Maintaining the communicative vibrancy of Nigeria's over 250 languages in the era of Artificial Intelligence

Ifeanyi Chukwunonso Ogbodo
Doctoral student, Faculty of Law and Administration
Adam Mickiewicz University in Poznań, Poland.
ifeogb@amu.edu.pl

Abstract

Nigeria covers a vast land area of 923,768 km², roughly equal to the combined size of Poland, Germany, the Netherlands, Belgium, and Italy, and it is home to over 250 distinct ethnic groups. Each group has its own history, traditional political system (ranging from centralized emirates to acephalous village democracies), art forms, and origin myths. Nigeria has 250 indigenous languages with 270 sub-variations, each reflecting unique ways of thinking, feeling, and socializing. In cities such as Lagos, Port Harcourt, and Abuja, one can hear 30–50 languages in a single day on the streets.

These languages are not static "codes" but constantly evolving: they borrow, blend, shrink, expand, and reinvent themselves through daily conversation, music, migration, urban youth culture, and digital messaging. This dynamism makes human communication in Nigeria especially resilient and creative, yet it also poses one of the biggest challenges for language technologies and artificial intelligence systems, which are mostly built on a few dominant languages.

This study explores how the fluid, context-sensitive, and deeply social nature of communication in Nigerian languages both enriches and complicates the development of inclusive AI tools, and offers recommendations.

I. Background

Nigeria has over 220 million people with diverse languages and cultures. However, AI tools such as Google Translate and ChatGPT are trained on major languages, including English, Chinese, and Spanish, partly because these languages have many online resources, so AI learns them easily. Nigeria's languages (such as Hausa, Yorùbá, Igbo, and smaller ones) lack online resources and as a result, AI struggles with them. If this continues, it could lead to language decline, with young people abandoning their ancestral languages and some languages gradually disappearing. This article examines Nigeria's language diversity and AI, highlighting challenges like limited data and opportunities, such as developing AI that understands Nigerian accents and mixed languages. It emphasizes the need to incorporate Nigerian languages into AI to ensure that the technology benefits everyone, preserves culture, and promotes inclusivity. Proper integration can help AI protect and celebrate Nigeria's rich diversity instead of damaging it.

II. Linguistic Diversity in Nigeria

Nigeria has more than 520 indigenous languages, reflecting its diverse ethnic groups. This number ranges from 250 to 370, depending on whether a language is classified as a core language of an ethnic group or as a variation. Based on the most recent Ethnologue data, Nigeria is home to 525 living languages, mainly from the Niger-Congo family, with some from Afroasiatic and Nilo-Saharan branches (Eberhard et al., 2023).

The three main ethnic groups in Nigeria are the Hausa (including Fulani), Yoruba, and Igbo. These ethnic groups collectively represent about 60–70% of the population. The Hausa are the largest group at around 30%, followed by Yoruba with approximately 15–16%, and Igbo also around 15%. The remaining population includes smaller groups such as the Tiv (2–3%), Kanuri, Ijaw, Edo, Urhobo, Nupe, and others (Simons & Fennig, 2018). People's locations often align with their ethnic groups; Hausa is primarily spoken in northern Nigeria, with different variations. In the southwest, Yoruba dominates with different variations; in the southeast, Igbo is common. Smaller groups have their own languages across the country, especially in central regions and along the coast. In cities such as Lagos, Abuja, and Port Harcourt, there is greater linguistic blending. Many people speak multiple languages, often switching between them during conversations. Nigerian Pidgin English, a lively mix of English and local languages, is widely used daily to communicate across groups.

Pidgin English primarily uses English words, simplifies its grammar significantly, and incorporates elements from local languages. It is not anyone's first language; instead, people learn it as a second language. For example, "How you dey?" means "How are you?"; "*I dey fine*" means "I'm fine"; "*Wetin dey happen*?" means "What's happening?"; "*I wan chop*" means "I want to eat."

Nigeria's diverse cultures are rich in songs, stories, proverbs, and traditions expressed in many forms, but this diversity is at risk. Some languages have disappeared, and many are endangered (Moseley, 2010). Young people in cities prefer English or Pidgin for school, work, and modern life. As global connectivity increases through the internet and travel, smaller languages risk fading if not passed on to children.

III. The Dynamism of Nigerian Languages

Nigerian languages are continually changing and expanding. People in Nigeria readily borrow words from other languages, such as English, Arabic, French, and nearby African languages, and adapt them to their own speech. This trend of borrowing is not new. It has occurred for hundreds of years through trade, religion, and colonialism, and it continues today (Yakpo, 2024; Faraclas, 1996).

Ifeanyi Ogbodo, *Preserving the Communicative Dynamism of Nigeria's 250+ Languages in the Age of Artificial Intelligence*. In: Z. Vetulani, P. Paroubek, I. Kardava (Eds.), Human Language Technologies as a Challenge for Computer Science and Linguistics – 2025, pp. 63–65. Poznań: Adam Mickiewicz University Press, 2025.
© Author, 2025. https://doi.org/10.14746/amup.9788323245414.12. Open Access chapter, distributed under the terms of the CC licence (BY-NC-ND, https://creativecommons.org/licenses/by-nc-nd/4.0/).

i. New Words and Slang from Music, Movies, and Social Media

In cities like Lagos, Abuja, and Port Harcourt, young people are always coming up with fresh words and phrases in their everyday chats. Afrobeats music and Nollywood movies disseminate catchy expressions that children and adolescents love to repeat. Music can make a piece of slang famous nationwide almost overnight. On platforms such as WhatsApp, Instagram, and TikTok, users create shortcuts and combine words. For example, “Sapa” means being broke, and “japa” means moving abroad to find better opportunities. Notably, these new words still conform to the language's traditional rules when discussing modern life. This keeps Nigerian languages feeling alive, fun, and full of energy.

ii. Migration and City Life

Many Nigerians move from rural villages to big cities in search of better opportunities. When they do, they often mix their home languages with those they encounter, creating novel ways of speaking. Nigerian Pidgin started as a simple way for people who spoke different languages to communicate. Over time, it has grown into a whole language with its own grammar, proverbs, and rich vocabulary. Today, more than 100 million people speak it, and it continues to evolve by incorporating new words and expressions from young people, music, and social media (Mensah, 2011).

iii. Why this Change is a Strength

This continuous evolution facilitates communication among Nigerians, making it adaptable and inventive, enabling individuals to convey intricate emotions, humor, or social contexts in a natural and compelling manner. Furthermore, it aids in preserving languages amid the dominance of English, as speakers prefer adaptation to abandoning their native language.

iv. The Downside

When schools, dictionaries, or AI systems teach or use a language, they need everyone to adhere to a single consistent version with the same spelling and grammar rules. This clarifies and makes things easier to understand. However, languages change quickly over time. The way people speak and pronounce words keeps evolving, so it is hard to agree on one single "correct" way forever. Additionally, in Nigeria, different regions have their own ways of speaking the same language, and new slang emerges all the time, especially among young people. This natural variety makes it even more challenging to establish a fixed standard.

v. Language Shift Among Young People

Many children and teenagers favor speaking English or Pidgin, considering them modern, practical, or “cool.” They tend to use indigenous languages less frequently at home or among friends, resulting in a decline in fluent speakers over time, particularly for less widely spoken languages. Consequently, this trend poses a risk to the preservation of cultural knowledge, including proverbs, folktales, history, and traditional values, as the language gradually diminishes.

IV. Challenges of Building AI for Nigerian Languages

Given the background, it is not easy to develop AI tools such as speech recognition, translation applications, or chatbots for Nigerian languages compared with those for languages like Polish or English. Furthermore, there is a noticeable bias in dominant Large Language Models that reinforces the exclusion of African languages. This accounts for the limited training data available for African languages (Muhammad et al., 2023).

i. Dearth in data:

Most AI models perform well on common languages due to abundant data (Joshi et al., 2020). The paucity of available data on Nigerian languages would ordinarily pose a challenge considering that many of the languages are mainly oral and lack written online content. Illustratively, AI learns from billions of words in English across various sources, whereas Nigerian languages often lack such digital data, and the available data are inconsistent. There are existing projects like Masakhane that aim to improve datasets for small languages, but progress has been slow, primarily because they are volunteer-dependent (Adelani et al., 2021; Nekoto et al., 2020).

ii. Diacritics and Spelling:

There is also the complexity of tone in characters of many Nigerian languages, which AI often mishandles. Many characters use special marks, such as accents, to indicate sounds or tones. In Igbo, a pitch (high or low tone) can completely change a word's meaning, even when the spelling is the same. A classic example is the word ákwá (high tone on both syllables: á-kwá), which means "cry" or "weeping"; àkwá (low tone on the first syllable, high on the second: à-kwá) means "egg". It is worth noting that people often omit these special marks when typing on phones or computers, resulting in inconsistent spelling. Part of the reason for this willful omission is that Nigerians do not use keyboards with keys for the thousands of special characters that span across indigenous languages. The Nigerian official language is English, and keyboards and phone keypads are designed with the English language alphabet, which is non-tonal. Lastly, Nigerians frequently switch between languages within a single sentence, for example, from English to Pidgin to Yoruba in a 12-word sentence, which confuses AI systems.

iii. Computational and Ethical Constraints

Training AI requires advanced technology and reliable power, which researchers in Nigeria often lack, making the operation of large models expensive. Secondly, to whom should the ownership of data comprising voices

and texts from communities be attributed? This may raise the issue of benefit sharing. There is also the issue of whether participants understood their privacy rights and consented appropriately; conversely, an overemphasis on these issues can result in the collection of limited data, which may lead to an AI model that perpetuates harmful stereotypes or overlooks cultural nuances.

Western-centric data can lead AI to misinterpret Nigerian proverbs, humor, and social norms, thereby limiting its usefulness in education, health, and government services. However, community projects and startups are gathering better data and developing local tools. With more investment, collaboration, and ethical focus, AI can become inclusive of Nigeria's diverse languages.

V. Opportunities for AI to evolve.

The manner in which Nigerians communicate is distinctive. It amazes foreigners that tone changes the meanings of words in context, and new words regularly emerge from music, social media, and daily life. This fluidity adds fun and creativity to communication. It can also help make AI smarter and better for everyone around the world, especially in places where people blend tongues, by handling the dynamics of switching between languages, comprehending accents, and understanding words in contexts of jokes, proverbs, and cultural hints (Orife et al., 2020).

There are a few existing projects on this. One is the NaijaVoices, which has over 1,800 hours of spoken recordings from thousands of Nigerian speakers in Hausa, Igbo, and Yorùbá, covering ages, accents, dialects, and ranging from casual to formal speech. The promoters intend to model AI into understanding natural speech and accents, improving voice tools like Siri or Google Assistant for Nigerians, and aiding local language apps for health, farming, and education.

Another one is N-ATLAS modelled to understand Yorùbá, Hausa, Igbo, and Nigerian-accented English. It converts speech to text, generates subtitles, summarizes talks, and engages in natural conversation. It fosters AI inclusivity for Nigerians in their languages. MasakhaNER was developed to teach AI to recognize names of people, places, and organizations in African languages. It improves AI's understanding of local news, searches, and chats, extracting important info more accurately (Emezue et al., 2022) The impact of these projects, which were aimed at preserving cultural knowledge, supporting education, promoting economic inclusion, and preserving language, has not yet been felt. There is a need for more projects in this regard because of the promising prospect of keeping Nigeria's rich cultures alive using AI.

VI. Recommendations

i. There is a need for the creation of open-source datasets through crowdsourcing and partnerships that address the lack of high-quality data for Nigerian languages. Crowdsourcing involves native speakers contributing voice recordings and text, capturing diverse accents and slang. Open sourcing prevents tech companies from hoarding data and allows local developers to create affordable health and education tools.

ii. There is also a need for investments in the standardization of Nigerian languages while respecting their linguistic variation. Nigerian languages have dialects, tones, and flexible spellings, like skipping diacritics in Yorùbá. Total standardization could erase diversity, but guidelines are necessary for AI. Governments and universities should fund linguists to develop flexible standards, similar to Bible translations and news corpora, enabling AI to handle fluid communication without losing creativity. There is need for fostered collaborations among academia, industry, and communities.

iii. Enforcing ethical data practices that emphasize sovereignty is needful. This ensures that Nigerians control their data. Rules should be created that mandate local storage, community consent, and benefit sharing, such as jobs or tools for contributor

VII. References

Adelani, D. I., et al. (2021). MasakhaNER: Named Entity Recognition for African Languages. Transactions of The Association for Computational Linguistics, 9, 1116–1131.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2023). Ethnologue: Languages of the World (26th ed.). SIL International.

Emezue, C., et al. (2022). NaijaVoices: Speech Datasets for Nigerian Languages. Lacuna Fund.

Faraclas, N. (1996). Nigerian Pidgin. Routledge

Joshi, P., et al. (2020). The State and Fate of Linguistic Diversity and Inclusion in the NLP World. Proceedings of ACL.

Mensah, E. (2011). Lexicalization in Nigerian Pidgin. Concentric: Studies in Linguistics.

Moseley, C. (Ed.). (2010). Atlas of the world's languages in danger (3rd ed.). UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000187026

Orife, I., et al. (2020). Masakhane – Machine Translation for Africa. arXiv preprint arXiv:2003.11529.

Simons, Gary F. and Charles D. Fennig (Editors). (2018). Ethnologue: Languages of the World. Twenty-first edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Yakpo, K. (2024). West African Pidgin: World Language Against the Grain. Journal of Language Contact.