# AI in STEM with African Languages

Building understanding, trustworthy AI use and learner participation

Dr Seani Rananga | NSTF-proSET | 19 August 2026

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact




Photo: JeffNyoka / Wikimedia Commons / CC BY-SA 4.0

# Introduction: STEM learning begins with understanding language



- Before learners can solve a word problem, follow an experiment or evaluate evidence, they must understand what they read and hear.

- The Foundation Phase is therefore not separate from future AI and STEM participation; it establishes the language and reasoning on which later learning depends.

- When reading for meaning is weak, AI may produce more text without producing more understanding.

- The starting question is: can AI help learners access STEM concepts in the languages through which they already make meaning?

DBE Early Grade Reading Study; DBE Reading Benchmarks (2022) | Photo: FintonCassidy, Wikimedia Commons, CC BY-SA 3.0

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# South Africa’s reading evidence shows why AI must build comprehension

## GRADE 3

Curriculum expectation: read for meaning by the end of the Foundation Phase

## 15%

of assessed Grade 3 learners could not read one word in the passage (Funda Uphumelele)

## 81%

of South African Grade 4 learners were below the PIRLS Low International Benchmark (2021)

**For STEM, AI must help learners interpret instructions, evidence and explanations—not merely generate answers.**

DBE Funda Uphumelele National Survey (2025); DBE (2022); PIRLS South Africa (2021)

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# A STEM learner must read, reason and explain



- Science requires learners to follow procedures, identify claims and connect observations to explanations.
- Mathematics requires understanding the language of a problem before selecting an operation or model.
- Engineering and technology require interpreting constraints, instructions, diagrams and safety information.
- AI should strengthen these practices—not bypass them by supplying answers that learners cannot evaluate.

NSTF/proSET Concept Document (2026) | Photo: Peace Corps, Wikimedia Commons, public domain

SADiLaR South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# AI can support the literacy-to-STEM pathway

• Speech and text: read instructions aloud, support pronunciation and allow questions through voice.

• Multilingual explanation: restate a concept in a familiar language and connect it to the formal terminology used in class.

• Scaffolding: break a complex task into smaller steps and provide examples at the learner's pace.

• Multimodal learning: combine words, images, audio, demonstrations and simulations.

• Feedback: help learners revise an explanation—but the teacher and learner must still verify accuracy.




UNESCO (2023); Kasneci et al. (2023) | Photo: JeffNyoka, Wikimedia Commons, CC BY-SA 4.0

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# African languages make AI support usable in the classroom



- Learners can ask questions in the language they use at home and in community life.
- Teachers can bridge home-language understanding with English technical terminology instead of forcing a choice between them.
- Speech technologies can support learners whose strongest language practices are oral.
- Local examples and terminology can connect abstract STEM concepts to familiar environments.
- The aim is additive multilingualism: strengthen understanding in African languages while expanding access to the wider scientific register.

UNESCO (2024); DBE Language in Education Policy | Photo: Peace Corps, Wikimedia Commons, public domain

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

“Language as culture is the collective memory bank of a people’s experience in history.”

Ngũgĩ wa Thiong’o, Decolonising the Mind (1986, p.15)

# The forum asks whether AI strengthens or weakens scientific thinking

• AI can help solve technical problems, but learners must understand how an answer was produced.

• A fluent response can contain fabricated facts, weak reasoning or an incorrect translation.

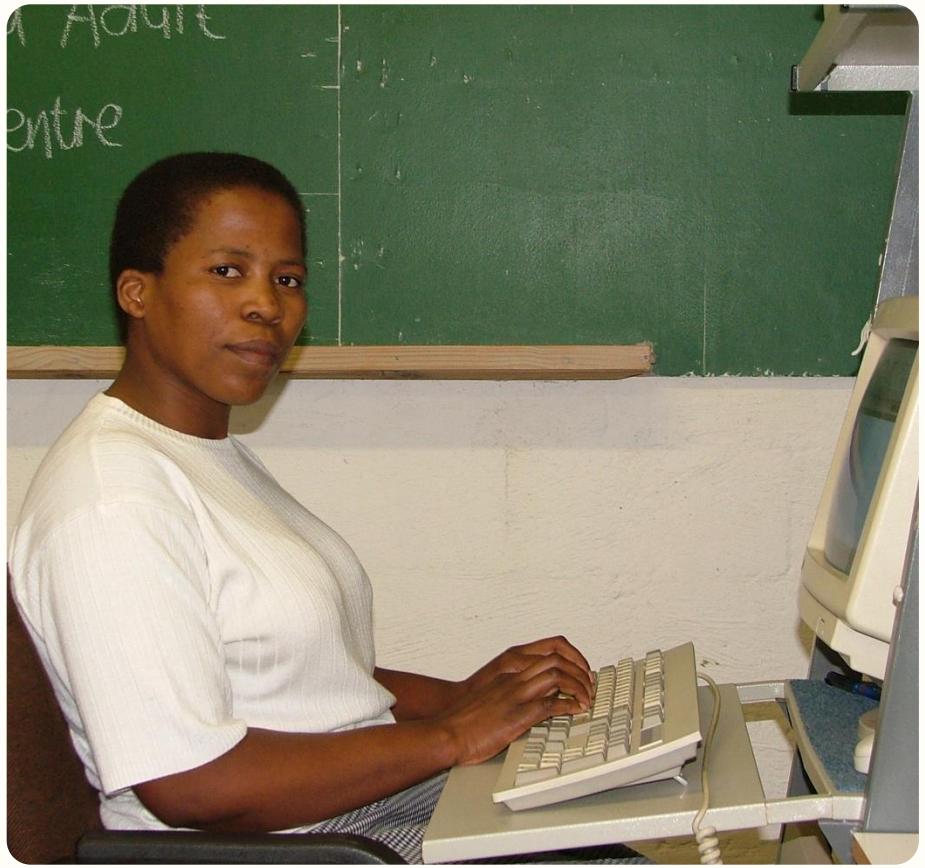
• Best practice must therefore include verification, source checking, comparison and communication of uncertainty.

• This is where multilingual misinformation becomes a practical AI-in-STEM teaching problem.




NSTF/proSET Concept Document (2026) | Photo: JeffNyoka, Wikimedia Commons, CC BY-SA 4.0

# Multilingual misinformation is a STEM education problem



- Misinformation is false or misleading information shared regardless of intent; disinformation is deliberately deceptive.

- A scientific claim may circulate as English or isiZulu text, a Sepedi and etc, voice note, a meme, video or code-switched post.

- If an AI system performs poorly in the learner's language, it may explain, translate or classify the claim incorrectly.

- Learners therefore need both multilingual access and the scientific habits required to test what AI tells them.

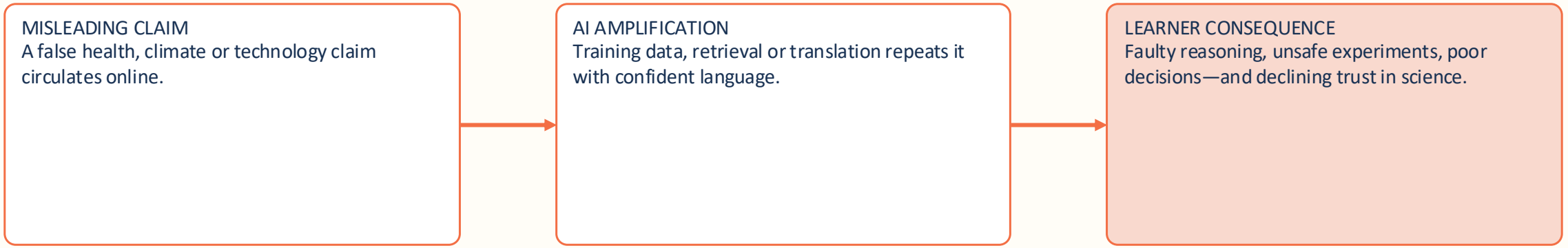
UNESCO (2023); Rananga (2026) | Photo: Marius140%, Wikimedia Commons, CC BY-SA 4.0

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

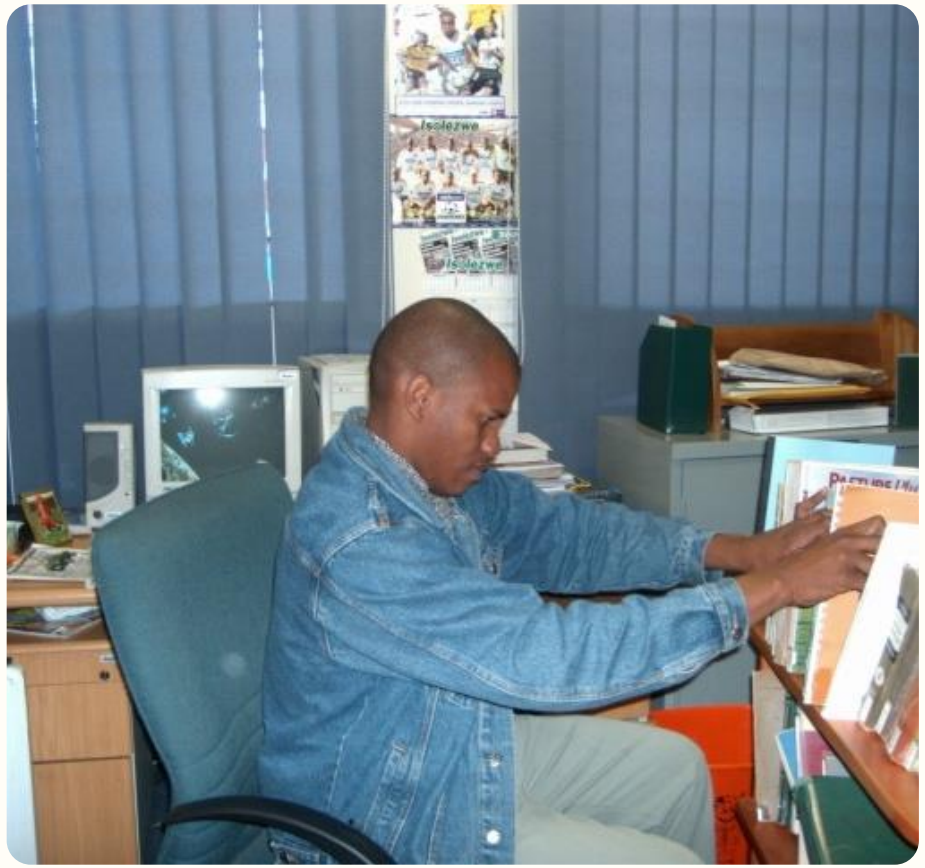
# A false claim can move from a phone into a STEM lesson




**The educational response: teach verification, make evidence visible, and evaluate AI in every language learners use.**

UNESCO (2023); Huang et al. (2025)

# My PhD research tested multilingual misinformation detection



- The study focused on COVID-19 claims in English, isiZulu and Sepedi.
- It connected translation fidelity, language representation, classifier performance and explainability.
- The dataset contained 7,590 labelled English claims; validation and test translations were checked by native-language speakers.
- The purpose was not only to label a post, but to understand where an AI decision becomes more or less reliable.

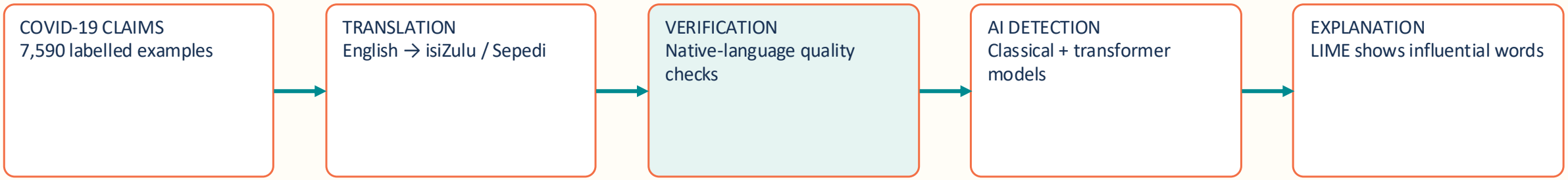
Rananga (2026), PhD thesis, North-West University | Photo: Bianca Lawrence53, Wikimedia Commons, CC BY-SA 4.0

SADiLaR South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# The PhD study turned multilingual misinformation into a testable AI pipeline




**Research developed at NWU; its methods now inform continuing work and collaboration at UP/DSFSI. The STEM question is how language, data and model design change reliability.**

Rananga (2026)

# One claim, three languages, one prediction task

ENGLISH
“Drinking hot water cures COVID-19.”

PREDICTION
MISINFORMATION

ISIZULU
Reveal on next click

WAITING

SEPEDI
Reveal on next click

WAITING

**Illustrative system view: the same scientific claim is evaluated across languages while the model output remains explainable and open to human review.**

Illustrative claim; workflow based on Rananga (2026)

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# One claim, three languages, one prediction task

ENGLISH
“Drinking hot water cures COVID-19.”

PREDICTION
MISINFORMATION

ISIZULU
“Ukuphuza amanzi ashisayo kwelapha i-COVID-19.”

PREDICTION
MISINFORMATION

SEPEDI
Reveal on next click

WAITING

**Illustrative system view: the same scientific claim is evaluated across languages while the model output remains explainable and open to human review.**

Muhammadi Kamoun, Mkululi Sidumo Ndlangalavi (2026)

# One claim, three languages, one prediction task

ENGLISH
“Drinking hot water cures COVID-19.”

PREDICTION
MISINFORMATION

ISIZULU
“Ukuphuza amanzi ashisayo kwelapha i-COVID-19.”

PREDICTION
MISINFORMATION

SEPEDI
“Go nwa meetse a go fiša go alafa COVID-19.”

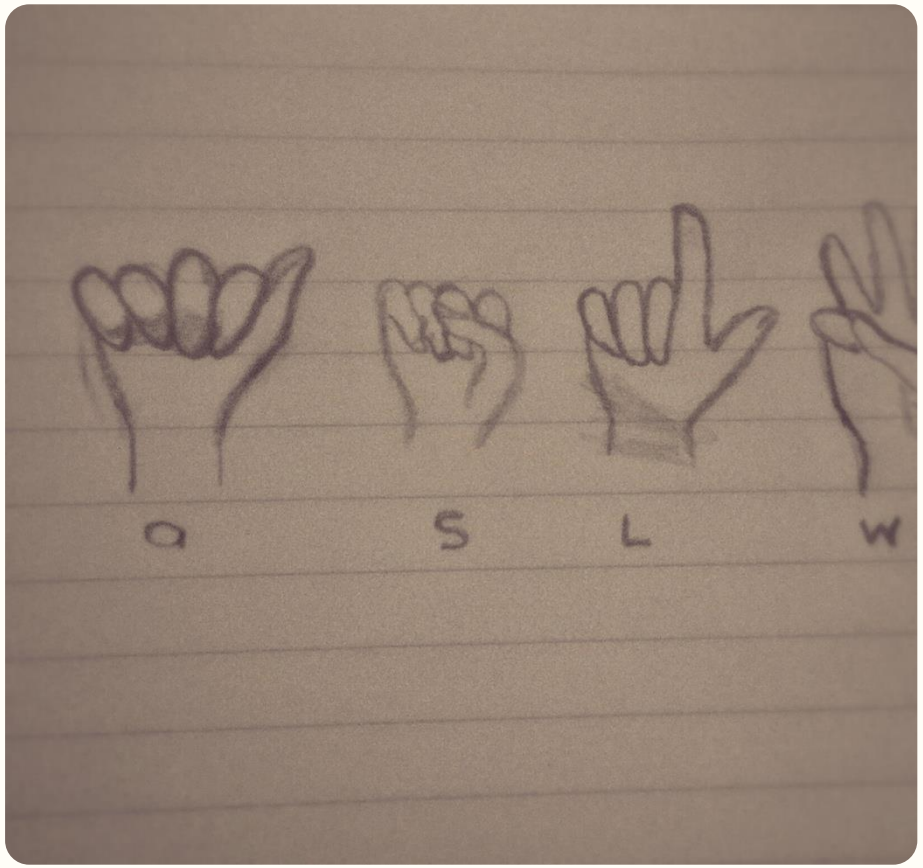
PREDICTION
MISINFORMATION

**Illustrative system view: the same scientific claim is evaluated across languages while the model output remains explainable and open to human review.**

Multilingual misinformation detection

# The next phase is multilingual and multimodal misinformation

- Languages: extending towards all South African official languages.
- Modalities: text, images, audio, video and South African Sign Language.
- Domains: public health, elections, water crises and other high-impact information environments.
- Goal: detect harmful claims while preserving linguistic, cultural and visual context.



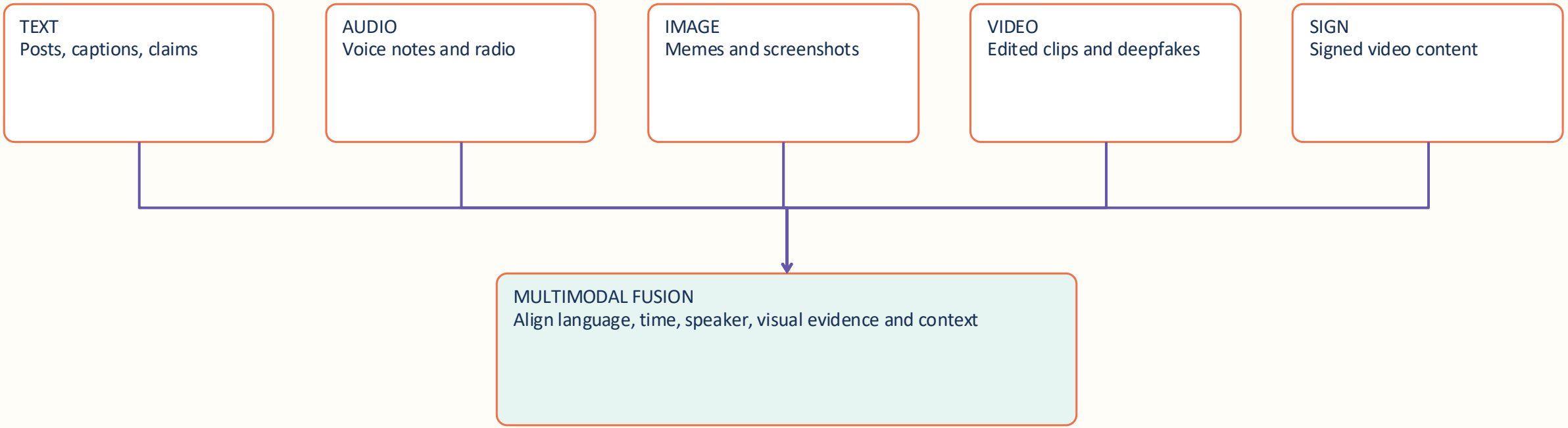

UP/DSFSI research programme | Photo: Wikimedia Commons, CC BY 4.0

SADiLaR South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

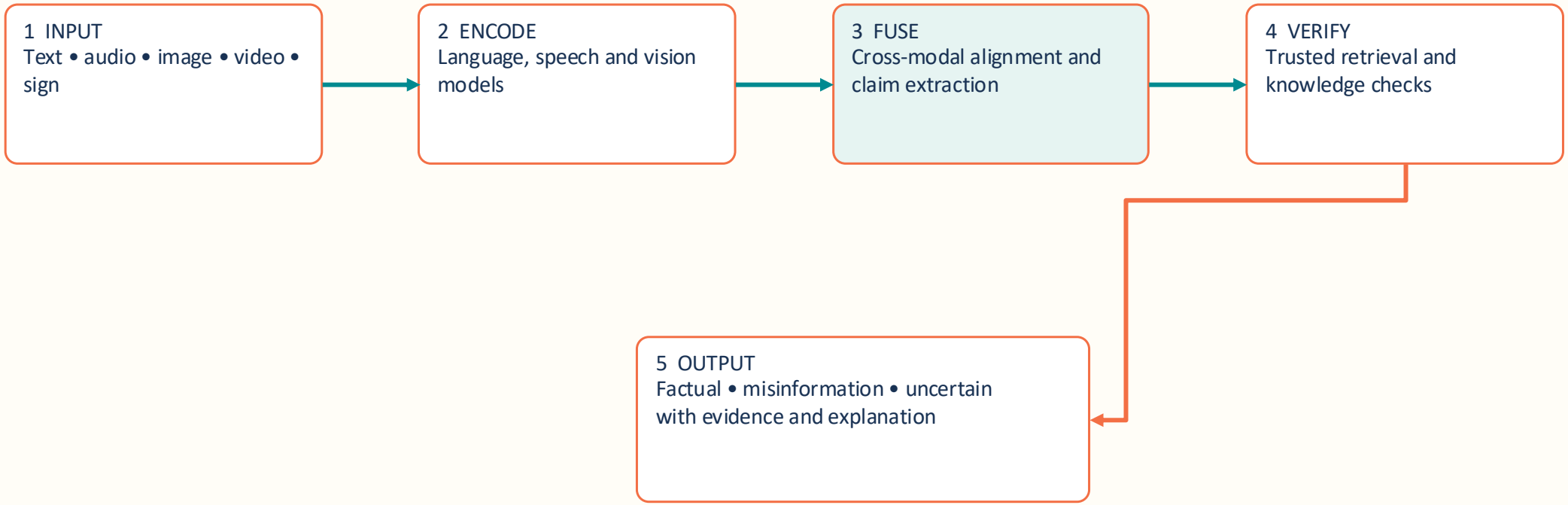
Data Science for Social Impact

# Multimodal misinformation requires evidence from several channels



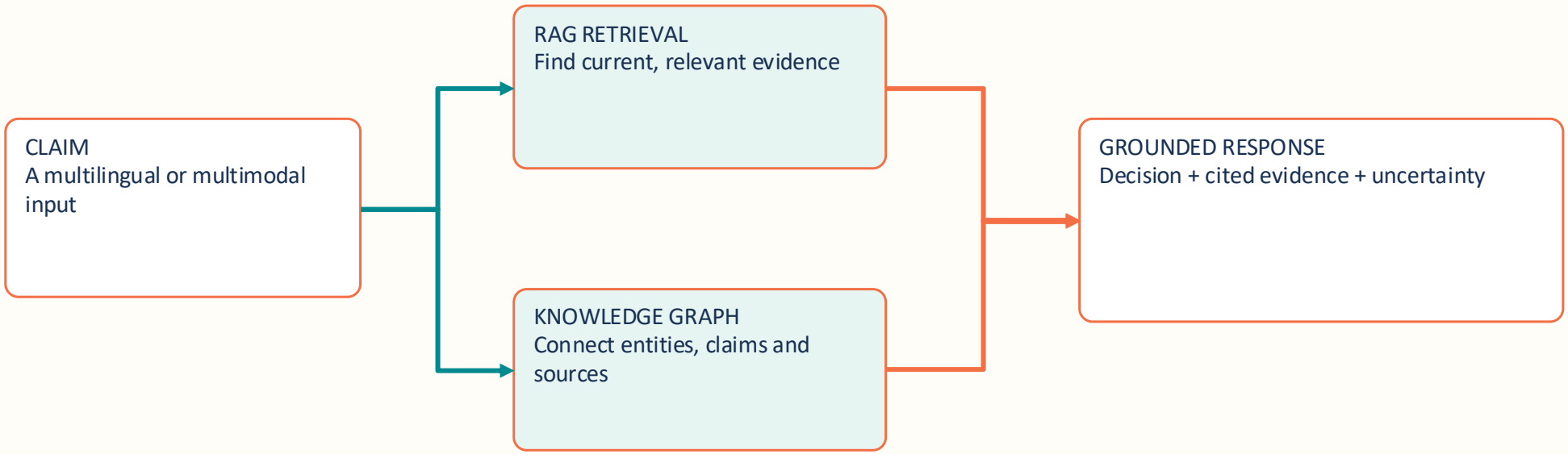

**A misleading claim may be carried by the words, the image, the speaker, the edit—or the interaction among them.**

# The multimodal research architecture combines detection, evidence and explanation




**Human review remains essential for high-risk, ambiguous and culturally sensitive cases.**

# RAG and knowledge graphs ground AI responses in trusted evidence




**In STEM education, the same architecture can teach learners to trace evidence rather than accept an answer because it sounds confident.**

# SADiLaR can connect African languages to AI-in-STEM practice

- Develop and curate text, speech and multimodal language resources.
- Support technical and scientific terminology that teachers and learners can use.
- Provide machine-readable resources for multilingual tools, datasets and evaluation.
- Connect language specialists, computational researchers, educators and communities.
- Build capacity so African languages are used to create technology—not only translated after systems are built.



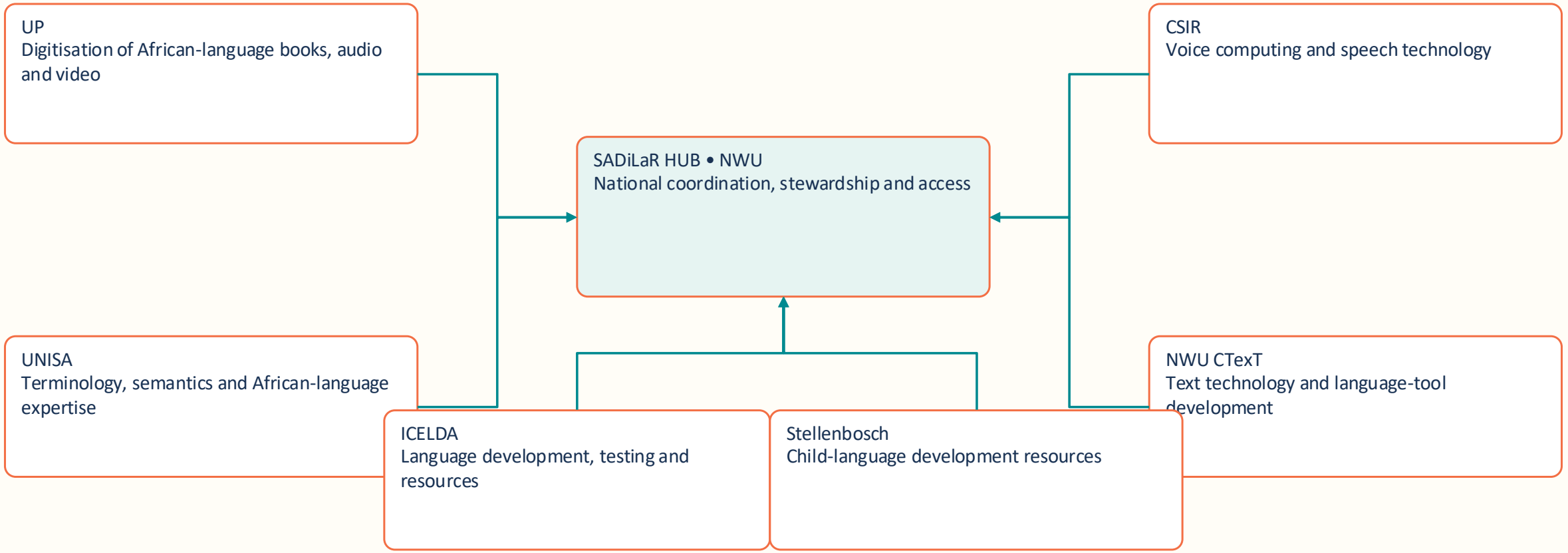
SADiLaR Nodes and HESSP (2026) | Photo: Peace Corps, Wikimedia Commons, public domain

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# SADiLaR works through a national network of complementary expertise




**Together, the network connects language knowledge, reusable data and computational capability—the foundations required for multilingual AI-in-STEM.**

SADiLaR Nodes (2026)

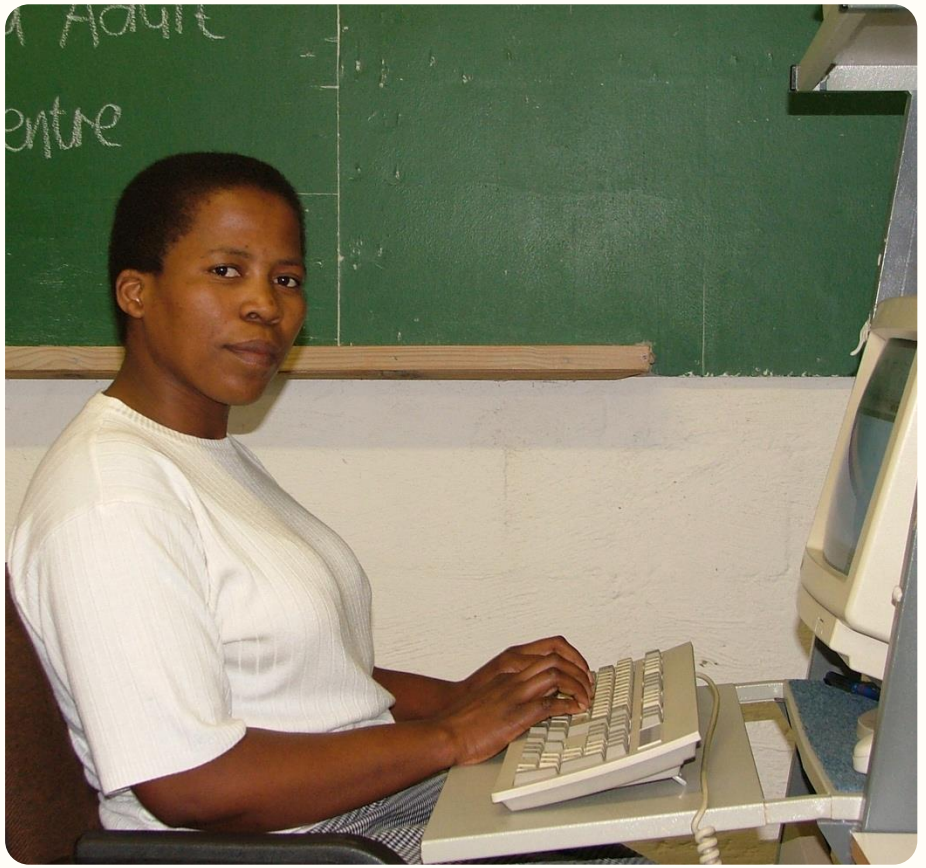
# Terminology links understanding, teaching and computation

• Learners need accessible explanations of concepts such as algorithm, probability, bias, evidence and uncertainty.

• Teachers need agreed terms, definitions and examples that can be used consistently.

• AI systems need those terms encoded in dictionaries, corpora, datasets and interfaces.

• Terminology development therefore connects classroom understanding with the computational representation of African languages.



SADiLaR HESSP; NSTF Language and STEM Education (2017) | Photo: Marius140%, Wikimedia Commons, CC BY-SA 4.0

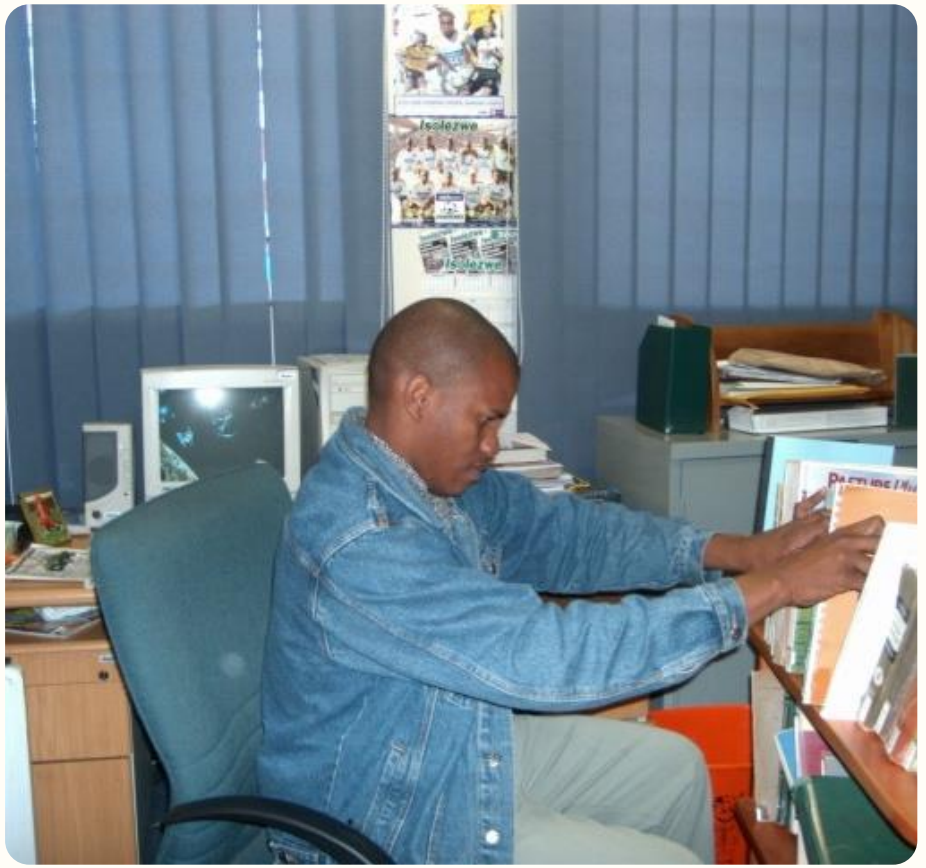
# Hundzula brings language and computational researchers together

• The UP–SADiLaR collaboration brings African languages into the computational space through focused, retreat-based research.

• Current participation is mainly at researcher level; it is not yet a school programme.

• The collaboration can produce corpora, dictionaries, annotations, models, prototypes and research capacity.

• Its value for AI-in-STEM is that educational tools require both language knowledge and computational expertise.



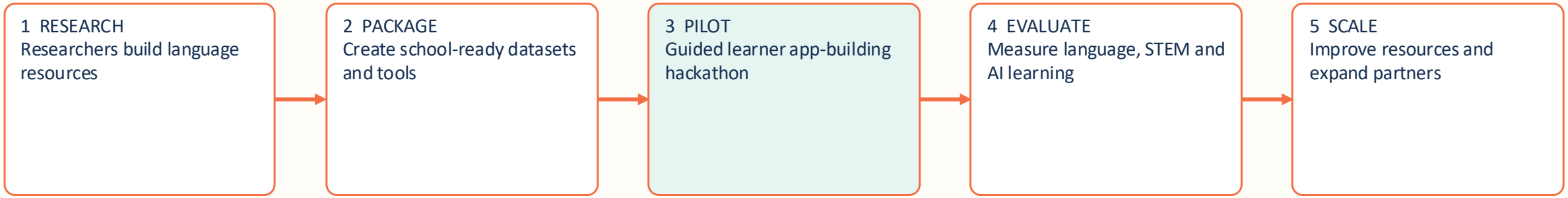
Supplied Hundzula presentation (2026) | Photo: Bianca Lawrence53, Wikimedia Commons, CC BY-SA 4.0

SADiLaR South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# A Hundzula-to-school action plan can move from research to participation




**Current position: researcher collaboration. Proposed next step: a small, ethically governed learner pilot—not a claim that a school programme already exists.**

# A learner hackathon can move from AI use to AI creation

- Learners identify a language barrier in health, environment, education or community information.
- They build a small glossary, dataset, chatbot, speech tool, fact-checking interface or multimodal prototype.
- Teachers, language experts and computational researchers guide the work.
- Projects are evaluated for scientific accuracy, language quality, usability, ethics and cultural fit.
- The aim is not a once-off event, but a pathway from learner ideas to reusable language resources.




Proposed direction; not yet an active school programme | Photo: JeffNyoka, Wikimedia Commons, CC BY-SA 4.0

SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact

# What SADiLaR can contribute next




- Co-develop African-language STEM and AI terminology with educators and learners.
- Create classroom-ready multilingual examples using health, water, climate and technology claims.
- Develop resources that support text, speech, images, video and South African Sign Language.
- Pilot a guided learner hackathon with clear educational and ethical outcomes.
- Evaluate whether learners understand, reason, explain and create—not merely whether they accessed AI.

Synthesis of NSTF/proSET objectives and SADiLaR contribution | Photo: JeffNyoka, Wikimedia Commons, CC BY-SA 4.0

# Use African languages to connect foundational literacy, scientific reasoning and the ability to create with AI.

## Thank you|Ndi a Livhuwa|Ke a Leboga

**Dr Seani Rananga**
**seani.rananga@up.ac.za**

SADiLaR contribution to the AI-in-STEM discussion

# References


• Adelani, D.I. et al. (2022). Africa-centric transfer learning for named entity recognition. EMNLP.

• Huang, L. et al. (2025). A survey on hallucination in large language models. ACM Computing Surveys.

• Joshi, P. et al. (2020). The state and fate of linguistic diversity and inclusion in NLP. ACL.

• Kasneci, E. et al. (2023). ChatGPT for good? Opportunities and challenges for education. Learning and Individual Differences.

• Nekoto, W. et al. (2020). Participatory research for low-resourced machine translation. Findings of EMNLP.

• Ngũgĩ wa Thiong'o (1986). Decolonising the Mind. James Currey.
• DBE (2025). Funda Uphumelele National Survey: Policy Summary and national results. education.gov.za.

• Howie, S. et al. (2023). PIRLS 2021 South African Main Report. University of Pretoria.

• DBE (2022). Benchmarking Early Grade Reading Skills in South Africa. education.gov.za.

• NSTF/proSET (2026). Artificial Intelligence in STEM Education: Concept Document and Draft Programme.

• Phoobane, P. & Gudyanga, R. (2026). Mapping Artificial Intelligence Research in STEM Education in South Africa. JCSR, 8(1), 23–43.

• Rananga, S. (2026). Multilingual misinformation detection in low-resource languages in South Africa through LLM fine-tuning. PhD thesis, NWU.
• FintonCassidy. Fiona1on1Sep13. Wikimedia Commons, CC BY-SA 3.0.

• JeffNyoka. Children using tablets at City of Joburg libraries. Wikimedia Commons, CC BY-SA 4.0.

• Peace Corps. Zambian kids learning how to use computers. Wikimedia Commons, public domain.

• Marius140%. Adult education / English class for a Rainbow Nation. Wikimedia Commons, CC BY-SA 4.0.

• Bianca Lawrence53. Researching information for transfer of knowledge. Wikimedia Commons, CC BY-SA 4.0.

• South African Sign Language image. Wikimedia Commons, CC BY 4.0.

• Image licences and source pages checked on Wikimedia Commons, 19 August 2026.


SADiLaR
South African Centre for Digital Language Resources

African Institute for Data Science and Artificial Intelligence (AfriDSAI)

Data Science for Social Impact