

AI and African Languages in Education

What we're learning

THE TWO QUESTIONS THIS BRIEF EXPLORES

01

What **opportunities and risks** does AI present for African languages in South African schools, including the goals of Mother Tongue-based Bilingual Education (MTbBE)?

02

How might AI **support or undermine** the development, use, and resourcing of African languages in schools?

PHOTO CREDIT: UNICEF SOUTH AFRICA

DESIGNED & CONVENED BY

human**studios**

STRATEGIC PARTNERS



TECHNICAL AND FUNDING PARTNERS



Drawing existing evidence into practical *insight*.

This is one of nine themes in **Lefa AI**, a rapid research series on what AI means for South African basic education. It has intentionally been produced quickly, to keep up with the rate of change and strengthen decision making around AI in basic education.

That speed came with a trade-off: this series doesn't contribute new research, but draws together existing literature and expert experience into practical insight. Throughout, we've aimed to **weigh opportunities and risks fairly**, not argue for or against AI in schools.

1

rapid review

of existing literature on AI and indigenous languages, AI in language learning, and MTbBE

4

expert interviews

with South African academics, an AI specialist and an education programme manager

Where this conversation starts

Mother Tongue-based Bilingual Education

South Africa has committed to teaching children in their home language. Mother Tongue-based Bilingual Education (MTbBE) is an approach where learners are taught and assessed in their home language while gradually learning an additional language, such as English. At the same time AI is moving into the basic education sector.

A thin evidence base

Global evidence on AI in education is growing fast, but focuses on high income countries and is primarily based on English. For South Africa's indigenous African languages, **the evidence base for meaningful AI integration is limited.**

This research set out to map what is known, identify where the real opportunities lie, and **flag where the risks are greatest**.

Key findings

01

AI performance in African languages is constrained by a lack of available language data.

South Africa's indigenous languages are considered “**low-resource**” languages for AI because they lack sufficient, high-quality digital language datasets. Existing datasets are often narrow in scope and are not designed for education. Subject matter experts also highlighted complex grammar, multiple dialects and widespread code-switching, which AI tools cannot yet manage well.

SOME INDIGENOUS LANGUAGES HAVE MUCH MORE DIGITAL DATA AVAILABLE THAN OTHERS

MORE DIGITAL LANGUAGE DATA AVAILABLE FOR AI TRAINING

isiZulu

isiXhosa

Sepedi

VERY LIMITED DIGITAL LANGUAGE DATA AVAILABLE FOR AI TRAINING

Tshivenda

Xitsonga

Siswati

The scarcer the data, the longer it will take to reach reliable performance.

Sources: [Mussandi & Wichert 2024](#); [Marivate et al. 2023](#); expert interviews.

02

Personalised, adaptive learning has produced real gains, but not yet in an African language.

SPAIN

+1 month

reading progress

For every session of personalised adaptive learning
5-year longitudinal study

INDIA

Higher

maths and Hindi scores

Learners using an AI-powered adaptive platform

NIGERIA

1.5 - 2 yrs

of schooling gained in English with an AI tutor

SOUTH AFRICA

≈1 year

English literacy gain

23,000 Grade 3 learners
266 under-resourced schools

Across all four studies, learners made substantial gains from personalised AI support. **The largest improvements were consistently seen among learners who had fallen furthest behind**, suggesting that AI can provide support at an appropriate level for each learner.



However, none of these gains have been demonstrated in an African language, yet. While the potential is promising, it cannot be assumed that these results will transfer automatically to South Africa's multilingual classrooms.

Sources: [Cuevas-Ruiz et al. 2025](#); [Muralidharan et al. 2019](#); [De Simone et al. 2025](#); [Borole et al. 2025](#).

03

AI-assisted reading assessment is one of the most promising near-term uses of AI in South African classrooms.

Automatic Speech Recognition (ASR) tools listen to children reading aloud and assess their reading automatically. They are one of the most promising near-term applications of AI because they **support teachers by providing timely assessment and feedback, rather than replacing professional judgment.**

GHANA

Accurate oral reading assessment in real classrooms

An off-the-shelf ASR tool accurately assessed children's English oral reading fluency despite noisy classrooms and regional accents.

SOUTH AFRICA

AI reading assessment in an African language

Researchers built an open isiXhosa child speech dataset and trained AI models to assess early-grade pronunciation of words and letters. An important step towards practical reading assessment in large South African classrooms.

Sources: [Henkel et al. 2025](#); [Chevtchenko et al. 2025](#).

04

AI should support, not replace, teacher and expert judgement.

Across both the literature and interviews there was strong agreement that **human oversight is essential**. AI-generated content requires **linguistic, pedagogical and cultural review**, particularly in African languages where datasets remain limited.

“ AI outputs can be multilingual but monocultural.

DR JILL WALKER RETTBERG

AI can generate educational content in an African language while still reflecting assumptions, examples and cultural perspectives drawn largely from Western training data. Human expertise remains essential to ensure AI supports, rather than undermines, meaningful learning in South African classrooms.

Sources: [UNESCO 2023](#); [Ferrera 2023](#); [Rettberg 2022](#); Prof Leketi Makalela interview.

05

How a tool is designed decides whether learners actually learn.

In a Turkish study reported by the OECD (2026), learners with AI tutors did better in practice exercises but the same or worse in closed-book tests, they were leaning on the tool instead of doing the thinking, what the OECD calls “**cognitive offloading**.” By contrast, AI tools designed to prompt learners to explain, reason and solve problems for themselves, a **Socratic design**, supported deeper learning.

Professor Leketi Makalela adds that in a multilingual classroom this design has to be deliberate: an AI tool should be built around **planned, lesson-aligned use of two languages**.

Sources: [OECD 2026](#); [Mohebbi 2025](#); [De Simone et al. 2025](#); [Burns 2026](#); Prof Leketi Makalela interview.

Considerations for South African basic education

01

Build tools for MTbBE, don't borrow them from English.

MTbBE increases demand for high-quality bilingual resources, and AI could help meet that demand. The risk is that most available AI tools are not built for African languages, cultures or classrooms. Unless they are intentionally redesigned for the South African context, they risk reproducing the linguistic, cultural and pedagogical assumptions embedded in the data on which they were trained, rather than supporting meaningful multilingual learning.

02

Manage bias, rather than avoid AI because of it.

AI can carry and amplify cultural and linguistic biases, for example favouring dominant dialects. As Professor Leketi Makalela notes, these biases already appear in materials made by people carrying the “logic of English” into African language content; the response is expert oversight and curation at every step.

03

Invest in African language data sets.

Without investment in curriculum-aligned African language datasets, AI tools are unlikely to perform reliably in South African classrooms. This is the single constraint sitting underneath every other consideration on this page.

04

Address South African copyright law.

South Africa's current copyright rules are stricter than those in many other countries. They restrict the ethical use of high-quality existing materials, like dictionaries, textbooks and grammars, to train better AI models. South Africa's attempt to modernise its copyright law has stalled. Given how few good African language resources exist, this is an urgent gap.

Questions for further *exploration*

- Given that the schools most in need of multilingual support are often the least equipped digitally, what investment would stop AI from **widening rather than narrowing** existing inequalities?
- What skills, training, and support do South African teachers need to use AI as a collaborative resource for bilingual lesson preparation, particularly where teachers may not speak the home languages of all learners in their class?
- Do South African education organisations already hold valuable African language datasets that could support African language AI development, and what ethical, legal and governance considerations would need to be addressed first?



Language is a legacy, passed from *one generation* to the next. AI must be built with African languages, and the cultures and worldviews they carry, at its heart.

NOTES ON THIS RESEARCH WWW.LEFAAI.HUMANSTUDIOS.ORG

SCOPE AND METHOD

This brief explores the opportunities and risks that AI presents for African languages in South African basic education, including its implications for Mother Tongue-based Bilingual Education (MTbBE).

This rapid research brief draws on a rapid literature review and four key informant interviews with South African academics, an AI specialist, and an education programme manager. Classroom teachers and learners were not interviewed, so this brief does not include their direct perspectives.

EVIDENCE

Literature was selected for relevance to African languages, basic education and classroom practice; sources are cited in full beside each finding.

Interviews were conducted remotely, recorded with consent and analysed thematically. Quotations are used with permission.

LIMITATIONS

No AI tool was tested or piloted in a South African classroom, in any African language. No causal claims about impact are made.

Not an exhaustive review. AI research in education is moving quickly and further work may have been published after this brief was finalised.

Findings drawn from studies conducted outside South Africa may not transfer to South African multilingual classrooms.

USE AND RIGHTS

Views expressed are those of the programme and its contributors, and do not necessarily reflect the positions of NASCEE, JET Education Services, DGMT or UNICEF.

Photography © UNICEF South Africa. Images are illustrative and the individuals shown are not connected to the findings.

For corrections or to request the underlying source list, contact the Lefa AI programme team at connect@humanstudios.org.

© 2026 Human Studios. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). <https://creativecommons.org/licenses/by-nc-sa/4.0/>

DESIGNED & CONVENED BY



STRATEGIC PARTNERS



TECHNICAL AND FUNDING PARTNERS

