

From Swahili to Zulu, African techies develop AI language tools



To build the model, the project will also draw on the expertise of more than 7,000 fellows from Nigeria's tech talent programme, a government scheme to train three million people in skills such as coding and programming. File photo.

Image: 123RF/ALEXANDER KIRCH

When the Nigerian government announced plans in April to develop a multilingual AI tool to boost digital inclusion across the West African nation, 28-year-old computer science student Lwasinam Lenham Dilli was thrilled.

Dilli had struggled to scrape datasets from the internet to build a large language model (LLM), used to power AI chatbots, in his native Hausa language as part of his final year project at university.

languages will not be forgotten or left out of the AI ecosystem.

The world has been swept up in a whirlwind of AI mania, with tools such as OpenAI's ChatGPT, Meta's Llama 2, and Mistral AI captivating millions globally with their ability to generate human-like text.

However, for many tech-savvy African, the excitement has been tempered by a frustrating reality: when languages like Hausa, Amharic, or Kinyarwanda are entered into the chat, many of the advanced systems falter, often producing nonsensical responses.

Technology experts warn the lack of LLMs in African languages will lead to the exclusion of millions of people on the continent, increasing the digital and economic divides.

The Nigerian government-led initiative to develop a multilingual LLM aims to level the playing field.

"The LLM will be trained on five low-resource languages and accented English to ensure stronger language representation for development of artificial intelligence solutions," Nigeria's digital economy minister Bosun Tijani said in April.

The government will partner with Nigerian AI startups, and local data will be collected by volunteers who are fluent in any of five Nigerian languages, namely Yoruba, Hausa, Igbo, Ibibio and West African lingua franca, Pidgin.

To build the model, the project will also draw on the expertise of more than 7,000 fellows from Nigeria's tech talent programme, a government scheme to train three million people in skills such as coding and programming.

Silas Adekunle, co-founder of Awarri, an AI startup that is part of the initiative, said building a nuanced AI tool that understood Nigeria's unique language and cultural landscape presented many challenges.

source machine learning to translate African languages and Lelapa AI, a research lab, has pioneered VulaVula, a for-profit language processing tool that translates, transcribes and analyses languages in English, Afrikaans, Zulu and Sesotho.

"The scale of the project, specially with limited resources, has required us to be creative in how we train the model, gather the data, compute and label what we have."

Africa is home to more than 2,000 languages spoken across 54 countries, according to Unesco.

However, most African languages remain underrepresented on the internet. English dominates the digital space, accounting for around 50% of all websites, followed by Spanish, German, Japanese and French.

Along with the Nigerian government initiative, there are also a small but growing number of African startups rising to the challenge of developing AI tools in languages such as Swahili, Amharic, Zulu and Sesotho.

In Kenya, for instance, health tech firm Jacaranda Health has pioneered the first LLM operating in Swahili to improve maternal healthcare in East Africa.

Built on Meta's Llama 3 system, UlizaLlama (AskLlama) aims to refine Jacaranda Health's SMS service for low-income Swahili-speaking expectant mothers who have queries ranging from dietary concerns and foetal movement to exercise during pregnancy.

The platform provides pre-written automated responses, but once UlizaLlama is integrated by the end of June, it will tailor responses to individual needs, offering more detailed pregnancy guidance and emergency support.

...were shooting for about 85% accuracy to start with and a faster response time. At the moment it takes a few minutes to respond, but we are hoping to get that down to less than a minute in the future."

In South Africa, the Masakhane initiative is using open-source machine learning to translate African languages.

Lelapa AI, a South African AI research lab, has pioneered VulaVula, a for-profit language processing tool that translates, transcribes and analyses languages in English, Afrikaans, Zulu and Sesotho.

However, AI experts said building LLMs in African languages poses significant challenges, ranging from availability of data to ethical concerns over consent, compensation and copyright.

Many African languages are low-resource languages, meaning there is a scarcity of data to train the models effectively, unlike high-resource languages such as English or French.

Michael Michie, co-founder of Everse Technology Africa, an AI startup building intelligence into data protection and privacy, said collecting the data needed to train LLMs also raised ethical questions.

In many African communities oral tradition predominates, and certain communities may not be interested in sharing their language to train LLMs and this should be respected.

"There are no regulations or laws in African countries that address issues related to consent, privacy and compensation to communities when collecting data to train AI tools. This needs to be addressed," said Michie.

"There are questions of who owns the language and who benefits. There needs to be guidelines to prevent exploitation and ensure the development of the LLMs benefits the people they are meant to serve," he said.

Open-source initiatives such as Creative Commons, which allow creators to legally share their work with specified conditions such as

science at the University of Pretoria and co-founder of Elapa AI.

However, if everything is open source, it may be harder to properly reimburse and acknowledge the original contributors to these language models, he said.

"A lot of people are working on LLMs because of the prestige. That's where the money is, but we need to make sure our languages are being taken care of."

[Thomson Reuters Foundation](#)



READER LETTER | Let AI robots replace politicians for smarter choices

There has been much in the papers lately about the potential dangers of artificial intelligence and the real danger of natural stupidity, mainly in ...

OPINION *1 month ago*



Microsoft introduces smaller AI model

Microsoft on Tuesday announced a cost-effective, small language artificial intelligence model that can perform tasks such as content creation and ...

NEWS *1 month ago*



Baidu says AI chatbot 'Ernie Bot' has amassed 200-million users

China's Baidu said on Tuesday its artificial intelligence chatbot "Ernie Bot" has garnered more than 200-million users as it seeks to remain China's ...

NEWS *2 months ago*



ANDREAS PESCHKE | Meaning of freedom that must never be taken for granted but defended



next year, says Tesla's Musk

Tesla CEO Elon Musk on Monday predicted development of artificial intelligence that was smarter than the smartest human probably by next year, or by ...

BUSINESS *2 months ago*
