

Global Journal of Arts Humanity and Social Sciences
ISSN: 2583-2034
Abbreviated key title: Glob.J.Arts.Humanit.Soc.Sci
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link: <https://gsarpublishers.com/journal-gjahss-home/>

Volume - 5 | Issue - 01 | Jan 2025 | Total pages 30-34 | DOI: 10.5281/zenodo.14672016

Empowering African Languages Through Machine Translation and Artificial Intelligence

By

Ozioma Anyawuiké¹, Iroetugo Edith Ruth²

^{1,2}French language and translation Department of Modern European Languages, Nnamdi Azikiwe University Awka



Article History

Received: 01- 01- 2025
Accepted: 14- 01- 2025
Published: 16- 01- 2025

Corresponding author
Ozioma Anyawuiké

Abstract

Thanks to their artificial potential - deep learning AI and machine translation - it appears that African indigenous languages, several of which fit the description of low-resource and endangered, are set to profit. Artificial Intelligence (AI) technologies, including voice recognition software, natural language processing (NLP), and neural machine translation (NMT), give inventive strategies for documenting, recording and reviving endangered languages in a continent that boasts of over 2,000 spoken languages. Projects such as Mozilla's Common Voice and the Masakhane project demonstrate the role that collaborative efforts in communities can play in developing AI models that are linguistically appropriate for African languages. This paper explores how AI powered solutions can facilitate digital inclusion, encourage multilingual education, and eradicate language barriers by allowing people to use technology in their own preferential language. But there are still a lot of issues, like standardized orthographies, a lack of digital sources, and regional differences in AI development. In addressing these challenges, it is paramount to adopt ethical AI frameworks that appreciate Africa's diversity, local participation and sustained funding.

Keywords: Artificial intelligence (AI), Indigenous languages, Neural machine translation (NMT), Natural language processing (NLP), Digital inclusion

Introduction

The development of machine translation and artificial intelligence (AI) offers an outstanding opportunity to manage multilingual issues and conserve and advance indigenous African languages. More than 2,000 languages are spoken on the African continent, but many of them are categorized as "low-resource languages," which means they don't have enough digital data or standardized orthographies to enable strong technological integration. As urbanization and globalization promote the dominance of widely spoken languages like Arabic, English, and French, these languages are in danger of being extinct. Artificial intelligence (AI) tools like voice recognition software, natural language processing (NLP), and neural machine translation (NMT) are in an ideal position to narrow this gap and effectively address Africa's linguistic diversity. For instance, NLP techniques can help create educational materials and digital archives for endangered languages, while NMT can facilitate real-time translation between African languages, promoting communication across linguistic

barriers. Moreover, speech recognition software designed for tonal and morphologically complex languages—such as Yoruba, Zulu, and Amharic—is opening the door for increased technological accessibility by enabling millions of people to communicate with digital platforms in their mother tongues.

The Role of AI in Preserving Indigenous African Languages

The ability of AI to record and preserve endangered languages is among its most important accomplishments. According to UNESCO, almost 40% of the world's languages are in danger of going extinct, with many African languages being especially endangered as a result of globalization, urbanization, and the prevalence of colonial languages like Portuguese, French, and English. Many indigenous languages are at risk of extinction due to a lack of written documents and digital resources, which adds to these drawbacks. By digitizing and standardizing these languages for future generations, artificial intelligence (AI) techniques provide creative answers to this urgent problem. Neural machine



translation (NMT) systems, for instance, are able to interpret and translate written or spoken language into digital representations, generating collections that function as cultural archives. These initiatives are further strengthened by voice recognition software, which allows speakers to record oral traditions, idioms, and narratives—all of which are essential to many African cultures. In order to keep less popular languages alive and usable in the digital era, AI can also help with the creation of dictionaries, the development of new writing systems, and the integration of these languages into international communication networks. In addition to protecting linguistic history, these technologies promote inclusivity by enabling communities to maintain their identities and engage with the global digital environment.

AI is being actively used by projects like the *Masakhane* project, a pan-African research community, to develop machine translation systems for underrepresented languages including Swahili, Amharic, and Yoruba. The goal of these initiatives is to ensure the survival and significance of African languages in a globalized society by making them available online. The researchers stress the importance of developing AI at the grassroots level as a way to empower local populations and preserve cultural identity. *Masakhane*, for instance, uses collaborative strategies in which African linguists, engineers, and AI specialists collaborate to solve the distinctive linguistic patterns of African languages, like agglutination and tone systems, which are frequently disregarded in standard AI models. Founder of the African Science Academy, Tom Ilube emphasizes the extensive cultural impacts associated with projects like this: “Algorithms define the future, and people forget that algorithms are not just technical; they are political and cultural.” This viewpoint emphasizes just how important it is to create AI systems that respect and take into account local situations rather than enforcing external norms. Projects like *Masakhane*, which support African-driven AI solutions, ensure that technology advancements are in line with the diverse legacy of the continent by strengthening African countries' cultural foundations and promoting technological inclusivity.

African oral traditions can be preserved with the help of voice recognition software, especially for languages without written forms. Orality is still vital in passing down African history, culture, and identity, as Bukola Adeyemi Oyeniyi notes, with customs such as dirges, panegyrics, and bridal poetry playing a significant role in social and ceremonial life (Oyeniyi, 2019). These essential spoken words can be digitally preserved by AI-driven speech transcription and translation technologies, guaranteeing their accessibility and protecting them from the deterioration of cultural heritage so that future generations can enjoy and learn from them.

Promoting Multilingual Communication in Africa

With over 2,000 languages spoken in Africa, artificial intelligence (AI) technologies—in particular, machine translation (MT) and speech recognition systems—are essential for tackling the continent's wide range of languages. For groups that have historically lacked written representations of their languages, these tools aid in bridging communication gaps, maintaining oral

traditions, and promoting digital inclusion. Africans may now communicate with digital platforms in their native tongues thanks to artificial intelligence (AI) capabilities like Facebook's multilingual Natural Language Processing (NLP) algorithms and Google's Neural Machine Translation. These developments support accessibility and inclusivity, both of which are essential for building a more cohesive and just society throughout the continent. According to Silas Adekunle, CEO of Awarri, the first multilingual major language model in Nigeria, "Data is central to the development of all AI models and applications" (THISDAYLIVE, 2024). With support for languages including Hausa, Igbo, and Yoruba, Awarri's platform encourages user contributions that document translations, to improve AI's comprehension of regional languages and customs. Through its vast data gathering platform, this effort highlights the critical role artificial intelligence (AI) plays in conserving Nigeria's cultural legacy and illustrates how technology may close the gap between indigenous languages and contemporary digital platforms.

Given that over 2,000 languages, many of which are low-resource languages, coexist in Africa, these AI systems also address the continent's linguistic issues. By promoting interlanguage communication along with developing digital tools that support both economic activity and cultural expression and communication, artificial intelligence (AI) can empower speakers of different languages. It (AI) plays an indispensable part in supporting the preservation and advancement of Africa's distinctive linguistic landscape; as it encourages digital inclusion and ensures that indigenous languages are represented in the digital ecosystem. While facilitating multilingual techniques, AI tools have the revolutionary potential to completely change education, especially when it comes to African languages. Given that students are more likely to succeed academically when taught in their mother tongue, UNESCO supports the use of mother tongues in the classroom. The preservation of both linguistic and cultural identities is aided by the development of greater cognitive abilities and the maintenance of stronger cultural bonds among children who get their education in their mother tongue, according to UNESCO publications (UNESCO, 2024).

The development of educational materials in indigenous African languages can be facilitated by utilizing AI. Localized educational materials, such as e-learning platforms and textbooks, that are tailored to students' native tongues can be produced with the use of AI-driven technologies. In order to ensure that African languages are not left behind in the digital era, this would be a major step in closing the digital divide. According to UNESCO, which emphasizes multilingual education as a crucial component of inclusive, high-quality learning, encouraging instruction in native languages also aids in the creation of inclusive learning settings and gives communities the tools they need to protect their cultural heritage.

AI and machine translation play an essential role in tackling the language challenges that African entrepreneurs face when entering global markets. Because many Sub-Saharan Africans deal with significant challenges from international trade due to language



barriers, the capacity to communicate in various languages is critical to supporting economic growth. According to *Digital for Good*, “63% of Sub-Saharan Africans face challenges accessing global markets due to language barriers” (Global Voices, 2023). AI tools, such as machine translation and natural language processing (NLP), are already breaking down these barriers, as they enable more Africans to communicate with global partners, interact with customers, and engage in online commerce.

Through these technologies, African businesses can employ AI translation techniques to reach global markets, engage in e-commerce, and gain access to international investment. This promotes greater inclusivity, economic progress, and worldwide opportunity. Furthermore, AI's ability to facilitate multilingual communication can help close the digital divide, which has previously prevented many Africans from engaging in the global economy (Global Voices, 2023; SciDev.Net, 2023). In this context, AI-powered platforms such as Google Translate and other multilingual NLP technologies assist African firms in translating product descriptions, contracts, and marketing materials into other languages, hence simplifying cross-border trade. Furthermore, AI enables faster and more consistent customer service, making it easier for African entrepreneurs to increase their worldwide reach (SciDev.Net, 2023). This innovation not only expands economic prospects, but it also increases the representation of African languages on digital platforms, promoting inclusivity in worldwide communication.

Challenges

While the potential of artificial intelligence (AI) to preserve and promote African languages is immense, several significant challenges remain. A primary issue is the fact that many African languages are classified as “low-resource” languages. These languages often lack extensive digital corpora, standardized orthographies, and sufficient linguistic data, which are quite important for training AI models. As a result, AI systems struggle to understand and process these languages accurately. Without substantial digital content in these languages, AI tools such as natural language processing (NLP) and machine translation (MT) systems cannot be effectively developed or deployed. Also, many African languages lack sufficient digital data to train AI models effectively. For instance, neural machine translation relies on large datasets to generate accurate and context-aware translations. Without significant datasets, AI models may produce incomplete or inaccurate translations, compromising their effectiveness.

Moreover, much of the development of AI occurs outside of Africa, particularly in places where major world languages such as English, French, and Chinese are dominant. This regional mismatch in AI creation can result in solutions that do not adequately reflect the linguistic, cultural, or sociopolitical reality of African countries. For example, AI systems built in non-African contexts may be unsuitable for dealing with the tonal nuances and syntactical details found in many African languages, such as Yoruba and Zulu. Likewise, these AI technologies may not address local language variants or dialects, excluding a sizable proportion

of the population who speak non-standard forms of their native languages (Adewale, 2023; Adedeji & Oyekunle, 2020).

Addressing the Challenges of Low-Resource Languages

Despite Africa's vast linguistic diversity, many of its languages are underrepresented in digital spaces due to a lack of written records, standardized orthographies, and digitized linguistic resources. Various limitations impede the development of AI models capable of understanding and producing various languages, which is critical for creating inclusive technology that local populations can utilize in their native languages. To solve this, a number of initiatives have emerged centered on the creation of open-source datasets and AI models specific to African languages. One of the important initiatives is Mozilla's *Common voice* project, which intends to develop big, open-source speech datasets for a variety of languages. This effort encourages native speakers of African languages to donate their voices to the training of speech recognition models. Similarly, the *African Language Dataset Challenge* aims to generate comprehensive datasets that authentically reflect African languages. Both projects rely on crowdsourcing to collect data, ensuring that the diversity of African languages is represented in these digital archives. These datasets are not only useful for building AI applications like speech-to-text systems and machine translation tools, but they also ensure that African languages are effectively represented in the digital domain.

Moustapha Cisse, the head of Google's AI lab in Accra, emphasizes the importance of locally-driven initiatives in shaping the future of AI, particularly in Africa. He argues that “Africa's strength lies in its human resources, and initiatives that harness this potential will drive AI innovation in a culturally sensitive way” (Cisse, 2024). This sentiment emphasizes the importance of involving locals in the creation of culturally relevant AI technologies that are sensitive to the continent's particular linguistic issues. By leveraging on the knowledge and contributions of native speakers, these programs not only empower local people but also ensure that the AI systems created are more suited to the linguistic details and ambiguities of African languages. One such locally-driven initiative is *Masakhane*, a collaborative project focused on building natural language processing (NLP) models for African languages. *Masakhane* brings together researchers, linguists, and AI practitioners across Africa to collaborate on the creation of machine translation and speech recognition systems for underrepresented languages. This initiative is essential because it ensures that African voices, literally and figuratively, are integral to the AI landscape. *Masakhane* helps to ensure the long-term development of AI tools that represent the needs and ambitions of African communities by encouraging collaboration and knowledge-sharing across the continent. In summary, initiatives like Mozilla's *Common Voice*, the *African Language Dataset Challenge*, and *Masakhane* are critical in addressing the linguistic challenges of low-resource African languages. These initiatives, which involve local communities as

well as making use of the human resources on the continent, are empowering African populations and preserving their language legacy for future generations by ensuring that AI technologies are developed in a way that is culturally relevant.

Ethical and Social Considerations

While the potential of AI to preserve and promote African languages is undeniable, it is important to recognize the ethical considerations involved in its implementation. As AI technologies, such as machine translation and natural language processing, become more integral to communication, there are risks associated with over-relying on these systems. One of the primary concerns is that excessive dependence on AI-powered translation may discourage people, particularly younger generations, from learning and engaging with their indigenous languages. This could lead to the erosion of linguistic diversity over time, as individuals may favour languages that are more prominently represented in digital spaces, such as English, French, or other widely spoken languages.

Again, AI models are only as good as the data they are trained on, and this can sometimes lead to biases if the datasets are not sufficiently diverse or representative of the cultural and linguistic variety across Africa. For instance, machine learning models trained on data that predominantly comes from urban settings, or from only a handful of dominant languages, could perpetuate stereotypes or even misrepresent marginalized languages and cultures. As a result, these biased AI systems could reinforce existing inequalities, misrepresent local customs, and fail to account for the nuances that exist within African societies.

Tom Ilube's observation on the ethical development of AI is particularly relevant in this context: "AI must not only solve technical problems but also respect the cultural and political contexts in which it operates" (Ilube, 2024). Ilube's emphasis on cultural awareness emphasizes the need for AI systems to be developed with deep understanding and respect for the social and political dynamics of the regions they aim to serve. This means that African nations and local communities must play an active role in the development of AI technologies, ensuring that the tools reflect local needs, values, and cultural contexts.

The way forward lies in building collaboration between local researchers, governments, and global tech companies. By working together, these groups can create AI systems that are not only technologically sound but also culturally relevant and equitable. Local researchers can ensure that indigenous languages are properly represented in AI datasets, and governments can help enforce policies that prioritize inclusivity and fairness. Global tech companies, on their part, must ensure that their AI tools are adaptable to the unique challenges faced by African societies, such as limited internet access and the need for multilingual support.

This collaborative approach would help create AI systems that do more than simply replicate existing global models. Instead, they would serve as tools that empower African communities to preserve their languages, promote cultural heritage, and promote greater social inclusion. Through such careful, culturally-aware AI

development, the continent can take advantage of the opportunities presented by artificial intelligence while mitigating its potential harms.

Future Directions and Opportunities

The future of AI in Africa lies in building sustainable and inclusive development strategies that prioritize the continent's unique needs. Strengthening collaborations between governments, academic institutions, and private sectors is key to achieving these goals. These partnerships can drive investments in critical AI infrastructure, capacity building, and education. Google's AI lab in Ghana exemplifies the potential of public-private collaboration in advancing AI technologies tailored to African languages and contexts. Similarly, Mozilla's Common Voice project, which gathers voice data in underserved languages like Kiswahili and Igbo, highlights how global initiatives can support localized AI solutions.

Building awareness and increasing participation in AI research among Africans is another vital step. By engaging more researchers, developers, and communities in open-source AI projects, Africa can position itself as a leader in creating culturally relevant and impactful AI solutions. Countries like Nigeria and Kenya are already contributing to this vision. Nigeria's multilingual Large Language Model (LLM) initiative integrates local languages into AI systems, demonstrating how tailored solutions can address the continent's linguistic diversity.

Moreover, there are significant opportunities to harness AI for socio-economic development. From improving healthcare delivery through AI diagnostics to enhancing education via localized language learning tools, the potential applications of AI in Africa are vast. Governments can further promote these developments by creating policies that encourage innovation while safeguarding cultural integrity. With these strategies in place, Africa is poised not only to adopt AI technologies but also to shape their evolution in ways that align with its diverse cultural and social realities.

Conclusion

Artificial intelligence (AI) offers a transformative opportunity to preserve and promote African indigenous languages while addressing the continent's linguistic diversity. Tools like neural machine translation (NMT), natural language processing (NLP), and voice recognition enable communities to digitize and preserve their linguistic heritage. Initiatives such as *Masakhane* and Mozilla's *Common Voice* highlight the power of locally-driven, collaborative efforts in advancing AI solutions tailored to African languages. Despite challenges like limited digital resources and geographic disparities in AI development, partnerships among governments, academia, and private sectors can drive sustainable progress. Through the prioritisation of inclusive, culturally aware AI initiatives, Africa can harness technology to preserve its languages, foster socio-economic growth, and enhance digital inclusion.

References

1. Google Opens AI Research Center in Ghana." *Google*



- Africa Blog*, 2024. (<https://africa.googleblog.com>).
2. Mozilla Common Voice: Kiswahili and African Languages." *Mozilla Foundation*, accessed December 3, 2024. (<https://voice.mozilla.org/en>).
 3. Nigeria Launches First Multilingual AI Language Model." *Inclusion Times*, April 22, 2024. (<https://www.inclusiontimes.com/nigeria-launches-first-multilingual-ai-language/>).
 4. Masakhane: Using AI to Bring African Languages into the Global Conversation." *Digital for Good | RESET.ORG*. Accessed November 8, 2024. (<https://en.reset.org>).
 5. Adedeji, Olamide, and Abiola Oyekunle. "AI and Language Preservation in Africa: Bridging the Technological Divide." *African Technology Review*, no. 15 (2020): 45-58.
 6. Adekunle, Silas. "Data Is Central to the Development of All AI Models and Applications." *THISDAYLIVE*, April 22, 2024. (<https://www.thisdaylive.com>).
 7. Adewale, Olumide. "Challenges in Developing AI for African Languages: A Nigerian Perspective." *Journal of African Linguistics*, no. 8 (2023): 12-25.
 8. Cisse, Moustapha. "The Role of AI in Empowering Africa's Languages." *Nature News*. Accessed November 8, 2024. (<https://www.nature.com>).
 9. Global Voices. "How Africans Are Bridging the Language Digital Divide." May 2024. <https://globalvoices.org>.
 10. Hanson, Vicki L. "Advances in AI-enabled language translation hold special promise for the developing world." ITU. November 2024.
 11. Ilube, Tom. *Archives of IT: Oral History Collection*. November 8, 2024. <https://archivesit.org.uk/interviews/tom-ilube>.
 12. Masakhane. "Masakhane: Machine Translation for Africa by Africans." November 8, 2024. (<https://www.masakhane.io>).
 13. Oyeniyi, Bukola Adeyemi. "Orality, History, and Historical Reconstruction." In *New Perspectives on African Oral Traditions and Folklore*, edited by Clement A. Akangbe, 105-130. Ibadan: Oyo University Press, 2019.
 14. SciDev.Net. "Digital Tool Aims to Translate 2,000 African Languages." June 2024. <https://scidev.net>.
 15. UNESCO. "Multilingual Education: Empowering Minority Languages." *UNESCO Reports*. Accessed November 19, 2024. (<https://unesdoc.unesco.org>).