Towards Algorithic Justice: Training Large Language Models (LLMs) on Indigenous African Languages, Ethics and Culture

---

Kariuki Muigua

Table of Contents

Towards Algorithic Justice?: Training Large Language Models (LLMs) on Indigenous African Languages, Ethics and Culture

Kariuki Muigua*

Abstract

This paper discusses how algorithic justice can be realised by training Large Language Models (LLMs) on indigenous African languages, ethics and culture. The paper observes that LLMs are powerful AI tools with the potential to understand, process and generate human language. Due to their immense capacity, the paper notes that LLMs are disrupting various industries through faster, accurate and more informed decision-making. Despite their potential, the paper notes that the adoption of LLMs raises several risks and challenges. In particular, the paper observes that the use of LLMs in Africa and the Global South fuels biases and discrimination undermining algorithic justice. In light of inherent biases and discrimination, the paper posits that there is need to train LLMs on indigenous African languages, ethics and culture towards algorithic justice.

1.0 Introduction

Artificial Intelligence (AI) has developed into a powerful and transformative technology that is revolutionizing virtually all spheres of life. It has been observed that AI is increasingly driving key developments in technology and business[1]. This technology is being widely employed across a broad range of industries enhancing innovation, efficiency and speed of tasks[2]. It has been observed that the availability of large amounts of training data and advances in affordable high computing power are fueling AI's growth with positive impacts on various industries[3]. In particular, it has been observed that the positive impacts of AI are being witnessed across many sectors including education, transportation, finance, data storage, communications, healthcare, environmental governance and law enforcement providing numerous opportunities to unlock Sustainable Development[4].

Harnessing AI therefore provides immense opportunities to utilise technology to foster Sustainable Development for people and planet. It has been observed that AI holds the potential to address complex development challenges from enhancing access to quality education and improving health care, to driving scientific innovation and climate action[5]. Due to its benefits, it has been correctly noted that if appropriately and ethically adopted, AI can drive innovation, promote inclusivity, reduce inequalities and

* PhD in Law (Nrb),SC, FCIArb (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) [June, 2026].

[1] World Intellectual Property Organization., 'Artificial Intelligence and Intellectual Property' Available at https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index (Accessed on 10/06/2026)

[2] Ibid

[3] Ibid

[4] 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.html (Accessed on 10/06/2026)

[5] Organisation for Economic-Cooperation and Development., 'Artificial Intelligence' Available at https://www.oecd.org/en/topics/artificial-intelligence.html (Accessed on 10/06/2026)

help accelerate progress towards the Sustainable Development Goals (SDGs)[6]. Despite its advantages, AI can fuel several risks and challenges if not appropriately and ethically harnessed. For instance, it has been observed that AI can embed biases, cause and widen socio-economic inequalities, violate human rights and undermine environmental sustainability[7].

Algorithmic bias is a fundamental concern when utilising AI across several industries. It has been correctly noted that since AI algorithms are only as unbiased as the data sets they are trained on, algorithmic decision-making may perpetuate existing biases and discrimination[8]. In particular, since most AI tools and systems are designed in the Global North, they may perpetuate bias and prejudices in the Global South unless they are tuned to fit with local circumstances[9]. This situation can be seen through racial, language and gender biases that have been noted in some algorithmic decision-making processes due to historically discriminatory and biased datasets[10].

Algorithmic bias in AI is undesirable since it fuels discrimination, human right violations, inequalities and injustices with disproportionate impact on vulnerable and marginalized populations. Consequently fostering algorithic justice is key towards ensuring the safe, ethical, and appropriate adoption of AI for development[11]. The concept of algorithic justice recognises that while AI tools can drive innovation, governance and development, they can also harm vulnerable and marginalized people, and threaten human rights[12]. Algorithic justice notes that if unchecked and unregulated, AI systems can amplify racism, sexism, ableism and other forms of discrimination[13]. At its core, algorithic justice involves tackling biases by ensuring fairness, inclusivity, transparency, accountability, human rights and non-discrimination in algorithmic decision-making[14]. Fostering algorithic justice is therefore necessary towards safe, ethical and appropriate adoption of AI as a transformative technology towards development.

This paper discusses how algorithic justice can be realised by training Large Language Models (LLMs) on indigenous African languages, ethics and culture. The paper observes that LLMs are powerful AI tools with the potential to understand, process and generate human language. Due to their immense capacity, the paper notes that LLMs are disrupting various industries through faster, accurate and more informed decision-making. Despite their potential, the paper notes that the adoption of LLMs raises several risks and challenges. In particular, the paper observes that the use of LLMs in Africa and the Global South fuels

[6] United Nations., 'Global Dialogue on AI Governance' Available at https://www.un.org/global-dialogue-ai-governance/en (Accessed on 10/06/2026)
[7] United Nations Educational, Scientific and Cultural Organization., 'Ethics of Artificial Intelligence' Available at https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (Accessed on 10/06/2026)
[8] Orero. L.O., & Kaaniru. J., 'Automated Decision-Making Policies in Africa' Available at https://cipit.strathmore.edu/wp-content/uploads/2023/08/Policy-Brief-Design-Automated-Decision-Making.pdf (Accessed on 10/06/2026)
[9] Ibid
[10] 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 10/06/2026)
[11] Algorithmic Justice., Available at https://www.siop.org/wp-content/uploads/legacy/docs/White%20Papers/justice.pdf?ver=2020-05-07-085828-327 (Accessed on 10/06/2026)
[12] Ibid
[13] Ibid
[14] United Nations Educational, Scientific and Cultural Organization., 'Ethics of Artificial Intelligence' Op Cit

biases and discrimination undermining algorithic justice. In light of inherent biases and discrimination, the paper posits that there is need to train LLMs on indigenous African languages, ethics and culture towards algorithic justice.

2.0 Impact of Large Language Models (LLMs) on Algorithic Justice

Large language models (LLMs) have been described as a category of deep learning models which trained on huge amounts of data, thus making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks[15]. LLMs have also been defined as AI systems to designed to learn particular aspects of one or more languages in order to generate coherent and context-relevant language[16]. These aspects include grammar, syntax and semantics[17]. LLMs can also be described as a type of AI tools designed to understand and generate human-like text based on the input they receive[18].

It has been observed that LLMs are built using deep learning techniques allowing them to process vast amounts of data and learn complex patterns in a particular language[19]. By analyzing huge amounts of data, LLMs are able to effectively learn patterns and structures of language, enabling them to predict and generate plausible and coherent text[20]. These models work as key statistical prediction machines that are able to repeatedly and effectively predict the next word in a sequence[21]. It has been observed that LLMs learn patterns in their text and generate language that follows those patterns[22].

Due to their immense capacity, LLMs have been described as a conversational gateway between humans and technology[23]. In particular, it has been observed that LLMs represent a milestone in how humans interact with technology since they are the first AI systems with the capacity to handle unstructured human language at scale, allowing for natural communication between humans and machines[24]. Through their capabilities, LLMs can be adapted to a variety of applications, such as text generation, summarising, translating and answering questions among others[25]. In particular, LLMs can be utilised to perform a wide

[15] What are Large Language Models (LLMs)?., Available at https://www.ibm.com/think/topics/large-language-models (Accessed on 10/06/2026)
[16] European Data Protection Supervisor., 'Large Language Models (LLM)'., Available at https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en (Accessed on 10/06/2026)
[17] Ibid
[18] AI Demystified: Introduction to Large Language Models., Available at https://uit.stanford.edu/service/techtraining/ai-demystified/llm (Accessed on 10/06/2026)
[19] Ibid
[20] What is a Large Language Model (LLM)?., Available at https://ask.library.arizona.edu/faq/407985 (Accessed on 10/06/2026)
[21] What are Large Language Models (LLMs)?., Op Cit
[22] Ibid
[23] World Economic Forum., 'Why Large Language Models are the future of manufacturing' Available at https://www.weforum.org/stories/2024/04/why-large-language-models-are-so-important-for-the-future-of-the-manufacturing-industry/?gad_source=1&gad_campaignid=22228224717&gbraid=0AAAAAoVy5F7O8qSC37xayF84Ju8OSQP4Q&gclid=Cj0KCQjwlqTRBhCBARIsANrkrxhnd33Tg3nmp2OXIPQhJsjs4CZDbBSehbgRoFU92QbI0FozisNPCwAaAhlKEALw_wcB (Accessed on 10/06/2026)
[24] What are Large Language Models (LLMs)?., Op Cit
[25] European Data Protection Supervisor., 'Large Language Models (LLM)' Op Cit

range of language-related tasks including language translation, creative writing, text summarization and answering of questions[26]. Further, since they are able to generate coherent and context-specific text, LLMs can be useful in automated reporting, chatbots and content creation[27].

LLMs are therefore a revolutionary technology. They represent a shift from traditional AI models which focus on perception and understanding due to their ability to generate human-like content[28]. These models are being employed in a wide range of industries such as AI assistants in customer service, transcription, healthcare, procurement management and manufacturing[29]. It has been pointed out that when given agentic capabilities, LLMs can perform, with varying degrees of autonomy, various tasks that would otherwise be performed by humans[30].

Despite their effectiveness in natural language processing and generation, LLMs suffer from several limitations and risks. For instance, LLMs can be misused to generate and spread misleading and harmful content including deepfakes and disinformation[31]. Overreliance on LLMs can also raise legal and ethical concerns including violation of intellectual property rights and data privacy especially when they are used to generate content based on copyrighted materials or personal data[32]. In particular, it has been observed that LLMs can inherit and worsen biases present in the data they are trained on[33]. For example, if the training data includes biased or prejudiced language, the LLM model may produce outputs that reflect these biases, leading to harmful, discriminatory and prejudicial outcomes[34]. In particular, it has been observed that LLMs models in Africa are mostly trained in English undermining indigenous languages in the continent[35]. This often leads to discriminatory outcomes in key sectors including education, healthcare and agriculture since these models neglect unique data sets that reflect cultural, ethical and linguistic needs in the continent[36]. It is therefore imperative to train LLMs on indigenous African languages, ethics and culture in order to harness the transformative power of AI for development.

3.0 Training Large Language Models (LLMs) on Indigenous African Languages, Ethics and Culture

LLMs provide immense opportunities towards harnessing the revolutionary power of AI for development. Their ability to understand and generate natural language has led to LLMs being widely applied in different fields and domains including translation, customer care (chatbots), natural language processing, education

[26] AI Demystified: Introduction to Large Language Models., Op Cit
[27] Ibid
[28] Large Language Models Explained., Available at https://www.nvidia.com/en-us/glossary/large-language-models/ (Accessed on 10/06/2026)
[29] What are Large Language Models? LLMs Explained., Available at https://cohere.com/blog/large-language-models (Accessed on 10/06/2026)
[30] What are Large Language Models (LLMs)?., Op Cit
[31] AI Demystified: Introduction to Large Language Models., Op Cit
[32] Ibid
[33] Ibid
[34] Ibid
[35] African-language Large Language Models (LLMs) present major opportunities for the continent., Available at https://medium.com/code-for-africa/african-language-large-language-models-llms-present-major-opportunities-for-the-continent-8a92a69518b3 (Accessed on 10/06/2026)
[36] Ibid

(language training), creation of artistic works, and healthcare[37]. However, it has been observed that as the AI revolution transforms the world, millions of people in Africa and the Global South cannot tap its full promise since the languages they speak are not built into LLMs[38]. This perpetuates biases, discrimination and prejudices[39].

It is therefore necessary to train LLMs on indigenous African languages, ethics and culture in order to foster algorithic justice. It has been observed that failure to integrate indigenous African languages into LLMs prevents millions of people in the continent from interacting and harnessing the power of the digital world[40]. In light of this concern, it has been correctly observed that in Africa and the Global South, AI holds immense potential to drive development, especially if models including LLMs are trained to understand and function in indigenous and local languages[41]. Indigenous languages in Africa carry cultural knowledge, context and ethics that can guide the development of AI models that fit with local needs and circumstances[42]. Consequently, it has been pointed out that developing LLMs that understand and generate text and audio in indigenous African languages can help preserve the linguistic heritage, ethics and culture of the people of Africa[43].

LLMs trained on African languages, ethics and culture can increase access to information for millions of people in the continent. This can transform key sectors including *education* by enabling students to learn in indigenous languages[44]; *healthcare* through better communication between patients and medical professionals[45]; *environmental governance* through the use of indigenous and traditional ecological knowledge towards sound environmental protection[46]; *agriculture* by providing farmers with access to information on weather, best farming practices, and market prices in their own languages[47]; and access to government services for all linguistic groups[48]. Training LLMs models on indigenous African languages can also support cultural preservation by ensuring that the languages, culture and ethics of the people of Africa are harnessed and documented in the digital world for posterity[49].

---

[37] European Data Protection Supervisor., 'Large Language Models (LLM)' Op Cit
[38] Millions on the African continent can't fully benefit from the AI revolution. This Princeton course aims to change that., Available at https://research.princeton.edu/news/millions-african-continent-can%E2%80%99t-fully-benefit-ai-revolution-princeton-course-aims-change (Accessed on 11/06/2026)
[39] Ibid
[40] Ibid
[41] Code for Africa., 'African-language Large Language Models (LLMs) present major opportunities for the continent' Available at https://medium.com/code-for-africa/african-language-large-language-models-llms-present-major-opportunities-for-the-continent-8a92a69518b3 (Accessed on 11/06/2026)
[42] Ibid
[43] Ibid
[44] Ibid
[45] Ibid
[46] United Nations Development Programme., 'People-Centric AI for Conserving Biodiversity' Available at https://www.undp.org/sites/g/files/zskgke326/files/2025-12/people-centric-ai-for-conserving-biodiversity.pdf (Accessed on 11/06/2026)
[47] Code for Africa., 'African-language Large Language Models (LLMs) present major opportunities for the continent' Op Cit
[48] Ibid
[49] Ibid

Training LLMs on indigenous African languages, ethics and culture is therefore crucial towards fostering algorithic justice by enhancing access to information for millions of people in the continent. Consequently, it is imperative for developers to collaborate with local communities and indigenous peoples in order develop AI technologies including LLMs that are not only scientifically robust but also socially and contextually appropriate[50]. It is also necessary to bolster Africa's institutional and human capacity in AI through training, investments and research in order to ensure the availability of competent and qualified individuals and institutions that can guide the development of AI models including LLMs trained on indigenous African languages, ethics and culture[51].

4.0 Conclusion

With AI systems including LLMs transforming the world, fostering algorithic justice is key towards tackling biases, prejudices and discrimination in order to ensure that all people, communities and nations harness the power of technology for development. In particular, with AI being widely adopted in Africa, it is imperative to train LLMs on indigenous African languages, ethics and culture towards algorithic justice.

[50] United Nations Development Programme., 'People-Centric AI for Conserving Biodiversity' Op Cit

[51] Code for Africa., 'African-language Large Language Models (LLMs) present major opportunities for the continent' Op Cit

References

African-language Large Language Models (LLMs) present major opportunities for the continent., Available at https://medium.com/code-for-africa/african-language-large-language-models-llms-present-major-opportunities-for-the-continent-8a92a69518b3

AI Demystified: Introduction to Large Language Models., Available at https://uit.stanford.edu/service/techtraining/ai-demystified/llm

Algorithmic Justice., Available at https://www.siop.org/wp-content/uploads/legacy/docs/White%20Papers/justice.pdf?ver=2020-05-07-085828-327

Code for Africa., 'African-language Large Language Models (LLMs) present major opportunities for the continent' Available at https://medium.com/code-for-africa/african-language-large-language-models-llms-present-major-opportunities-for-the-continent-8a92a69518b3

European Data Protection Supervisor., 'Large Language Models (LLM)'., Available at https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en

Large Language Models Explained., Available at https://www.nvidia.com/en-us/glossary/large-language-models/

Millions on the African continent can't fully benefit from the AI revolution. This Princeton course aims to change that., Available at https://research.princeton.edu/news/millions-african-continent-can%E2%80%99t-fully-benefit-ai-revolution-princeton-course-aims-change

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/

Orero. L.O., & Kaaniru. J., 'Automated Decision-Making Policies in Africa' Available at https://cipit.strathmore.edu/wp-content/uploads/2023/08/Policy-Brief-Design-Automated-Decision-Making.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.html

Organisation for Economic-Cooperation and Development., 'Artificial Intelligence' Available at https://www.oecd.org/en/topics/artificial-intelligence.html

United Nations Development Programme., 'People-Centric AI for Conserving Biodiversity' Available at https://www.undp.org/sites/g/files/zskgke326/files/2025-12/people-centric-ai-for-conserving-biodiversity.pdf

United Nations Educational, Scientific and Cultural Organization., 'Ethics of Artificial Intelligence' Available at https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

United Nations., 'Global Dialogue on AI Governance' Available at https://www.un.org/global-dialogue-ai-governance/en

What are Large Language Models (LLMs)?., Available at https://www.ibm.com/think/topics/large-language-models

What are Large Language Models? LLMs Explained., Available at https://cohere.com/blog/large-language-models

What is a Large Language Model (LLM)?., Available at https://ask.library.arizona.edu/faq/407985

World Economic Forum., 'Why Large Language Models are the future of manufacturing' Available at https://www.weforum.org/stories/2024/04/why-large-language-models-are-so-important-for-the-future-of-the-manufacturing-industry/?gad_source=1&gad_campaignid=22228224717&gbraid=0AAAAAoVy5F7O8qSC37xayF84Ju8OSQP4Q&gclid=Cj0KCQjwlqTRBhCBARIsANrkrxhnd33Tg3nmp2OXIPQhJsjs4CZDbBSehbgRoFU92QbI0FozisNPCwAaAhlKEALw_wcB

World Intellectual Property Organization., 'Artificial Intelligence and Intellectual Property' Available at https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index