Text Classification

Sentiment Analysis (f1)

News Topic Classification (acc)

Intent Classification (acc)

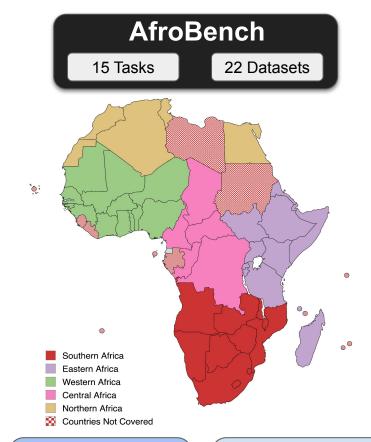
Hate Speech Classification (f1)

Natural Language Inference (acc)

Token Classification

POS Tagging (F1)

NER (F1)



Reasoning

Mathematics (em)

Summarization

Summarization (bertscore)

Question Answering

Cross-Lingual QA (f1)

Reading Comprehension (f1)

Knowledge QA (acc)

MMLU (acc)

Machine Translation

Machine Translation (chrf)

Diacritics Restoration

Diacritics Restoration (chrf)