Uirtus (Peer-reviewed Journal of Arts and Humanities)
Available online at https://uirtus.net/
E-mail: soumissions@uirtus.net / revueuirtus@gmail.com
Vol. 5, No. 2, August 2025, Pages: 38-57
DOI: https://doi.org/10.59384/uirtus.2025.2947
ISSN: 2710 - 4699

The Role of Machine Translation in African Languages: Potentials and Pitfalls

Ulrich Douo

Article history:

Submitted: June 5, 2025
Revised: July 2, 2025
Accepted: July 9, 2025

Keywords:
Machine translation, African languages, Natural language processing, language technology, low-resource languages

Abstract

This article examines the evolving role of machine translation (MT) in the context of African languages, highlighting its potential for linguistic equity and digital inclusion. With over 2,000 languages spoken across the continent, MT technologies present a unique opportunity to bridge linguistic divides, enhance access to information and services, and promote cultural and educational inclusion across diverse linguistic communities. However, deploying MT in African contexts is fraught with challenges. These include the scarcity of high-quality linguistic data, the structural and typological complexity of many African languages, and the lack of culturally adaptive translation models. The majority of extant MT systems are optimized for high-resource languages, frequently resulting in low accuracy and cultural misrepresentation when applied to African languages. The article posits a multidisciplinary, locally grounded approach to MT development, prioritizing the creation of inclusive datasets, investing in low-resource language technologies, and integrating cultural and contextual awareness into translation models. It is imperative that these issues are addressed in order to ensure that MT can serve as a tool for linguistic equity and digital inclusion across the continent.

Corresponding author:
Ulrich Douo,
University of Ghana
Email : udouo@ug.edu.gh
ORCID : 0009-0001-5323-3217

UIRTUS
— UIRTUS —
vol. 5, no. 2, August 2025 ISSN 2710-4699 Online

Introduction

Language occupies a pivotal position in human interaction, influencing access to knowledge, identity, and power. In multilingual societies such as those across the African continent, where over 2,000 languages are spoken, language can either facilitate inclusive development or exacerbate marginalization. In the contemporary era, characterised by the proliferation of digital technologies, the role of machine translation (MT) has become increasingly pronounced in mediating communication and facilitating access to information. MT systems, which automatically translate text or speech from one language into another, are now embedded in many everyday technologies, including search engines, social media, and mobile applications. Nevertheless, the development and deployment of MT for African languages remains limited and uneven, giving rise to critical questions regarding both its potential and its limitations. (Abbasi; Marivate, and Sefara).

The advent of artificial intelligence (AI) and natural language processing (NLP) has precipitated a paradigm shift in the capabilities of machine translation (MT), particularly in the context of high-resource languages such as English, Mandarin, French, and Spanish. The enhancement of these languages is facilitated by the availability of substantial digitised corpora, robust linguistic resources, and considerable investment in computational infrastructure.

Conversely, a considerable number of African languages are regarded as under-resourced or low-resource languages, characterised by an absence of standardised orthographies, linguistic instruments, and digital data (Joshi, et al.). This technological disparity signifies that while global MT systems continue to advance, numerous African languages remain excluded from digital and AI-driven ecosystems.

Notwithstanding the challenges encountered, there is an increasing recognition of MT's potential to promote linguistic inclusion across Africa. Using MT to translate digital content, including but not limited to health information, educational materials, news, and public service announcements, into local languages has the potential to enhance access to critical information and services (Hershcovich). In multilingual states where government and institutional communication frequently occurs in a limited number of dominant or colonial languages, MT offers a scalable solution for addressing linguistic inequities. Furthermore, emerging techniques in low-resource MT, including transfer learning, multilingual neural models and community-led data initiatives, offer promising pathways to overcome existing limitations (Nekoto, et al.).

However, it must be noted that these potential benefits are accompanied by significant drawbacks. Many MT systems encounter challenges when processing the complex morphology, rich tonal systems and context-dependent meanings that are characteristic of African languages. The quality of translations is often inadequate, leading to semantic errors, grammatical inconsistencies, and cultural misinterpretations. Moreover, the training data for these systems frequently exhibits biases or inconsistencies, resulting in translations that have the potential to perpetuate harmful stereotypes or distort critical meanings, particularly within sensitive contexts such as healthcare and the legal domain (Bender, et al.). Ethical concerns also arise regarding the prioritisation of languages, the development and control of technology, and the meaningful inclusion of local communities in these processes.

A socio-political dimension must also be considered in this context. The promotion of specific African languages through MT may inadvertently result in the marginalisation of those with fewer resources, thereby exacerbating existing linguistic hierarchies. Furthermore, the increasing automation of translation work has the potential to displace human translators and interpreters, particularly those working in community and educational contexts. In the absence of a participatory and context-aware approach, there is a risk that MT will become a tool of linguistic dominance rather than empowerment.

The present study investigates the complex interplay between the potential benefits and disadvantages of MT in the African context. Drawing on interdisciplinary perspectives from the fields of computational linguistics, sociolinguistics and development studies, the study explores the current state of MT initiatives in Africa, highlighting successful case studies and analysing ongoing challenges. A particular emphasis is placed on issues of data scarcity, ethical development, and community involvement. The objective of this study is to contribute to a more inclusive, equitable, and sustainable approach to machine translation, with the aim of supporting Africa's vast linguistic diversity in the digital age.

1. Literature Review and Theoretical Framework

The field of machine translation (MT) has witnessed rapid advancements, particularly with the advent of neural machine translation (NMT) systems. However, much of the focus in research and development has historically centered on widely spoken languages, often neglecting African languages, which are linguistically rich but underrepresented in computational resources (Martinus). The growing digital divide has motivated scholars and technologists to examine how MT can be harnessed to both empower and potentially misrepresent African linguistic communities.

Studies have highlighted that African languages, such as Swahili, Yoruba, and Amharic, possess complex morphological structures, tonal features, and diverse dialects that pose significant challenges to conventional MT models (Yimam). These languages are also low-resource, meaning they lack the large parallel corpora typically needed to train high-performing MT systems (Nekoto, et al.). Efforts such as the Masakhane project have emerged to bridge this gap by promoting community-driven research and dataset development for African languages. This decentralized approach not only improves translation quality but also ensures that the cultural and linguistic nuances of African languages are preserved.

Despite promising progress, several pitfalls remain. One major issue is the propagation of linguistic bias and inaccuracies, especially when MT tools are trained on insufficient or poor-quality data (Sennrich). Inaccurate translations can lead to misunderstandings, misinformation, or even the erosion of indigenous knowledge systems. Moreover, the dominance of English in training data can lead to Anglocentric translations that distort original meanings (Nekoto, et al.). Researchers warn against the overreliance on MT for critical domains such as health, legal communication, and education, where accuracy is paramount.

In addition, questions of language standardization arise. Many African languages exist more in spoken than written form, with multiple dialects coexisting. MT systems may inadvertently prioritize dominant dialects, marginalizing others and reinforcing linguistic hierarchies (Bird). These challenges necessitate interdisciplinary collaboration among linguists, technologists, and local communities to develop ethical and effective MT tools.

This study is anchored in two theoretical frameworks: **Linguistic**

Relativity and **Technological Appropriation Theory**.

Linguistic Relativity, derived from the Sapir-Whorf hypothesis, posits that language shapes thought and influences how individuals perceive and interact with the world (Whorf). In the context of MT, this theory underscores the importance of preserving linguistic structures and cultural contexts within African languages. If MT systems distort these elements, they not only alter communication but also compromise the cognitive and cultural frameworks of speakers. Thus, the accurate representation of African languages in MT systems is not merely a technical concern, but a cognitive and sociocultural one.

Technological Appropriation Theory, on the other hand, examines how users adapt and repurpose technology within their own cultural and contextual frameworks (Bar, et al.). African communities actively reshape MT tools to fit local needs. This theory helps to explain how grassroots initiatives like Masakhane gain traction by leveraging open-source tools and community collaboration, fostering innovations that large tech firms often overlook. Together, these frameworks emphasize the complex interplay between language, culture, and technology. They offer a lens through which the potentials (e.g., linguistic inclusion, improved access to information) and pitfalls (e.g., misrepresentation, marginalization) of MT in African languages can be critically assessed.

2. The Potentials of MT for African Languages

Africa is one of the most linguistically diverse continents, with over 2,000 languages spoken across its 54 countries. Despite this richness, most African languages are underrepresented in digital technologies, education systems, and government institutions. Machine Translation (MT), which uses artificial intelligence to translate text or speech between languages, offers a promising pathway to bridge linguistic gaps. This paper explores the potentials of MT in fostering linguistic inclusion and access, preserving and digitizing African languages, and promoting economic and social benefits.

2.1. Linguistic Inclusion and Access

Linguistic exclusion has long hindered equitable access to education, public services, and digital content in Africa. Machine translation has the potential to democratize access to information by enabling the translation of content into local languages. This can be particularly transformative in multilingual countries where government documents, educational materials, and health communication are often only available in colonial languages such as English, French, or Portuguese (De Pauw, et al.). During health emergencies, such as the COVID-19 pandemic, the lack of accessible information in local languages posed serious risks.

MT systems can play a critical role in translating public health messages into multiple African languages quickly and accurately (Senghor, et al.). Similarly, in the education sector, translating textbooks and online learning platforms into mother tongues through MT enhances comprehension and inclusiveness (Bird). Furthermore, linguistic inclusion through MT supports civic participation and legal access. Citizens are more likely to engage with governmental processes when they can access documents and services in their first language. This fosters trust, transparency, and accountability in governance.

2.2. Language Preservation and Digitization

Many African languages are endangered or under-documented, facing the risk of extinction due to globalization, urbanization, and generational language shifts. MT contributes to language preservation by creating digital resources and corpora from oral and written content in local languages (Besacier, et al.). As these languages are digitized and integrated into MT systems, they gain increased visibility and utility in digital spaces. Incorporating local languages into MT systems also involves compiling parallel corpora, developing orthographies, and engaging native speakers in the annotation process. These activities not only preserve linguistic data but also validate and empower communities by recognizing their linguistic heritage (Bird). MT tools, combined with speech recognition and text-to-speech technologies, can also facilitate the development of language learning applications. These tools provide educational resources for both native speakers and language learners, contributing to intergenerational transmission and revitalization efforts.

2.3. Economic and Social Benefits

MT for African languages opens wide economic opportunities. Businesses can use MT to localize advertisements, product information, and customer services, enabling them to engage with a broader and more linguistically diverse consumer base (Abbott). By reducing language barriers, MT encourages entrepreneurship and market expansion. In addition, the development and deployment of MT technologies create new job opportunities in the fields of linguistics, language data curation, and AI research. African universities and tech startups are increasingly contributing to the creation of language technologies tailored to local needs, driving innovation and capacity building. Socially, MT fosters greater inclusivity in digital spaces. African users gain access to global information and media in their own languages, promoting cultural exchange and knowledge equity (Bamgbose). This digital inclusion is crucial in ensuring that technological advancements benefit all populations, not just those fluent in dominant global languages.

Moreover, MT can enhance international collaboration in academic and scientific research. By translating scholarly works into and out of African languages, researchers can contribute to and benefit from global discourse more effectively (Besacier, et al.). Machine Translation offers transformative possibilities for African languages and societies. By promoting linguistic inclusion, preserving and digitizing endangered languages, and generating economic and social benefits, MT stands as a powerful tool for development and cultural sustainability. However, realizing these potentials requires overcoming challenges such as data scarcity, dialectal variation, and infrastructure limitations. Collaborative efforts among governments, researchers, technologists, and local communities are essential to create ethical and effective MT systems that reflect and respect Africa's linguistic diversity.

2.4. Enhancing Public Services and Government Communication

Governments across Africa can use MT to disseminate information in multiple languages, especially in health, legal, and civic domains. For instance, during the COVID-19 pandemic, language barriers hindered the delivery of critical health messages. MT could enable real-time translation of such messages into indigenous languages, increasing outreach and compliance (Senghor, et al.). Machine Translation offers transformative possibilities for African languages and societies. By promoting linguistic inclusion, preserving and digitizing endangered languages, and generating economic and social benefits, MT stands as a powerful tool for development and cultural sustainability. However, realizing these potentials requires overcoming challenges such as data scarcity, dialectal variation, and infrastructure limitations. Collaborative efforts among governments, researchers, technologists, and local communities are essential to create ethical and effective MT systems that reflect and respect Africa's linguistic diversity.

3. The Pitfalls and Limitations

Despite its potential, machine translation for African languages faces serious challenges. These include limited and poor-quality language data, the loss of cultural meaning, technological bias, and ethical issues related to ownership and representation. This section examines these key limitations to highlight the need for more inclusive and context-aware MT systems.

3.1. Data Scarcity and Quality

One of the most significant challenges facing the development of effective machine translation (MT) systems for African languages is the acute scarcity and uneven quality of linguistic data. Unlike widely spoken global languages such as English, French, or Spanish, most African languages suffer from a lack of parallel corpora—that is, texts that are translated between two or more languages and aligned at the sentence or phrase level. Parallel corpora are essential for training statistical and neural machine translation models, as they provide the foundational material from which the systems learn equivalences between languages (Adewumi, et al.; Joshi, et al.). In many African contexts, even monolingual corpora—large collections of text in a single language—are limited or non-existent. Where such data does exist, it may be fragmented, not digitized, or of poor quality due to inconsistent orthographic standards or transcription errors (Nekoto, et al.). This severely limits the ability of researchers and developers to train reliable MT models, which typically rely on large, clean, and diverse datasets for high performance.

Furthermore, African languages are often highly tonal and morphologically rich. Tones—variations in pitch that can change the meaning of a word—are a central feature in languages such as Yoruba, Igbo, or Ewe. Yet, many written corpora fail to consistently mark tone, leading to ambiguities and inaccuracies during model training and translation output (Bird). Morphological complexity—such as agglutination, noun class systems, and verbal extensions—adds another layer of difficulty. A single word in a

Bantu language like Zulu or Shona, for example, may encode what would require a full sentence in English. This mismatch in structure creates significant hurdles for both alignment in parallel corpora and for the design of MT systems that can handle such features appropriately (Orife).

Dialectal variation further compounds the problem. Many African languages exist in multiple regional or sociolectal forms. For instance, Swahili has standardized forms used in education and media, but also numerous dialects with lexical and syntactic differences (De Pauw, et al.). An MT model trained on data from one variety may perform poorly when applied to another, resulting in mistranslations or reduced intelligibility. The lack of sufficient, clean, and well-annotated data—combined with linguistic features like tone, complex morphology, and dialect diversity—poses a fundamental limitation to the scalability and quality of machine translation in African languages. Addressing these issues requires not only technical innovation but also sustained investment in language documentation, standardization, and community-driven data collection efforts (Orife, et al.).

3.2. Cultural and Semantic Loss

Another major limitation of machine translation in the African context lies in the loss of cultural and semantic nuances during the translation process. Language is not merely a tool for communication; it is a carrier of culture, worldview, and local knowledge (Banda; Gyasi). Many African languages are rich in oral traditions, proverbs, idiomatic expressions, and culturally embedded references that are difficult—if not impossible—to render accurately through machine translation systems trained primarily on Western languages and structures (Nekoto, et al.). Idioms and proverbs, for example, are central to African modes of expression. In Akan, the proverb *"Dua koro gye mframa a, ebu"* ("A single tree that resists the wind will break") communicates collective strength and interdependence. A literal machine translation of this phrase would likely render it as a simple botanical fact, completely stripping it of its metaphorical and social meaning. Without cultural grounding, MT systems often fail to identify such figurative language, resulting in flat, literal, and misleading outputs (Akinlade).

Moreover, semantic mismatches are common when African concepts are translated into European languages. Many African languages encode ideas and values that have no direct equivalents in English, French, or Portuguese

(Bamgbose; KÍARÍBẸ̀Ẹ́). For instance, the Yoruba concept of *"Àṣẹ"* conveys divine authority, power, and the ability to make things happen—an idea that blends spirituality, intention, and action. Machine translation systems typically reduce such complex ideas to vague, generic words like "command" or "authority," failing to capture their depth and cultural resonance (Adebara, and Abdul-Mageed). These challenges also extend to ritual language, kinship systems, and naming conventions, which are culturally specific and context-dependent. For example, in many African languages, kinship terms reflect age, lineage, or social role, unlike in English where terms are often broad or neutral (Eberhard; Heine). MT systems, without cultural context, flatten such semantic distinctions, weakening the socio-cultural message.

In addition, non-standardized orthographies and multilingual realities—where individuals switch between languages in a single discourse (code-switching)—make it even harder for MT tools to recognize and translate culturally meaningful content. This multilingual flexibility, common in both urban and rural African settings, challenges the rigid language boundaries assumed by most MT architectures (Banda; Nekoto, et al.). Ultimately, the risk of cultural and semantic loss in machine-translated African texts is not just a technical issue—it is an epistemological and ethical one. When MT systems fail to preserve cultural meaning, they contribute to linguistic impoverishment and the erasure of worldviews, reinforcing asymmetries in global communication (Bird). Addressing this problem calls for more than just better algorithms—it requires collaboration with linguists, cultural experts, and local speakers to develop translation systems that are context-aware and culturally sensitive (Orife).

3.3. Technological Bias and Linguistic Hegemony

Machine Translation (MT) systems, as with all technological tools, are shaped by the data and priorities of the societies that develop them. In the context of African languages, a significant challenge arises from the technological bias inherent in most machine translation (MT) models, which often serve to perpetuate linguistic hegemony (Bird; Joshi, et al.). The crux of the issue pertains to the fact that the majority of commercial and research-driven MT systems such as Google Translate, DeepL, or Meta AI's translation projects are predominantly trained on extensive data sets originating from European and Asian languages, where abundant resources are readily available. The following languages have benefited from decades of standardisation, digital corpus development and state-supported linguistic infrastructure (Adebara, and Abdul-Mageed; Anastasopoulos, and Neubig). In contrast, most African languages are under-resourced or considered "low-resource" in computational linguistics, leading to imbalances in translation accuracy (Marivate, and Sefara). For instance, while English-French MT yields high accuracy, performance significantly drops when African languages are introduced (Nekoto, et al.).

As a result, MT tools often treat African languages as add-ons rather than central components. This creates a situation where English or French becomes the default pivot language, and translation flows are often unidirectional: from African languages to dominant colonial languages or vice versa (De Pauw, et al.). Such asymmetry reinforces linguistic hierarchies that marginalize African languages in both digital and institutional communication (Mutale). In addition, biases in training data often reflect and perpetuate global inequalities. For example, if a machine learning model is trained mostly on government reports, religious texts, or formal media content, it may overrepresent elite discourse while neglecting informal registers, youth language, or indigenous epistemologies (Akinlade). This can result in skewed or stereotyped representations of African linguistic and cultural identity.
Moreover, technological design choices—such as user interfaces that exclude scripts, diacritics, or tonal markers—can marginalize languages with complex orthographies or phonological systems. Even when African languages are included, they are often offered with limited functionalities, making them second-class digital citizens in the global language technology ecosystem (Banda; Bird). This linguistic hegemony extends to policy and economic dimensions. Most MT systems are built and owned by companies and institutions in the Global North, raising concerns about control over African linguistic futures, economic benefit, and cultural representation. African voices are often missing in these processes, risking the transformation of MT from a tool of access to a mechanism of digital linguistic colonialism (Adewumi, et al.; Orife).

The technological bias and linguistic hegemony embedded in current MT systems not only limit their usefulness for African languages but also raise deeper ethical concerns about equity, representation, and linguistic justice. Redressing this situation requires the decentralization of MT development, inclusive data policies, and strong support for African-led language technology initiatives that prioritize local linguistic realities and cultural autonomy (Marivate, and Sefara; Orife).

3.4. Ethical Concerns

Beyond the technical and linguistic challenges, machine translation (MT) in African languages raises a few ethical issues that merit serious attention. These concerns relate to ownership, representation, equity, and exploitation, particularly in the collection and use of linguistic data and the development of MT systems for underrepresented languages (Bird; Marivate, and Sefara). One central ethical issue is the ownership of linguistic data. In many African countries, linguistic resources such as corpora, glossaries, and dictionaries are often gathered by foreign researchers, international organizations, or private tech companies, sometimes without transparent agreements or appropriate benefit-sharing mechanisms (Adewumi, et al.). The resulting data is often stored and monetized abroad, limiting local access and control—raising questions about data sovereignty and linguistic rights (Esselink, and Vries; Orife).

Closely linked to this is the issue of intellectual property. When MT systems built using African language data are commercialized, the profits flow to multinational corporations, while the local speakers and linguists who contributed the data and knowledge are rarely compensated or even acknowledged (Bamgbose; Orife). This dynamic reinforces a neo-colonial model of digital extraction, where African knowledge is commodified without return. Another ethical concern is inclusion and agency. Many MT initiatives in Africa are top-down, driven by the technical agendas of external institutions rather than local priorities. Such approaches often fail to engage communities, understand multilingual realities, or account for local language politics, resulting in poor adoption or harmful outcomes (Banda; Nekoto, et al.). In critical domains such as health or legal information, mistranslations from poorly localized MT systems can have real-life consequences, from misinformation to violations of rights (Adebara, and Abdul-Mageed).

There is also the danger of language standardization without consent. MT developers may choose one dialect or orthographic norm over others due to data availability, thereby marginalizing minority varieties and reinforcing linguistic hierarchies (De Pauw, et al.)). For example, favoring Standard

Swahili over regional variants can silence linguistic diversity and cultural expression in multilingual regions (Mutale). Lastly, privacy and informed consent are often overlooked in data collection for MT. Oral recordings, text samples, or social media data may include sensitive personal or cultural content, collected without full understanding or consent from contributors. This poses risks not only to individuals but also to community trust in technological research and institutions (Bird; Marivate, and Sefara). The ethical concerns around MT in African languages go far beyond technical limitations. They demand the development of inclusive, transparent, and just practices that center African voices and values. Sustainable and responsible MT development must ensure community participation, equitable benefit sharing, and cultural respect at every stage of the process (Orife, et al.).

4. Case Studies and Examples

To better understand the real-world dynamics of machine translation (MT) in African languages, it is important to examine case studies and concrete examples that highlight both the progress and the challenges in this space. These examples demonstrate the diversity of approaches, the unevenness of development, and the practical consequences of applying MT technologies in African linguistic contexts (Adewumi, et al.; Bird).

4.1. Google Translate: Uneven Language Support

One of the most well-known MT platforms, Google Translate, illustrates the inequality in language representation. While Google supports a handful of African languages such as Swahili, Hausa, Yoruba, and Zulu, many others- including Xhosa, Wolof, Twi, and Edo- are absent or poorly represented (Marivate, and Sefara). Even for supported languages, the quality of translation is inconsistent, often failing to render grammatical accuracy or cultural nuances (Joshi, et al.; Nekoto, et al.). For instance, Swahili translations tend to be more reliable due to a relatively larger corpus and regional standardization, whereas Yoruba translations often suffer from tone misrepresentation, grammatical errors, and semantic ambiguities (Adebara, and Abdul-Mageed). This disparity reflects the broader issue of linguistic hierarchies in MT development, where resource availability influences language prioritization and system performance (Bamgbose; Bird).

4.2. Masakhane: A Decentralized, African-Led Initiative

In contrast to top-down MT development, Masakhane represents a grassroot, pan-African movement focused on building machine translation models for African languages by Africans, for Africans. Launched in 2019, Masakhane brings together linguists, researchers, and technologists from across the continent to collaboratively create open-source MT systems (Orife). What makes Masakhane impactful is its emphasis on local ownership, inclusivity, and ethical collaboration. The initiative supports over 40 African languages, using community-sourced data that respects linguistic variation and cultural diversity. Beyond tool development, it invests in human capacity building in computational linguistics across Africa. Masakhane's model demonstrates that decentralized, community-driven approaches can produce high-quality, context-aware MT while empowering local stakeholders (Nekoto, et al.).

4.3. Mozilla Common Voice: Community Data Collection

Mozilla's Common Voice project also contributes positively to African language technologies by enabling volunteers to record and donate voice data in their native languages. Languages such as Igbo, Kinyarwanda, Luganda, and Tigrinya have notably benefited (Foundation; Solaiman, et al.). Although Common Voice focuses primarily on speech data for applications like speech-to-text and voice recognition, it plays an important complementary role in multimodal resource development, essential for robust MT systems. Its participatory model encourages community engagement and ethical data collection, democratizing access to digital language tools and reinforcing open-source culture in African NLP (Bird; Esselink, and Vries).

4.4. Failures and Limitations in Practice

There are also cautionary examples where poor implementation of MT in African contexts led to negative consequences. In some e-governance platforms or public health campaigns, African language MT tools were deployed without adequate testing or cultural adaptation, leading to miscommunication and public mistrust (Akinlade). For instance, during the COVID-19 pandemic, MT-generated health advisories translated into local African languages frequently included semantic errors and mistranslations of technical terms, distorting critical public health messages (Marivate, and Sefara; Orife, et al.). Similarly, chatbots or digital assistants deployed in urban

African contexts have failed to understand dialectal variation, code-switching, or idiomatic expressions, thereby alienating users and widening the digital divide (Bird; Nekoto, et al.). These examples underline the complex interplay between technological capacity, linguistic diversity, and sociocultural context in African MT. They affirm that while innovations exist, success hinges on inclusive design, ethical practices, and continuous community collaboration (De Pauw, et al.; Orife).

5. Recommendations and Future Directions

To address the pitfalls outlined in previous sections and harness the full potential of machine translation (MT) for African languages, a multi-layered and collaborative approach is essential. This section proposes strategic recommendations and outlines promising future directions for building more inclusive, effective, and ethical MT systems tailored to Africa's linguistic realities.

5.1. Building Inclusive and Diverse Language Resources

The development of reliable MT systems begins with the creation of large, diverse, and high-quality corpora. For most African languages, this involves not just digitizing existing texts but also collecting, annotating, and standardizing data in collaboration with local communities (Bird; Marivate, and Sefara). Efforts like Common Voice (Foundation) and Bantu Speech Corpora initiatives show that open, participatory data collection is both feasible and ethical. To ensure inclusiveness, language resources must cover regional dialects, speech registers, genres, and modalities (text, audio, visual). Without such diversity, MT systems risk privileging elite or standardized forms, reinforcing digital inequality (Adewumi, et al.; Mutale).

5.2. Strengthening African-Led Research and Development

There is an urgent need to invest in African computational linguistics and AI talent. Initiatives like Masakhane, AI4D Africa, and the Data Science Africa Network provide successful examples of African-led research and innovation (Nekoto, et al.; Orife). Governments, universities, and development agencies should offer long-term support for training programs, research fellowships, and interdisciplinary collaboration. Moreover, the decentralization of MT model development- enabling local teams to define goals, build tools, and test applications in real-world contexts- promotes relevance, empowerment, and sustainability (Orife).

5.3. Promoting Ethical Standards and Data Justice

To avoid exploitative practices and protect community interests, MT development must be guided by ethical frameworks. These should include the following:

- Informed consent from data contributors (Esselink, and Vries).
- Transparent data governance, including clear licensing and community access (Foundation).
- Equitable benefit sharing with contributors and local institutions (Banda).
- Protection of minority dialects and cultural expressions from erasure or misuse (Joshi, et al.).

Ethical guidelines should be co-developed with African linguists, community leaders, and digital rights organizations to ensure legitimacy and accountability.

5.4. Designing Context-Aware Translation Systems

African MT systems must go beyond literal translation to handle code-switching, tone, morphology, and cultural references. This requires hybrid models that integrate rule-based systems, neural networks, and human-in-the-loop editing (Akinlade). In critical sectors like healthcare, education, and law, automated translation should always be reviewed by human experts to avoid misinformation and harm (Marivate, and Sefara). Models should also be domain-sensitive, adapting to genres such as oral narratives, bureaucratic texts, or informal social media posts, which are central to African language use.

5.5. Encouraging Policy and Institutional Support

Governments, regional bodies (e.g., African Union, AfCFTA), and educational institutions should:

- Officially recognize African languages as digital languages of science and technology (Bamgbose).
- Provide policy frameworks for language digitization, open data sharing, and inclusive tech development (Adewumi, et al.).

- Encourage public-private partnerships to scale ethical MT solutions, especially for marginalized communities.

Pan-African collaboration could promote cross-border linguistic research, create continental language data repositories, and strengthen regulatory protections for African digital linguistics.

Conclusion

The future of machine translation in African languages is filled with potential-but also with critical challenges. If guided ethically and inclusively, MT can help bridge language barriers, preserve linguistic diversity, and empower local communities across the continent. However, without thoughtful intervention, it risks reinforcing existing inequalities and eroding cultural richness. Africa's linguistic future in the digital age depends on building MT systems that reflect the continent's plurality of voices, complexity of expression, and cultural depth. This calls for collaboration between linguists, technologists, policymakers, and language communities, working together toward a common goal: making African languages visible, viable, and vibrant in the machine age. The integration of machine translation (MT) into African linguistic contexts presents both transformative opportunities and critical challenges. On one hand, MT technologies have the potential to significantly enhance access to information, education, and public services across a multilingual continent, supporting broader goals of inclusion, development, and cultural preservation. They can facilitate cross-border communication, contribute to the digital representation of under-resourced languages, and help bridge the language divide that has historically limited access to global and regional discourse.

However, the application of MT in African languages is constrained by several key pitfalls, including limited language data, the linguistic complexity and diversity of African languages, and a lack of culturally aware translation models. Most current MT systems are trained on high-resource languages, leading to performance disparities and potential misrepresentation of African languages and cultures. Addressing these challenges requires interdisciplinary collaboration among computational linguists, local language communities, policymakers, and technologists. Sustainable progress will depend on the development of inclusive datasets, investment in language technologies for low-resource languages, and the incorporation of sociocultural context into algorithmic design. Only by confronting these issues can machine translation evolve into a truly equitable tool that supports linguistic diversity and digital inclusion in Africa. MT will only fulfill its promise in Africa if it is made by and for African speakers."

Works Cited

Abbasi, R. A., et al. "The Challenges of Low-Resource Language Translation: A Review." *ACM Computing Surveys*, 2022, pp. 1–30, https://doi.org/10.1145/3485123.

Abbott, A., et al. "Data for African Languages: A Challenge and an Opportunity." *Proceedings of the AfricaNLP Workshop at EACL*, 2021.

Adebara, Ife, and Muhammad Abdul-Mageed. "Towards Afrocentric NLP for African Languages: Where We Are and Where We Can Go." *arXiv preprint* arXiv:2203.08351, 2022.

Adewumi, Tosin, et al. "AfriWOZ: Corpus for Exploiting Cross-Lingual Transferability for Generation of Dialogues in Low-Resource, African Languages." *arXiv preprint* arXiv:2204.08083, 2022.

Akinlade, O., I. Orife, and V. Marivate. "Towards a Nuanced Evaluation of African Language Machine Translation." *Proceedings of the AfricaNLP Workshop*, 2021.

Anastasopoulos, Antonios, and Graham Neubig. "Should All Cross-Lingual Embeddings Speak English?" *arXiv preprint* arXiv:1911.03058, 2019.

Bamgbose, Ayo. "African Languages Today: The Challenge of and Prospects for Empowerment under Globalization." *Selected Proceedings of the 40th Annual Conference on African Linguistics*, Cascadilla Proceedings Project, 2011, pp. 1–14.

Banda, Felix. "Critical Perspectives on Language Planning and Policy in Africa: Accounting for the Notion of Multilingualism." *Stellenbosch Papers in Linguistics PLUS*, vol. 38, 2009, pp. 1–11.

Bar, François, et al. "Mobile Technology Appropriation in a Distant Mirror: Baroque Infiltration, Creolization and Cannibalism." *Seminario sobre Desarrollo Económico, Desarrollo Social y Comunicaciones Móviles en América Latina*, 2007, pp. 20–21.

Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610–23.

Besacier, Laurent, et al. "Automatic Speech Recognition for Under-Resourced Languages: A Survey." *Speech Communication*, vol. 56, 2014, pp. 85–100.

Bird, Steven. "Decolonising Speech and Language Technology." *28th International Conference on Computational Linguistics (COLING 2020)*, Association for Computational Linguistics (ACL), 2020, pp. 3504–19.

De Pauw, Guy, et al. "Exploring the SAWA Corpus: Collection and Deployment of a Parallel Corpus English—Swahili." *Language Resources and Evaluation*, vol. 45, no. 3, 2011, pp. 331–44.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, editors. *Ethnologue: Languages of the World*. 2020.

Esselink, Bert, and Arjen-Sjoerd Vries. *A Practical Guide to Localization*. 2000.

Foundation, Mozilla. "Common Voice: A Platform for Multilingual Voice Data Collection." *Common Voice*, 2023, https://commonvoice.mozilla.org.

Gyasi, Kwaku Addae. *The Francophone African Text: Translation and the Postcolonial Experience*. vol. 48, Peter Lang, 2006.

Heine, Bernd, and Derek Nurse, editors. *African Languages: An Introduction*. 2000.

Hershcovich, Daniel, et al. "Participatory Approaches to NLP for African Languages: Lessons from Grassroots Collaborations." *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022, pp. 1025–36.

Joshi, Pratik, et al. "The State and Fate of Linguistic Diversity and Inclusion in the NLP World." *arXiv preprint* arXiv:2004.09095, 2020.

Kíaríbẹ̀ẹ́, Luqman Abísọ́lÁ. *A Comparative Performance Semiotics of Orin Kete and Orin Agbè among the Ìbàràpá and Òkè-Ògùn People of Yorùbá, Nigeria*. University of Ìbàdàn, 2023.

Marivate, Vukosi, and Tshephisho Sefara. "Improving Short Text Classification through Global Augmentation Methods." *Machine Learning and Knowledge Extraction*, vol. 4, Springer, 2020, pp. 385–99.

Martinus, Abbott. "Low-Resource Machine Translation: A Review of African Languages." *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*, 2021, pp. 384–95.

Mutale, K., I. Orife, and I. Adebara. "Language Technology and the Reproduction of Linguistic Inequity in Africa." *Proceedings of the AfricaNLP Workshop*, 2023.

Nekoto, Wilhelmina, et al. "Participatory Research for Low-Resourced Machine Translation: A Case Study in African Languages." *arXiv preprint* arXiv:2010.02353, 2020.

Orife, I., et al. "Masakhane: Named Entity Recognition for African Languages." *arXiv preprint* arXiv:2011.02347, 2020.

Orife, Iroro, et al. "Masakhane—Machine Translation for Africa." *arXiv preprint* arXiv:2003.11529, 2020.

Senghor, Abdou Simon, et al. "Towards a Transactional Medicine Approach to Combating Global Emerging Pathogens: The Case of COVID-19." *Global Public Health*, vol. 18, no. 1, 2023, p. 2272710.

Sennrich, Rico. "Challenges and Opportunities in Machine Translation of Low-Resource Languages." *Annual Review of Linguistics*, vol. 7, 2021, pp. 277–93.

Solaiman, Irene, et al. "Evaluating the Social Impact of Generative AI Systems in Systems and Society." *arXiv preprint* arXiv:2306.05949, 2023.

Whorf, Benjamin Lee. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* MIT Press, 2012.

Yimam, S. M., et al. "Exploring Amharic and Tigrinya for Machine Translation and Part-of-Speech Tagging." *LREC 2020 Proceedings*, 2020, pp. 1411–19.

About the Author

Ulrich Douo is a Lecturer in the Department of French at the University of Ghana, Legon. His research and work focus on comparative literature, feminism and translation studies. Dr. Douo is also a translator. Since 2018, he has been employed as a French translator at the United Nations Office for Project Services (UNOPS), a specialised agency of the United Nations. He also works with organisations such as the African Agricultural Technology Foundation (AATF), African Continental Free Trade Area (AfCFTA), ECOWAS, ICCA Consortium, Korea International Cooperation Agency - Ghana (KOICA), KPMG-Ghana, Department of African Studies & Merian Institute for Advanced Studies in Africa (MIASA) of the University of Ghana, Ruzizi III Energy Limited (Rwanda), West African Tax Administration Forum (WATAF), and Nigerian National Petroleum Corporation (NNPC).

– UIRTUS –
vol. 5, no. 2, August 2025 ISSN 2710-4699 Online

How to cite this article/Comment citer cet article:
MLA: Douo, Ulrich. "The Role of Machine Translation in African Languages: Potentials and Pitfalls." *Uirtus*, vol. 5, no. 2, 2025, pp. 38-57, https://doi.org/10.59384/uirtus.2025.2947.