

Transparency and Explainability for Artificial Intelligence (AI): Ensuring that AI Systems and their Decision-Making Processes can be Understood and Audited by Humans

Kariuki Muigua

Transparency and Explainability for Artificial Intelligence (AI): Ensuring that AI Systems and their Decision-Making Processes can be Understood and Audited by Humans

Table of Contents

Abstract.....	3
1.0 Introduction	3
2.0 Impact of Artificial Intelligence on Transparency and Explainability	5
3.0 Incorporating Transparency and Explainability into Artificial Intelligence Systems and Decision-Making Processes	6
4.0 Conclusion	8
References	9

Transparency and Explainability for Artificial Intelligence (AI)?: Ensuring that AI Systems and their Decision-Making Processes can be Understood and Audited by Humans

Kariuki Muigua*

Abstract

This paper examines how transparency and explainability can be integrated into Artificial Intelligence (AI). The paper observes that transparency and explainability are fundamental tenets of the rule of law that prevent arbitrary exercise of powers by ensuring that decision-making processes are understood by the affected parties. However, the paper notes that algorithmic decision-making through AI can violate the transparency and explainability requirements under the rule of law. The paper discusses how AI impacts transparency and explainability. In light of this concern, the paper advocates for reforms towards ensuring that AI systems and their decision-making processes can be understood and audited by humans in order to uphold the rule of law requirements of transparency and explainability.

1.0 Introduction

The growth of Artificial Intelligence (AI) provides opportunities towards bolstering the rule of law. The United Nations defines the rule of law as a principle of governance in which all persons, institutions and entities, both public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards¹. It has been observed that under the rule of law, all public powers act within the constraints set out by law, in accordance with the values of democracy and fundamental rights, and under the control of independent and impartial courts². This concept envisages ideal societies where no one is above the law; everyone is treated equally under the law; everyone is held accountable to the same laws; there are clear and

* PhD in Law (Nrb), SC, FCI Arb (Chartered Arbitrator), OGW, LL. B (Hons) Nrb, LL.M (Environmental Law) Nrb; Dip. In Law (KSL); FCPS (K); Dip. in Arbitration (UK); MKIM; Mediator; Consultant: Lead expert EIA/EA NEMA; BSI ISO/IEC 27001:2005 ISMS Lead Auditor/ Implementer; ESG Consultant; Advocate of the High Court of Kenya; Professor of Environmental Law and Conflict Management at the University of Nairobi, Faculty of Law; Member of the Permanent Court of Arbitration (PCA) [July, 2026].

¹ United Nations., 'What is the Rule of Law.' Available at <https://www.un.org/ruleoflaw/what-is-the-rule-of-law/> (Accessed on 23/07/2026)

² European Commission., 'What is the Rule of Law?' Available at https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/rule-law/what-rule-law_en (Accessed on 23/07/2026)

fair processes for enforcing laws; there is an independent judiciary; and human rights are guaranteed for all persons³.

It has been observed that AI is being widely embraced in governance to interpret, implement and enforce regulations towards advancing the rule of law⁴. AI systems are being utilised to enhance efficiency in administration processes, prevent arbitrary decisions, and strengthen the enforcement and protection of human rights in accordance with the rule of law⁵. In addition, it has been observed that AI tools and systems are being leveraged to strengthen access to justice and enhance the efficiency of judicial administrations towards upholding the rule of law⁶.

However, despite its efficacy in promoting the rule of law, if not harnessed appropriately, ethically and responsibly, AI can undermine the fundamental tenets of the rule of law including transparency, explainability, accountability and due process⁷. It is therefore necessary to inculcate the tenets of the rule of law including transparency and explainability into AI in order to effectively harness its transformative potential.

This paper examines how transparency and explainability can be integrated into AI. The paper observes that transparency and explainability are fundamental tenets of the rule of law that prevent arbitrary exercise of powers by ensuring that decision-making processes are understood by the affected parties. However, the paper notes that algorithmic decision-making through AI can violate the transparency and explainability requirements under the rule of law. The paper discusses how AI impacts transparency and explainability. In light of this concern, the paper advocates for reforms towards ensuring that AI systems and their decision-making processes can be understood and audited by humans in order to uphold the rule of law requirements of transparency and explainability.

³ American Bar Association., 'Rule of Law.' Available at https://www.americanbar.org/groups/public_education/resources/rule-of-law/ (Accessed on 23/07/2026)

⁴ Smuha. N.A., 'AI, the Rule of Law, and Algorithmic Rule by Law' Available at <https://www.icconnectblog.com/ai-the-rule-of-law-and-algorithmic-rule-by-law/> (Accessed on 23/07/2026)

⁵ Ibid

⁶ United Nations Educational, Scientific and Cultural Organization., 'Artificial Intelligence and the Rule of Law' Available at <https://www.unesco.org/en/artificial-intelligence/rule-law> (Accessed on 23/07/2026)

⁷ Judiciary of Kenya., 'Judiciary to Leverage AI to Enhance Justice' Available at <https://judiciary.go.ke/judiciary-to-leverage-ai-to-enhance-justice/> (Accessed on 23/07/2026)

2.0 Impact of Artificial Intelligence on Transparency and Explainability

Transparency and explainability are fundamental elements of the rule of law. For instance, transparency ensures that decision-making processes are open, inclusive and comprehensible⁸. It has been observed that transparency empowers citizens to exercise their fundamental rights, hold decision-makers accountable, and participate meaningfully in decision-making processes⁹. Transparency ensures open communication, informed decision-making and accountability, which play a vital role towards good governance¹⁰. Explainability on the other hand seeks to ensure that those affected by decisions understand how and why such decisions are made¹¹. It has been observed that explainability is key towards ensuring trust, transparency, and accountability in decision-making processes while giving those affected by a particular decision a sound basis to challenge such decisions¹².

Fostering transparency and explainability is therefore vital towards strengthening the rule of law and governance processes. With AI being widely integrated into governance processes, upholding transparency, explainability and accountability has become a pertinent concern towards ensuring that this technology is harnessed in a safe, ethical and appropriate manner. For example, it has been observed that algorithmic decision-making through AI makes it difficult to understand how decisions are made which undermines transparency, accountability, and due process requirements under the rule of law¹³. The secretive nature of algorithmic decision-making through AI raises pertinent concerns that can violate transparency¹⁴. In particular, it has been observed that inadequate transparency measures can lead to human right violations through biased AI models¹⁵. Algorithmic decision-making in hiring processes, healthcare, finance, law enforcement and justice systems has often resulted in biased and discriminatory

⁸ Open Government Partnership., 'Transparency' Available at <https://www.opengovpartnership.org/glossary/transparency/> (Accessed on 23/07/2026)

⁹ Ibid

¹⁰ Transparency: Definition, Meaning and Why it Matters., Available at <https://researchcentre.trtworld.com/dictionary/transparency/> (Accessed on 23/07/2026)

¹¹ AI Explainability., Available at <https://verifywise.ai/lexicon/ai-explainability> (Accessed on 23/07/2026)

¹² Ibid

¹³ Burgess. P., 'AI and the Rule of Law' Available at <https://www.bloomsbury.com/us/ai-and-the-rule-of-law-9781509963171/> (Accessed on 23/07/2026)

¹⁴ Key Human Rights Challenges of AI., Available at <https://ennhri.org/ai-resource/key-human-rights-challenges/> (Accessed on 23/07/2026)

¹⁵ What Does Transparency Really Mean in the Context of AI Governance?., Available at <https://www.oceg.org/what-does-transparency-really-mean-in-the-context-of-ai-governance/> (Accessed on 23/07/2026)

outcomes¹⁶. Without adequate transparency safeguards, the use of AI can therefore have negative impacts on society including through violating human rights¹⁷.

In addition, it has been observed that due to limited understanding about how AI models are trained, the data they are trained on and how they determine outcomes, explainability is a pertinent concern when harnessing AI in governance processes¹⁸. It has been correctly noted that since AI is having a huge influence over critical decisions in key sectors including healthcare, finance and access to justice, it is imperative for users to understand its outputs for justice and human rights¹⁹. Lack of explainability in AI decision-making processes can result in legal challenges, reduced user adoption, or model rejection²⁰. AI has been described as a '*black box*' since in most cases, it is difficult to explain, manage and regulate its outcomes due to the technology's increasing complexity²¹.

In light of the foregoing concerns, it is imperative to inculcate transparency and explainability into AI systems and their decision-making processes in order to promote the rule of law. In particular, it has been observed that since the rule of law seeks to prevent arbitrary exercise of powers, allowing AI to make decisions unchecked by humans violates this cardinal principle²². It is therefore necessary to ensure that AI systems and their decision-making processes can be understood and audited by humans towards fostering transparency and explainability requirements under the rule of law.

3.0 Incorporating Transparency and Explainability into Artificial Intelligence Systems and Decision-Making Processes

There is need to inculcate transparency and explainability into AI systems and decision-making processes towards the safe, ethical and appropriate adoption of this powerful technology. It has been observed that transparency is key towards unlocking the full potential of AI by fostering a clear understanding of AI tools and systems including their capabilities and limitations²³. By fostering transparency, it is possible to ensure clarity and openness about how AI models and systems function, make decisions, and affect

¹⁶ Ibid

¹⁷ Ibid

¹⁸ What is AI Transparency?., Available at <https://www.ibm.com/think/topics/ai-transparency> (Accessed on 23/07/2026)

¹⁹ AI Explainability., Op Cit

²⁰ Ibid

²¹ What is AI Transparency?., Op Cit

²² Burgess. P., 'AI and the Rule of Law' Op Cit

²³ Organisation for Economic Co-operation and Development., 'AI Principles' Available at <https://www.oecd.org/en/topics/sub-issues/ai-principles.html> (Accessed on 23/07/2026)

users and stakeholders²⁴. Consequently, it has been argued that transparency forms the basis towards building trust, ensuring accountability, and fostering a culture of ethical and appropriate use of AI and other technologies²⁵.

Embracing explainability is also key towards the safe and ethical use of AI. It has been correctly observed that for users and other stakeholders to trust that AI is making effective, accurate and fair decisions, they need visibility into how these models operate, the logic algorithmic decision-making and how AI models are evaluated for accuracy and fairness²⁶. Explainability also seeks to ensure that users know more about the data used to train AI models, including data sources and how data is processed, weighted and labeled²⁷. Through explainability, it is possible to understand algorithmic decision-making by clearly describing how and why AI systems and models make a specific decision or prediction²⁸. Explainability therefore allows users to comprehend and trust the decisions and predictions made by AI models²⁹. Through explainability, it is possible to ensure that AI decision-making processes are accurate, fair, non-biased and uphold human rights³⁰.

In order to foster transparency and explainability, it is imperative to ensure that AI systems and their decision-making processes can be understood and audited by humans. Users and organizations have been urged to adopt explainable AI systems and models that provide the rationale for their decision-making processes towards ensuring accuracy, transparency and accountability³¹. Further, there is need for developers to design AI systems with inherent transparency safeguards in order to enable users understand the logic behind algorithmic decision-making³². In addition, when AI systems are deployed in decision-making processes in key sectors such as healthcare, finance, hiring, education and justice systems, informing parties on why such tools are

²⁴ What Does Transparency Really Mean in the Context of AI Governance?., Op Cit

²⁵ Ibid

²⁶ What is AI Transparency?., Op Cit

²⁷ Ibid

²⁸ AI Explainability., Op Cit

²⁹ What is Explainable AI?., Available at <https://www.ibm.com/think/topics/explainable-ai> (Accessed on 23/07/2026)

³⁰ Ibid

³¹ What Does Transparency Really Mean in the Context of AI Governance?., Op Cit

³² Ibid

adopted, how they are utilised and the data relied on is key towards ensuring transparency, explainability, accountability and procedural fairness³³.

Human audits are also key towards ensuring that AI systems are transparent and explainable. Through human audits, it is possible to detect and address risks such as algorithmic biases, errors, discrimination and human right violations towards the safe, ethical and appropriate use of AI³⁴. It has been observed that human audit fosters transparency and addresses errors in AI systems towards building trust between technology and society³⁵. Human audits are strategic tools that can identify vulnerabilities and errors in AI systems, promote human rights, meet regulatory requirements, and build trust in the adoption of AI³⁶. Human oversight of AI systems and their decision-making processes including through regular audits and measures for review and appeal of AI decision-making is therefore vital towards its safe and ethical adoption in a transparent and explainable manner³⁷.

4.0 Conclusion

In an era marked by widespread adoption of AI, ensuring that this technology adheres to the tenets of the rule of law is key towards its safe, ethical and appropriate use³⁸. In particular, by fostering transparency and explainability, it is possible to ensure that AI is effectively harnessed in governance processes towards strengthening the rule of law. Achieving this goal involves ensuring that AI systems and their decision-making processes can be understood and audited by humans.

³³ United Nations Educational, Scientific and Cultural Organization., 'Guidelines for the use of AI systems in courts and tribunals' Available at <https://unesdoc.unesco.org/ark:/48223/pf0000396582> (Accessed on 23/07/2026)

³⁴ The crucial role of humans in AI oversight., Available at <https://www.cornerstoneondemand.com/resources/article/the-crucial-role-of-humans-in-ai-oversight/> (Accessed on 23/07/2026)

³⁵ Ibid

³⁶ What is an AI Audit?., Available at <https://www.ibm.com/think/topics/ai-audit> (Accessed on 23/07/2026)

³⁷ Global Arbitration Review., 'Artificial intelligence in arbitration: evidentiary issues and prospects' Available at <https://globalarbitrationreview.com/guide/the-guide-evidence-in-international-arbitration/3rd-edition/article/artificial-intelligence-in-arbitration-evidentiary-issues-and-prospects> (Accessed on 23/07/2026)

³⁸ Burgess. P., 'AI and the Rule of Law' Op Cit

References

AI Explainability., Available at <https://verifywise.ai/lexicon/ai-explainability>

American Bar Association., 'Rule of Law.' Available at https://www.americanbar.org/groups/public_education/resources/rule-of-law/

Burgess. P., 'AI and the Rule of Law' Available at <https://www.bloomsbury.com/us/ai-and-the-rule-of-law-9781509963171/>

European Commission., 'What is the Rule of Law?' Available at https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/rule-law/what-rule-law_en

Global Arbitration Review., 'Artificial intelligence in arbitration: evidentiary issues and prospects' Available at <https://globalarbitrationreview.com/guide/the-guide-evidence-in-international-arbitration/3rd-edition/article/artificial-intelligence-in-arbitration-evidentiary-issues-and-prospects>

Transparency and Explainability for Artificial Intelligence (AI): Ensuring that AI Systems and their Decision-Making Processes can be Understood and Audited by Humans

Judiciary of Kenya., 'Judiciary to Leverage AI to Enhance Justice' Available at <https://judiciary.go.ke/judiciary-to-leverage-ai-to-enhance-justice/>

Key Human Rights Challenges of AI., Available at <https://ennhri.org/ai-resource/key-human-rights-challenges/>

Open Government Partnership., 'Transparency' Available at <https://www.opengovpartnership.org/glossary/transparency/>

Organisation for Economic Co-operation and Development., 'AI Principles' Available at <https://www.oecd.org/en/topics/sub-issues/ai-principles.html>

Smuha. N.A., 'AI, the Rule of Law, and Algorithmic Rule by Law' Available at <https://www.icconnectblog.com/ai-the-rule-of-law-and-algorithmic-rule-by-law/>

The crucial role of humans in AI oversight., Available at <https://www.cornerstoneondemand.com/resources/article/the-crucial-role-of-humans-in-ai-oversight/>

Transparency: Definition, Meaning and Why it Matters., Available at <https://researchcentre.trtworld.com/dictionary/transparency/>

United Nations Educational, Scientific and Cultural Organization., 'Artificial Intelligence and the Rule of Law' Available at <https://www.unesco.org/en/artificial-intelligence/rule-law>

United Nations Educational, Scientific and Cultural Organization., 'Guidelines for the use of AI systems in courts and tribunals' Available at <https://unesdoc.unesco.org/ark:/48223/pf0000396582>

United Nations., 'What is the Rule of Law.' Available at <https://www.un.org/ruleoflaw/what-is-the-rule-of-law/>

What Does Transparency Really Mean in the Context of AI Governance?., Available at <https://www.oceg.org/what-does-transparency-really-mean-in-the-context-of-ai-governance/>

What is AI Transparency?., Available at <https://www.ibm.com/think/topics/ai-transparency>

What is an AI Audit?., Available at <https://www.ibm.com/think/topics/ai-audit>

Transparency and Explainability for Artificial Intelligence (AI): Ensuring that AI Systems and their Decision-Making Processes can be Understood and Audited by Humans

What is Explainable AI?., Available at
<https://www.ibm.com/think/topics/explainable-ai>