policymaking. Citizen participation is critical, as demonstrated

in UNESCO's 2024 "AI for Humanity" dialogue series, which stressed collective agency in AI development. Social trust is equally vital where citizens must believe that AI systems are aligned with societal values and can be held accountable when harm occurs. Educational institutions and research centres have also started embedding AI ethics and trust modules in curricula to shape responsible future developers [23]. These capacity-building efforts complement legal and technical interventions by fostering ethical literacy and cultural sensitivity in AI innovation.

II. AI WITH TRUST A UNIVERSAL FRAMEWORK

The section expands the discussion to the global governance architecture required to sustain AI with trust. We identify three pillars: (1) harmonisation of standards across jurisdictions, (2) multi-stakeholder governance frameworks, and (3) technological infrastructures that reinforce transparency. Harmonisation prevents regulatory fragmentation that could undermine trust, while multi-stakeholder governance ensures that no single actor monopolises decision-making. Technologies such as distributed ledgers, cryptographic proofs, and explainability tools operationalise transparency in ways that strengthen accountability.

We argue that the future of AI governance lies in integrating ethical design, transparent infrastructure, and collaborative regulation. Building "AI with trust" requires continuous dialogue among governments, academia, civil society, and industry. As AI continues to permeate global systems, governance that values trust will determine whether technology serves humanity or erodes it.

III. OPERATIONALISING TRUSTWORTHY AI IN PRACTICE

Turning the principles of trustworthy AI into everyday practice remains a global challenge. Trust is not merely a product of compliance; it emerges when technology demonstrably aligns with social values such as fairness, transparency, and accountability. Organisations increasingly recognise that implementing governance mechanisms such as codes of conduct, ethical review boards, and algorithmic audit trails, forms the foundation of credible AI adoption [15, 16].

Public institutions have begun translating abstract ethical commitments into measurable processes. The [4] provides one of the most comprehensive regulatory frameworks, introducing obligations based on risk categories such as prohibiting harmful uses, tightly regulating high-risk systems, and promoting voluntary codes for minimal-risk applications. Meanwhile, UNESCO's recommendation on the Ethics of Artificial Intelligence [22, 23] now adopted by more than 190 countries, guides national strategies toward human-centred design and social inclusion.

In practice, operationalising trust involves establishing AI governance frameworks that integrate ethical impact assessments throughout the development lifecycle. For example, the OECD's 2024 AI Policy Observatory highlights how governments in Finland, Singapore, and Canada embed algorithmic accountability measures through cross-agency committees and public reporting. These measures help ensure that algorithmic systems remain auditable, explainable, and aligned with human oversight [16, 17].

In Kenya, the development of the Draft National Artificial Intelligence Policy [33] and updates to the Data Protection Act [34] signify growing alignment with global norms. Kenya's approach emphasizes citizen rights, responsible data use, and multi-stakeholder participation, reflecting lessons from the African Union's Continental Strategy for Artificial Intelligence (2024–2030) [34]; [1].The AU strategy urges African states to strengthen institutional capacities, harmonise ethical standards, and integrate AI into public-sector transformation in ways that safeguard local contexts and indigenous knowledge systems [1].

IV. COMPARATIVE GLOBAL GOVERNANCE PATHWAYS

Across the world, AI governance models differ by political structure, cultural values, and technological maturity. Yet they converge on a shared aim, which is building systems people can trust. Europe has pursued a legally binding model through the AI Act, positioning itself as a global standard-setter in rights-based regulation. The Act's enforcement architecture, which comprise national supervisory authorities and an EU-wide AI Board, demonstrates how regional cooperation can protect citizens while encouraging innovation [5].

North America, by contrast, relies on a hybrid approach. The United States Executive Order on Safe, Secure, and Trustworthy AI [21] mandates federal agencies to apply safety standards without imposing a single nationwide law. Canada's Artificial Intelligence and Data Act (AIDA), expected to take effect in 2025, balances innovation incentives with enforceable obligations for transparency and harm mitigation [7].

In Asia, models vary widely. Japan's Social Principles of Human-Centric AI [28] focus on fostering trust through self-regulation and social responsibility, while China's Interim Measures for the Management of Generative AI Services [27] enforce stringent state oversight emphasizing national security and content control. Singapore continues to refine its Model AI Governance Framework [9] integrating business-friendly ethical guidelines into industry operations.

Within Africa, AI governance remains an evolving landscape. The African Union's Continental AI Strategy [1] calls for regional data-governance frameworks, ethical certification of AI tools, and inclusive participation in standard-

setting bodies. Countries such as Kenya, Rwanda, and South Africa are piloting regulatory sandboxes and public consultations to test adaptive policy tools. Kenya's collaboration with the Smart Africa Alliance [19] demonstrates a growing continental appetite for shared ethical baselines and cross-border cooperation.

Comparatively, the European model prioritizes protection; the American model stresses innovation; Asia's models reflect state-centred pragmatism; and Africa's approach increasingly champions' inclusivity and sustainable development. These differences underline the need for mutual recognition and harmonised global standards to prevent governance fragmentation and regulatory arbitrage [12; 20; 6].

V. Embedding Ethics into AI Design and Deployment

Ethical integration must shift from after-thought to design principle. Ethics-by-design encourages developers to embed moral reasoning into algorithms and data pipelines from the start. This approach reduces the risk of bias, discrimination, and opaque decision-making [11].

Several practical frameworks are now available. The IEEE 7000-2024 standard guides engineers in aligning AI system design with ethical values through traceability and stakeholder feedback. Meanwhile, ISO/IEC 42001:2024 introduces a management-system standard for AI governance, offering globally consistent processes for documenting ethical assurance and performance monitoring [10]

Corporate practice is catching up. Global firms such as Microsoft and Google have established internal "responsible AI" offices to oversee compliance, while African startups are experimenting with explainable machine-learning tools that support fairness audits for local languages and datasets [15]. These efforts suggest a movement from principle to implementation, with ethics forming an innovation asset rather than a compliance burden. Governments can strengthen ethical embedding by incentivising open-source transparency, funding interdisciplinary research, and supporting women and youth participation in AI ethics initiatives. [22, 23] notes that trust flourishes when citizens understand the value choices behind technology. Public-private partnerships and civil-society oversight can therefore play essential roles in holding AI systems accountable.

VI. Building Capacity, Literacy, and Public Confidence

Capacity development is indispensable for trusted AI governance. The OECD AI Skills Outlook [16,17] warns that unequal AI literacy could widen social and economic divides. Countries investing in AI education, regulatory training, and ethical research see stronger institutional readiness and higher public trust.

Kenya's universities, including Murang'a University of Technology, now integrate courses in data governance, machine ethics, and digital policy, reflecting continental priorities for human-capital development [1]. Across Africa, partnerships with the UN Economic Commission for Africa and Smart Africa Digital Academy [19] promote continuous learning for policymakers and technologists alike.

Beyond formal education, public confidence also depends on communication transparency. Engaging communities in algorithmic-impact assessments, inviting civil-society critique, and ensuring that grievance mechanisms are responsive to citizens' concerns are key to sustainable governance [22, 23].

VII. Future Pathways for Global Cooperation and Adaptive Governance

Looking forward, the global AI landscape requires adaptive, multilevel governance capable of keeping pace with technological dynamism. As new generative and autonomous systems emerge, flexibility becomes crucial. Future governance will likely blend binding regulations with voluntary ethical commitments, connected through global coordination forums such as the UN High-Level Advisory Body on AI [30] and the Global Partnership on AI [29].

Key directions include, Interoperable standards that enable cross-border compliance and ethical assurance, AI impact registries for transparency and early risk detection, Ethical certification models that allow consumers and investors to identify trustworthy systems, Participatory policymaking, incorporating citizen assemblies and indigenous perspectives to contextualize governance norms.

For Africa, leveraging continental cooperation through the African Union's planned AI Observatory and regional hubs can ensure representation in setting international norms rather than merely adopting them. Aligning local needs such as agricultural automation, climate monitoring, and public-health analytics with global trust standards will position African innovation as both ethical and competitive [1].

VIII. Ethical, Legal, and Societal Implications of AI Governance

Building trustworthy AI requires a comprehensive framework that balances innovation with accountability, human rights, and inclusivity. Ethical and legal governance of AI is no longer a peripheral debate, rather it is central to sustaining societal trust and global legitimacy. According to [11], over 80 AI ethics guidelines have emerged globally, yet disparities in implementation and enforcement persist, particularly across regions with differing data protection regimes. These disparities highlight the urgent need for harmonized frameworks that transcend national boundaries.

The ethical dimensions of AI governance revolve around fairness, transparency, explainability, and respect for human autonomy. The OECD AI Principles [16, 17] emphasize the need for human-centered values, robustness, and accountability as global standards guiding responsible AI deployment. Similarly, the UNESCO Recommendation on the Ethics of Artificial Intelligence [31] calls for governance models that integrate human rights and social well-being into AI systems' life cycles, promoting global fairness and sustainability Legal implications are equally significant. The [4] represents the world's first comprehensive legal framework categorizing AI systems by risk levels, prohibiting unacceptable applications while regulating high-risk ones. This model introduces obligations for transparency, traceability, and post-market monitoring. Other jurisdictions, such as Canada, Japan, and Kenya, are observing this development closely as they draft their own AI governance strategies [5].

Societal implications of artificial intelligence extend beyond regulation to deeper questions of equity, participation, and public trust. Trustworthy AI cannot exist in isolation from the communities it serves. As [3] observes, data systems often mirror the power relations and social inequalities of the societies that produce them. Therefore, creating AI systems that people can trust requires inclusive governance, where governments, academia, industry, and civil society work together to identify and correct biases while promoting fairness and transparency.

Blockchain technology contributes significantly to this goal by offering technical mechanisms that complement ethical governance. It's decentralised and cryptographically secure design allows for transparent data provenance, immutable audit trails, and verifiable accountability structures. According to [13], blockchain's strength lies in addressing long-standing challenges of information security and data privacy through distributed consensus and encryption, thereby reinforcing public confidence in digital ecosystems.

More recent research underscores blockchain's expanding role in trustworthy AI. In [26] proposes a blockchain-based architecture that provides full traceability and auditing of AI decision processes, ensuring that data used in training and deployment can be independently verified. Similarly, [8] introduce a Stakeholders-in-the-Loop model, in which blockchain supports collective oversight by allowing communities and regulators to verify AI outcomes transparently. Together, these perspectives illustrate that blockchain is not merely a technical safeguard, but a social contract technology that translates ethical principles into enforceable digital accountability.

When integrated thoughtfully, AI and blockchain form a synergistic trust framework, one where fairness and privacy are maintained by design, not as an afterthought. This convergence represents a meaningful step towards operationalizing "AI with trust", bridging the gap between abstract ethical ideals and measurable governance outcomes.

Moreover, public engagement and literacy play critical roles in ensuring that AI governance is not merely top-down but participatory. Initiatives such as AI for Good and Global Partnership on AI (GPAI) illustrate how multi-stakeholder dialogue can promote ethical alignment and policy coherence across borders. These collaborations exemplify the broader notion of "AI with trust" as a paradigm in which technical assurance, ethical responsibility, and social legitimacy converge.

In sum, embedding ethical, legal, and societal considerations into AI governance transforms abstract principles into actionable global standards. Blockchain's integration into these frameworks enhances transparency and traceability, offering a path toward a future where AI operates under verifiable trust, not assumed credibility.

A. Empirical Evidence: Blockchain as a Structural Trust Layer

According to [2] they argue that trust in AI cannot be sustained by algorithmic accuracy alone, but must instead rest on verifiable data integrity, privacy-preserving mechanisms, and transparent accountability domains where blockchain technologies can serve as the technical backbone.

The authors demonstrate that blockchain enhances AI trustworthiness in three key ways: (1)Data Integrity, by creating immutable audit trails for datasets and model training records, blockchain prevents silent tampering and allows regulators and auditors to verify AI provenance. (2) Security, where Blockchain decentralisation eliminates single points of failure and limits insider threats common in centralised AI systems. (3) Privacy, where through zero-knowledge proofs and federated learning on-chain, personal data can remain encrypted and decentralised, supporting compliance with privacy regulations such as GDPR and the Kenya Data Protection Act without hindering AI learning.

In [2] concludes that blockchain should be viewed not as an add-on but as a structural trust layer for AI governance frameworks such as ISO/IEC 42001, NIST AI RMF, and the EU AI Act, enabling "computational trust" that complements institutional oversight.

CONCLUSION

The practice of trustworthy AI governance is no longer an optional moral aspiration but a structural requirement for legitimate digital transformation. Effective governance demands alignment between principles and operations between

what societies value and how systems behave. As seen across regions, Europe's rights-based protection, North America's innovation-centred pragmatism, Asia's state-driven coordination, and Africa's inclusive development ethos all contribute distinctive perspectives. Kenya's emerging frameworks echo these lessons, blending continental solidarity with global ambition. Ultimately, building AI that earns and sustains public trust will depend on inclusive governance ecosystems where governments, industry, academia, and citizens share responsibility for shaping ethical outcomes.

ACKNOWLEDGMENT

The author acknowledges the use of AI assistance (ChatGPT, GPT-5 by OpenAI) in improving the grammar, phrasing, and overall readability of the manuscript. The conceptualization, arguments, analysis, and conclusions are solely the author's original work.

REFERENCES

[1]. African Union. (2024, July). Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Prosperity. African Union Commission. https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf

[2]. Chen, S., Bassi, L., & Ahmed, R. (2025). Building trust in AI: How blockchain enhances data integrity, security, and privacy. IEEE Computer Magazine, 58(2), 63–70. https://doi.org/10.1109/MC.2024.3505012

[3]. Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

[4]. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L (1689), 12 July 2024.

[5]. European Commission. (2024). Regulation (EU) 2024/1680: Artificial Intelligence Act. Official Journal of the European Union. Brussels: European Union.

[6]. Floridi, L., & Cowls, J. (2021). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 31(3), 375–388. https://doi.org/10.1007/s11023-021-09548-8

[7]. Government of Canada, Innovation, Science and Economic Development Canada. (2023, March 13). The Artificial Intelligence and Data Act (AIDA) – Companion Document. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

[8]. Göksal, Ş. İ., & Solarte-Vásquez, M. C. (2024). *The blockchain-based trustworthy AI supported by stakeholders-in-the-loop model (BCTrustAI.SL).* ResearchGate.

[9].Infocomm Media Development Authority. (2024, January). Model AI Governance Framework for Generative AI (Proposed Draft). Singapore: IMDA.

[10]. ISO. (2024). ISO/IEC 42001:2024 – Artificial Intelligence Management System Standard. Geneva: International Organization for Standardization.

[11]. Jobin, A., & Luengo-Oroz, M. (2024). Embedding ethics into AI design: From principles to practice. Nature Machine Intelligence, 6(1), 45–52.

[12]. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[13]. Ndung'u, R. N. (2022). Blockchain as a Solution of Information Security and Data Privacy Issues. Review. International Journal of Computer Applications Technology and Research, 11(08), 337-340.

[14]. Nguyen, T., Li, K., & Mensah, R. (2024). Trustworthy AI governance: Balancing innovation and regulation in emerging economies. Journal of Artificial Intelligence Policy, 9(2), 45–61. https://doi.org/10.1080/jaip.2024.0092

[15]. Plantinga, P., Shilongo, K., Mudongo, O., Umubyeyi, A., Gastrow, M., & Razzano, G. (2024). Responsible artificial intelligence in Africa: Towards policy learning. Data & Policy, 6, e72. https://doi.org/10.1017/dap.2024.60

[16]. Organisation for Economic Co-operation and Development (OECD). (2024, May). *OECD updates AI Principles to stay abreast of rapid technological developments.* Press release. https://www.oecd.org/en/about/news/press releases/2024/05/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.html OECD

[17]. Organisation for Economic Co-operation and Development (OECD). (2024). AI Policy Observatory Annual Report 2024. Paris: OECD Publishing. Retrieved from https://oecd.ai

[18]. O'Neill, C., & Karim, S. (2025). Algorithmic accountability and democratic oversight in AI governance. AI Ethics and Society, 7(1), 12–28. https://doi.org/10.1016/aiethics.2025.01.003

[19]. Smart Africa. (2025). Smart Africa Alliance Annual Review 2025. Kigali: Smart Africa Secretariat.

[20]. Roberts, H., Cowls, J., Morley, J., & Floridi, L. (2023). Global AI governance: Barriers and pathways forward. AI & Society, 38(4), 1207–1222. https://doi.org/10.1007/s00146-022-01556-x

[21]. The White House. (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Washington, D.C.: The White House. Retrieved from https://www.whitehouse.gov

[22]. UNESCO. (2024). Recommendation on the Ethics of Artificial Intelligence: Implementation Status 2024. Paris: United Nations Educational, Scientific and Cultural Organization.

[23]. UNESCO. (2024). AI for Humanity dialogue series: Building collective intelligence and trust.

[24]. UNESCO Digital Policy Report. https://unesdoc.unesco.org/

[25]. Usher, M., & Barak, M. (2024). Unpacking the role of AI ethics online education for science and engineering. International Journal of STEM Education, 11, Article 35. https://doi.org/10.1186/s40594-024-00493-4

[26]. Zhang, Y., & Alami, L. (2024). Blockchain for algorithmic transparency: Emerging frameworks for AI accountability. Global Technology Governance Review, 5(3), 77–94. https://doi.org/10.1080/gtgr.2024.0214

[27]. Zhang, P. (2024). Exploiting Blockchain to Make AI Trustworthy: A Software Architecture for Transparency and Auditing. ACM.

[28]. China Cyberspace Administration. (2024). Interim Measures for the Management of Generative Artificial Intelligence Services. Beijing: Cyberspace Administration of China. Retrieved from https://www.cac.gov.cn

[29]. Japan Cabinet Office. (2024). Social Principles of Human-Centric AI (Revised 2024 Edition). Tokyo: Government of Japan.

[30]. Global Partnership on Artificial Intelligence (GPAI). (2025). Annual Report 2025: Advancing Responsible AI Through Global Collaboration. Paris: GPAI Secretariat. Retrieved from https://gpai.ai

[31]. United Nations. (2024). Interim Report of the UN High-Level Advisory Body on Artificial Intelligence. New York: United Nations. Retrieved from https://www.un.org/ai-advisory-body

[32]. United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO. Retrieved from https://unesdoc.unesco.org

[33]. Ministry of Information, Communications and the Digital Economy (MICDE). (2024). Draft National Artificial Intelligence Policy of Kenya. Nairobi: Government of Kenya. Retrieved from https://ict.go.ke

[34]. Office of the Data Protection Commissioner (ODPC). (2025). Data Protection Act (Amendment) 2025: Strengthening Digital Trust and AI Governance. Nairobi: Government of Kenya. Retrieved from https://www.odpc.go.ke