

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

Kariuki Muigua

Table of Contents

Abstract.....	3
1.0 Introduction	3
2.0 Artificial Intelligence Errors: Causes and Effects	4
3.0 Building Artificial Intelligence Liability Frameworks for Justice and Equity	6
4.0 Conclusion.....	7
References	8

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

Kariuki Muigua*

Abstract

This paper discusses how AI errors can be effectively avoided and addressed through liability frameworks. The paper observes that despite its positive benefits, the wide use of AI brings along with it the risk of errors and mistakes which can violate human rights. It discusses some of the common AI errors and their impact on human rights and development. Due to the risk of errors and mistakes in AI, the paper argues that there is need to build liability frameworks in order to effectively harness this transformative technology while mitigating its risks and challenges. In order to achieve this goal, the paper suggests how effective AI liability frameworks can be designed for justice and equity.

1.0 Introduction

The emergence of Artificial Intelligence (AI) is a pivotal moment for the global community not only a technological breakthrough but also a powerful tool that can accelerate progress towards the 17 Sustainable Development Goals (SDGs)¹. AI is a powerful technology that is transforming industries, fostering innovation and redefining how we solve complex global problems². Through appropriate approaches and policies, AI has the potential to advance human rights, unlock economic development, protect the environment and foster social progress³.

AI can therefore foster progress towards Sustainable Development for the benefit of both humanity and nature. For example, AI is improving healthcare, finance, education, law enforcement, access to justice, agriculture, and environmental governance including climate action and biodiversity conservation thus fostering progress towards the SDGs for people and planet⁴. Consequently, it has been suggested that with targeted cooperation, public-private investments and appropriate governance, the era of AI offers countries a chance to leapfrog traditional development stages towards building more prosperous, vibrant, resilient, sustainable and equitable societies⁵.

However, despite its transformative potential, the development and adoption of AI is associated with several risks and challenges that can undermine the quest for just, inclusive, fair and equitable societies.

* 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 Global Compact., 'Artificial Intelligence and the Sustainable Development Goals: Operationalizing technology for a sustainable future' Available at <https://unglobalcompact.org/compactjournal/artificial-intelligence-and-sustainable-development-goals-operationalizing> (Accessed on 17/07/2026)

² Ibid

³ Ibid

⁴ United Nations., 'Artificial Intelligence (AI)' Available at <https://www.un.org/en/global-issues/artificial-intelligence> (Accessed on 17/07/2026)

⁵ United Nations Development Programme., 'AI for Development: A Positive Tipping Point for People & Planet' Available at <https://www.undp.org/indonesia/blog/ai-development-positive-tipping-point-people-planet> (Accessed on 17/07/2026)

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

The United Nations points out that while AI has the potential to address some of humanity's greatest challenges, the rapid AI development inevitably brings with it a host of risks and challenges that need to be carefully navigated⁶. Consequently, it has been correctly opined that there is need for adequate safeguards in order to harness the transformative potential of AI while tackling its risks and challenges⁷.

This paper discusses how AI errors can be effectively avoided and addressed through liability frameworks. The paper observes that despite its positive benefits, the wide use of AI brings along with it the risk of errors and mistakes which can violate human rights. It discusses some of the common AI errors and their impact on human rights and development. Due to the risk of errors and mistakes in AI, the paper argues that there is need to build liability frameworks in order to effectively harness this transformative technology while mitigating its risks and challenges. In order to achieve this goal, the paper suggests how effective AI liability frameworks can be designed for justice and equity.

2.0 Artificial Intelligence Errors: Causes and Effects

Despite its transformative potential, AI can cause errors and mistakes if not carefully harnessed. For instance, it has been observed that since humans are the centre of development and deployment of AI, human decisions are where most AI errors and mistakes originate⁸. In particular, it has been correctly observed that since developers determine how AI models are designed, the data they are trained on, and how these systems are tested, incomplete or biased data sets can lead to errors, mistakes and discriminatory outcomes⁹. Due to biased data sets, AI technologies including facial recognition and language modelling have shown prejudice against racial and ethnic minorities, leading to injustices and discrimination through false arrests and accusations¹⁰.

In addition to developers, it has been observed that organisations deploying AI models without sufficient safeguards can create the risk of mistakes and errors¹¹. In particular, it has been observed that deploying AI models in high-risk areas such as healthcare, justice systems and financial services without determining their accuracy can result to flawed decisions with severe consequences¹². For example, AI errors in healthcare can have devastating impacts on the human right to health including through wrong diagnoses¹³. In addition, it has been observed that AI mistakes and errors can worsen existing biases in justice systems, particularly regarding race, sex or socioeconomic status potentially violating human rights including

⁶ United Nations., 'Artificial Intelligence (AI)' Available at <https://www.un.org/en/global-issues/artificial-intelligence> (Accessed on 17/07/2026)

⁷ Ibid

⁸ Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Available at <https://www.oneadvanced.com/resources/who-is-responsible-for-ai-mistakes/> (Accessed on 18/07/2026)

⁹ Ibid

¹⁰ Key Human Rights Challenges of AI., Available at <https://ennhri.org/ai-resource/key-human-rights-challenges/> (Accessed on 18/07/2026)

¹¹ Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Op Cit

¹² Ibid

¹³ Pujari. S et al., 'Artificial intelligence for global health: cautious optimism with safeguards' Available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC10225938/pdf/BLT.23.290215.pdf> (Accessed on 18/07/2026)

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

through false arrests or denial of bail¹⁴. Further, AI errors in financial services have seen people being excluded from loan eligibility on basis of their race or gender¹⁵.

Apart from errors flowing from human decisions, the technical architecture of AI can also lead to mistakes. For instance, it has been observed that AI models learn from data, identify patterns and generate outputs without fixed rules¹⁶. Consequently, these models can behave unpredictably, especially in unfamiliar situations leading to errors and mistakes¹⁷. There have been instances where AI models have made reference to non-existent legal cases, created new languages, performed insider trading, spread fake news and polarizing content, encouraged users to break the law, and given harmful health advice, demonstrating that these systems can cause errors and mistakes especially in unfamiliar and unpredictable conditions where training data is limited¹⁸.

In light of these concerns, it has been observed that AI can lead to hallucinations which are incorrect or misleading results that AI models generate¹⁹. AI hallucination has been described as a phenomenon where AI models generate convincing, contextually coherent but entirely fabricated or biased outcomes that are independent of the user's input or previous context²⁰. Due to the phenomenon of hallucination, the responses generated by AI may seem plausible, but in real sense, they are incorrect, misleading and biased²¹. It has been observed that AI hallucinations are often caused by several factors, including insufficient training data, incorrect assumptions made by AI models, or biases in the data used to train such models²².

The high incidence of errors in AI is a negative situation that undermines the effective use of this technology for development. These errors can undermine human rights especially due to algorithmic biases in key sectors including justice systems, law enforcement and financial services²³. Algorithmic biases also fuel injustices, inequalities and discrimination against vulnerable and marginalized populations including people of colour, women and indigenous and local communities²⁴. In addition, AI errors in healthcare can jeopardize human life and the right to health²⁵. Further, AI hallucinations including historical inaccuracies, wrong financial information, scientific inaccuracies, and flawed legal advice can have devastating effects on

¹⁴ Organisation for Economic Co-operation and Development., 'Governing with Artificial Intelligence' Available at https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/ai-in-justice-administration-and-access-to-justice_f0cbe651.html (Accessed on 18/07/2026)

¹⁵ Olurunju. N., 'African algorithmic governance: Benefit of a community-based approach' Available at <https://researchictafrica.net/2022/04/03/african-algorithmic-governance-benefit-of-a-community-based-approach/> (Accessed on 18/07/2026)

¹⁶ Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Op Cit

¹⁷ Ibid

¹⁸ When AI goes wrong: 13 examples of AI mistakes and failures., Available at <https://www.evidentlyai.com/blog/ai-failures-examples> (Accessed on 18/07/2026)

¹⁹ What are AI Hallucinations?., Available at <https://cloud.google.com/discover/what-are-ai-hallucinations> (Accessed on 18/07/2026)

²⁰ Ozer. M., 'Is Artificial Intelligence Hallucinating?' Available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC11681264/> (Accessed on 18/07/2026)

²¹ Ibid

²² What are AI Hallucinations?., Op Cit

²³ Olurunju. N., 'African algorithmic governance: Benefit of a community-based approach' Op Cit

²⁴ Ibid

²⁵ Pujari. S et al., 'Artificial intelligence for global health: cautious optimism with safeguards' Op Cit

people relying on this technology for key decision-making in vital sectors such as finance, business, justice and healthcare²⁶. Tackling AI errors is therefore necessary towards harnessing this transformative technology for development while minimizing its risks and challenges.

3.0 Building Artificial Intelligence Liability Frameworks for Justice and Equity

Due to the high incidence of mistakes and errors caused by a combination of factors such as human decision-making and AI hallucinations, it is imperative to build accountability and liability frameworks for justice and equity. AI mistakes and errors can have devastating impacts including human rights violations and social injustices through algorithmic biases and discrimination²⁷. Tackling AI errors and mistakes is therefore key towards harnessing the transformative potential of this technology for development and prosperity.

Fostering accountability and liability is vital towards the safe, ethical, responsible, inclusive and appropriate development and use of AI. It has been observed that through accountability and liability, it is possible to ensure that AI systems benefit all people without discrimination, respect human rights, are fair, transparent and explainable, and are robust, secure and safe²⁸. In particular, it has been observed that sound accountability measures ensure that all actors govern and manage risks throughout their AI's lifecycle ranging from planning and design, to data collection and processing, to model building and validation, to deployment, operation and monitoring²⁹. By strengthening accountability and building liability frameworks, it is possible to ensure that AI is developed, deployed, and utilized in such a manner that responsibility for bad outcomes can be assigned to liable parties³⁰. In particular, due to the opacity and complexity of algorithmic decision-making, liability frameworks are key towards ensuring that AI errors and mistakes are adequately addressed towards fostering accountability³¹.

In order to build effective AI liability frameworks, it is necessary to adopt shared responsibility³². Given the dynamic nature of this powerful technology and the number of actors involved, treating accountability and liability as the responsibility of one actor is not effective since mistakes and errors can arise at any stage of AI's lifecycle. Consequently, embedding ethical considerations at every stage of AI can ensure that those who violate such requirements are held liable³³. For example, developers can be held liable for negligence in the design and misrepresentation of the capabilities of AI especially where models are developed biased and discriminatory data sets that can lead to errors³⁴. In addition, organisations deploying AI, including financial institutions, schools, health providers and justice systems, can be held liable for

²⁶ AI Hallucinations—Understanding the Phenomenon and Its Implications., Available at <https://www.coursera.org/articles/ai-hallucinations> (Accessed on 18/07/2026)

²⁷ Olurunju. N., 'African algorithmic governance: Benefit of a community-based approach' Op Cit

²⁸ Organisation for Economic Co-operation and Development., 'Advancing Accountability in AI: Governing and Managing Risks throughout the Lifecycle for Trustworthy AI' Available at https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf (Accessed on 18/07/2026)

²⁹ Ibid

³⁰ AI Accountability., Available at <https://www.carnegiecouncil.org/explore-engage/key-terms/ai-accountability> (Accessed on 18/07/2026)

³¹ Ibid

³² Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Op Cit

³³ Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Op Cit

³⁴ Ibid

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

negligence in deployment especially where such deployment is done without adequate testing and transparency and accountability measures³⁵. Liability can also flow to end users of AI systems since they have a responsibility to question the accuracy of AI tools, understand the limitations of this technology and validate recommendations in order to avoid poor decision-making in vital sectors including healthcare, financial services and justice processes³⁶.

Effective liability frameworks therefore recognize that all actors throughout the AI's lifecycle can be responsible for its errors and mistakes. Consequently, it has been suggested that there is need to build sound liability frameworks including through rules and regulations that promote shared accountability among all actors, give parties affected by AI errors and mistakes an avenue for appeal and review of such decisions through liability claims, and ensure adequate damages depending on the source and gravity of AI errors and mistakes³⁷.

4.0 Conclusion

AI is a powerful and transformative technology that can accelerate progress towards Sustainable Development. However, without adequate safeguards, AI can also cause errors and mistakes that can violate human rights and worsen social injustices. It is therefore imperative to effective liability frameworks in order to avoid and address AI errors for justice, equity and development.

³⁵ Ibid

³⁶ Ibid

³⁷ Artificial Intelligence Liability Directive., Available at <https://www.ai-liability-directive.com/> (Accessed on 18/07/2026)

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

References

AI Accountability., Available at <https://www.carnegiecouncil.org/explore-engage/key-terms/ai-accountability>

AI Hallucinations—Understanding the Phenomenon and Its Implications., Available at <https://www.coursera.org/articles/ai-hallucinations>

Artificial Intelligence Liability Directive., Available at <https://www.ai-liability-directive.com/>

Human vs. AI: Who is responsible for AI mistakes? Accountability, law & best practices., Available at <https://www.oneadvanced.com/resources/who-is-responsible-for-ai-mistakes/>

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

Olurunju. N., 'African algorithmic governance: Benefit of a community-based approach' Available at <https://researchictafrica.net/2022/04/03/african-algorithmic-governance-benefit-of-a-community-based-approach/>

Organisation for Economic Co-operation and Development., 'Advancing Accountability in AI: Governing and Managing Risks throughout the Lifecycle for Trustworthy AI' Available at https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf

Organisation for Economic Co-operation and Development., 'Governing with Artificial Intelligence' Available at https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/ai-in-justice-administration-and-access-to-justice_f0cbe651.html

Ozer. M., 'Is Artificial Intelligence Hallucinating?' Available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC11681264/>

Pujari. S et al., 'Artificial intelligence for global health: cautious optimism with safeguards' Available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC10225938/pdf/BLT.23.290215.pdf>

United Nations Development Programme., 'AI for Development: A Positive Tipping Point for People & Planet' Available at <https://www.undp.org/indonesia/blog/ai-development-positive-tipping-point-people-planet>

United Nations Global Compact., 'Artificial Intelligence and the Sustainable Development Goals: Operationalizing technology for a sustainable future' Available at <https://unglobalcompact.org/compactjournal/artificial-intelligence-and-sustainable-development-goals-operationalizing>

United Nations., 'Artificial Intelligence (AI)' Available at <https://www.un.org/en/global-issues/artificial-intelligence>

Who is to Blame for Artificial Intelligence Errors: Building Liability Frameworks for Justice and Equity

United Nations., 'Artificial Intelligence (AI)' Available at <https://www.un.org/en/global-issues/artificial-intelligence>

What are AI Hallucinations?., Available at <https://cloud.google.com/discover/what-are-ai-hallucinations>

When AI goes wrong: 13 examples of AI mistakes and failures., Available at <https://www.evidentlyai.com/blog/ai-failures-examples>