UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Make today matter

www.up.ac.za

Africa-centric NLP models nearly double classification accuracy on isiZulu climate narratives — generic English models systematically miss African climate voices.

Mapping Climate Asymmetries in South African Media: Benchmarking Generic and Africa-centric NLP Models on English and isiZulu Climate Discourse

INTRODUCTION

- Climate change impacts Africa disproportionately, yet African voices are underrepresented in global climate policy narratives South African media shapes public climate discourse, but no systematic multilingual analysis exists especially across English and isiZulu NLP and LLMs offer tools to detect asymmetries but are widely-used models fit for African language contexts?

METHODS

1. Corpus: 186 articles (94 English, 92 isiZulu) from 8 SA media sources, 2005–2026
2. 5-theme taxonomy: Climate Perception, Economic Development, Environmental Risk, Energy Transition, Policy Debate
3. Zero-shot classification benchmarking: DistilBERT-MNLI (generic) vs mDeBERTa-XNLI (Africa-centric multilingual).
4. KDD pipeline + PCS framework (Predictability, Computability, Stability); fully local, reproducible.

Mapule Madi, Sisipho Twalo

RESULTS

The Africa-centric model nearly doubles classification accuracy on isiZulu (F1: 0.672 vs. 0.331) while the generic model leads on English (F1: 0.584 vs. 0.327) — a clear cross-over pattern confirming that model choice is not a neutral technical decision, but one that directly determines whose climate voice gets heard and whose gets erased.

	Generic (DistilBERT)	Africa-centric (mDeBERTa)
English F1	0.584	0.327
isiZulu F1	0.331	0.672
Mean F1	0.528	0.470

DISCUSSION

Using only generic English-trained models is a methodological bias, it erases isiZulu perspectives and reproduces the colonial asymmetries this field seeks to address. isiZulu frames climate through community vocabulary (abantu, izikhukhula, amanzi); any policy tool blind to this will under-serve the communities most at risk.The representation gap runs deeper than vocabulary — structural differences in how climate is expressed across languages persist even when bilingual taxonomies are introduced.Optimal solution: a hybrid pipeline routing English documents to the generic model and isiZulu documents to the Africa-centric model.

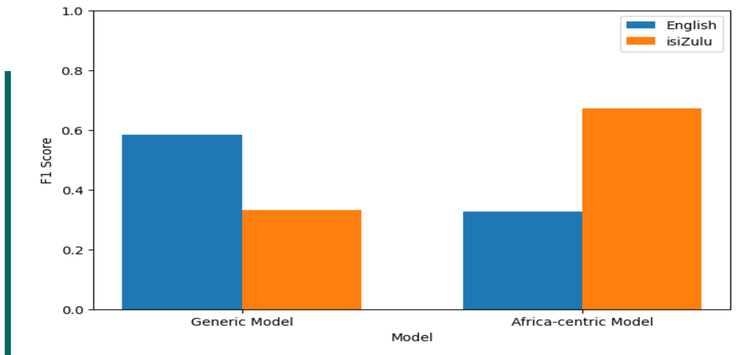

Figure: F1-scores of generic and Africa-centric models across English and isiZulu documents, showing language-dependent differences in classification performance.

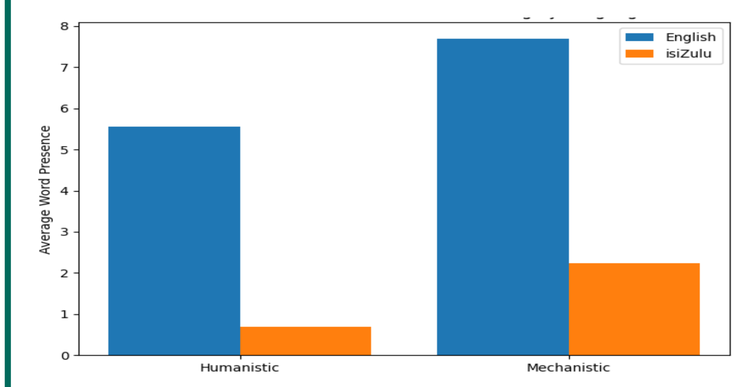

Figure 2 :Humanistic vs. mechanistic climate framing across English and isiZulu discourse.

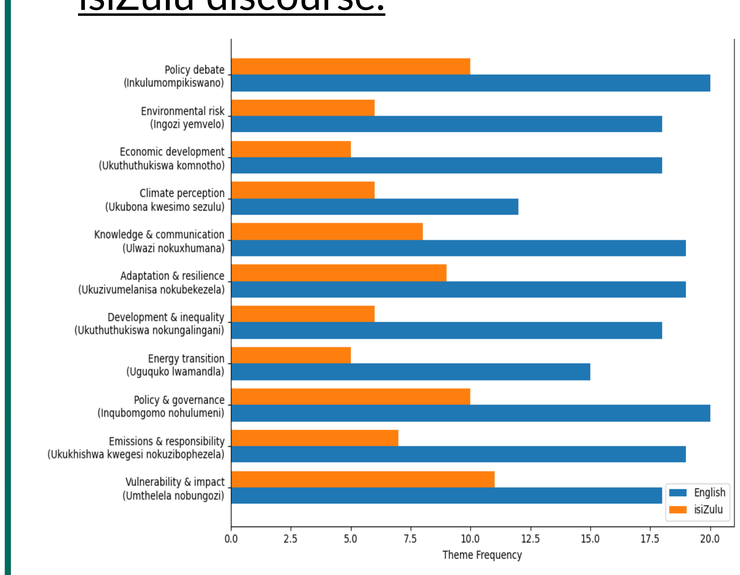

Figure 3: Climate theme coverage compared across English and isiZulu using bilingual thematic labels.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Make today matter

Make today matter

Take a picture to **download** the **full paper**