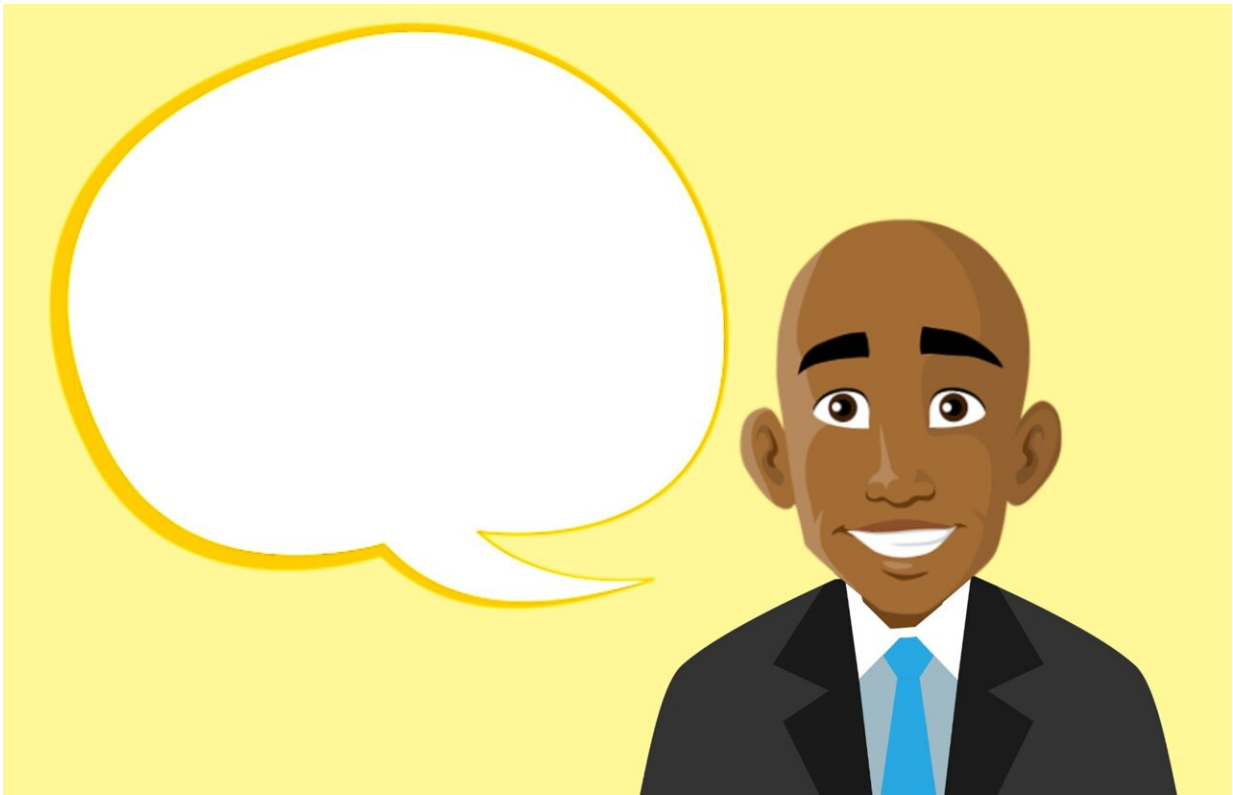# A roadmap to help AI technologies speak African languages

August 11 2023

Credit: Pixabay/CC0 Public Domain

From text-generating ChatGPT to voice-activated Siri, artificial intelligence-powered tools are designed to aid our everyday life—as long as you speak a language they support. These technologies are out of reach for billions of people who don't use English, French, Spanish or

other mainstream languages, but researchers in Africa are looking to change that.

In a study published in the journal *Patterns*, scientists draw a roadmap to develop better AI-driven tools for African languages.

"It doesn't make sense to me that there are limited AI tools for African languages," says first author and AI researcher Kathleen Siminyu of the Masakhane Research Foundation, a grassroots network of African scientists who aim to spur accessible AI tools for those who speak African languages. "Inclusion and representation in the advancement of language technology is not a patch you put at the end—it's something you think about up front."

Many of these tools rely on a field of AI called natural language processing, a technology that enables computers to understand human languages. Computers can master a language through training, where they pick up on patterns in speech and text data. However, they fail when data in a particular language is scarce, as seen in African languages.

To fill the gap, the research team first identified key players involved in developing African language tools and explored their experience, motivation, focuses, and challenges. These people include writers and editors who create and curate content, as well as linguists, software engineers, and entrepreneurs who are crucial in establishing the infrastructure for language tools.

Interviews with the key players revealed four central themes to consider in designing African language tools:

- First, bearing the impact of colonization, Africa is a multilingual society where African language is central to people's cultural

identities and is key to societal participation in education, politics, economy, and more.

- Second, there is a need to support African content creation. This includes building basic tools such as dictionaries, spell checkers, and keyboards for African languages and removing financial and administrative barriers for translating government communications to multiple national languages, which includes African languages.
- Third, the creation of African language technologies will benefit from collaborations between linguistics and computer science. Also, there should be focus on creating tools that are human centered, which help individuals unlock greater potential.
- Fourth, developers should be mindful of communities and ethical practices during the collection, curation, and use of data.

"There's a growing number of organizations working in this space, and this study allows us to coordinate efforts in building impactful language tools," says Siminyu. "The findings highlight and articulate what the priorities are, in terms of time and financial investments."

Next, the team plans to expand the study and include more participants to understand the communities that AI language technologies may impact. They will also address barriers that may hinder people's access to the technology.

The team hopes their study could serve as a roadmap to help develop a wide range of language tools, from translation services to misinformation-catching content moderators. The findings may also pave the way to preserve indigenous African languages.

More information: Kathleen Siminyu, Consultative Engagement of Stakeholders Towards a Roadmap for African Language Technologies, *Patterns* (2023). DOI: 10.1016/j.patter.2023.100820.

Provided by Cell Press