

Sixth Workshop on African Natural Language Processing (AfricaNLP2025)

Multilingual and Multicultural-aware LLMs

Vienna, Austria
31 July 2025

ISBN: 979-8-3313-2399-8

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Yankari: Monolingual Yoruba Dataset</i>	
Maro Akpobi	1
<i>Supervised Machine Learning based Amharic Text Complexity Classification Using Automatic Annotator Tool</i>	
Gebregziabihier Nigusie	7
<i>On the Tolerance of Repetition Before Performance Degradation in Kiswahili Automatic Speech Recognition</i>	
Kathleen Siminyu, Kathy Reid, ryakitimboruby@gmail.com ryakitimboruby@gmail.com, bmwasaru@gmail.com bmwasaru@gmail.com and chenai@chenai.africa chenai@chenai.africa	15
<i>Enhancing AI-Driven Farming Advisory in Kenya with Efficient RAG Agents via Quantized Fine-Tuned Language Models</i>	
Theophilus Lincoln Owiti and Andrew Kiprop Kipkebut	24
<i>Pretraining Strategies using Monolingual and Parallel Data for Low-Resource Machine Translation</i>	
Idriss Nguépi Nguéack, Mara Finkelstein and Toadoum Sari Sakayo	31
<i>Designing and Contextualising Probes for African Languages</i>	
Wisdom Aduah and Francois Meyer	39
<i>Building a Functional Machine Translation Corpus for Kpelle</i>	
Kweku Andoh Yamoah, Jackson Weako and Emmanuel Dorley	52
<i>Exploring Transliteration-Based Zero-Shot Transfer for Amharic ASR</i>	
Hellina Hailu Nigatu and Hanan Aldarmaki	64
<i>Fine-tuning Whisper Tiny for Swahili ASR: Challenges and Recommendations for Low-Resource Speech Recognition</i>	
Avinash Kumar Sharma, Manas Pandya and Arpit Shukla	74
<i>Who Wrote This? Identifying Machine vs Human-Generated Text in Hausa</i>	
Babangida Sani, Aakansha Soy, Sukairaj Hafiz Imam, Ahmad Mustapha, Lukman Jibril Aliyu, Idris Abdulmumin, Ibrahim Said Ahmad and Shamsuddeen Hassan Muhammad	82
<i>Automatic Speech Recognition for African Low-Resource Languages: Challenges and Future Directions</i>	
Sukairaj Hafiz Imam, Babangida Sani, Dawit Ketema Gete, Bedru Yimam Ahmed, Ibrahim Said Ahmad, Idris Abdulmumin, Seid Muhie Yimam, Muhammad Yahuza Bello and Shamsuddeen Hassan Muhammad	89
<i>SabiYarn: Advancing Low Resource Languages with Multitask NLP Pretraining</i>	
Oduguwa Damilola John, Jeffrey Otoibhi and David Okpare	95
<i>Retrieval-Augmented Generation Meets Local Languages for Improved Drug Information Access and Comprehension.</i>	
Ahmad Ibrahim Ismail, Bashirudeen Opeyemi Ibrahim, Olubayo Adekanmbi and Ife Adebara	108
<i>Story Generation with Large Language Models for African Languages</i>	
Catherine Nana Nyaah Essuman and Jan Buys	115

<i>Command R7B Arabic: a small, enterprise-focused, multilingual, and culturally aware Arabic LLM</i>	
Yazeed Alnumay, Alexandre Barbet, Anna Bialas, William Michael Darling, shaan@cohere.com shaan@cohere.com, joan@cohere.com joan@cohere.com, Kyle Duffy, stephaniehowe@cohere.com ste- phaniehowe@cohere.com, Olivia Lasche, Justin Seonyong Lee, anirudh@cohere.com anirudh@cohere.com and jennifer@cohere.com jennifer@cohere.com	126
<i>Challenges and Limitations in Gathering Resources for Low-Resource Languages: The Case of Me- dumba</i>	
Tatiana Moteu Ngoli, Mbuh Christabel and Njeunga Yopa	136
<i>YodiV3: NLP for Togolese Languages with Eyaa-Tom Dataset and the Lom Metric</i>	
Bakoubolo Essowe Justin, Kodjo François Xegbe, Catherine Nana Nyaah Essuman and Afolá Kossi Mawouéna Samuel	143
<i>Challenging Multimodal LLMs with African Standardized Exams: A Document VQA Evaluation</i>	
Victor Tolulope Olufemi, Oreoluwa Boluwatife Babatunde, Emmanuel Bolarinwa and Kausar Yetunde Moshood	150
<i>MOZ-Smishing: A Benchmark Dataset for Detecting Mobile Money Frauds</i>	
Felermimo D. M. A. Ali, Henrique Lopes Cardoso, Rui Sousa-Silva and saide.saide@unilurio.ac.mz saide.saide@unilurio.ac.mz	158
<i>In-Domain African Languages Translation Using LLMs and Multi-armed Bandits</i>	
Pratik Rakesh Singh, Kritarth Prasad, Mohammadi Zaki and Pankaj Wasnik	167
<i>HausaNLP: Current Status, Challenges and Future Directions for Hausa Natural Language Processing</i>	
Shamsuddeen Hassan Muhammad, Ibrahim Said Ahmad, Idris Abdulmumin, Falalu Ibrahim Lawan, Sukairaj Hafiz Imam, Yusuf Aliyu, Sani Abdullahi Sani, Ali Usman Umar, Tajuddeen Gwa- dabe, Kenneth Church and Vukosi Marivate	176
<i>Beyond Generalization :Evaluating Multilingual LLMs for Yorùbá Animal Health Translation</i>	
Godwin Adegbhingbe, Anthony Soronnadi, Ife Adebara and Olubayo Adekanmbi	192
<i>Evaluating Robustness of LLMs to Typographical Noise in Yorùbá QA</i>	
Paul Okewunmi, Favour James and Oluwadunsin Fajemila	195
<i>Swahili News Classification: Performance, Challenges, and Explainability Across ML, DL, and Trans- formers</i>	
Manas Pandya, Avinash Kumar Sharma and Arpit Shukla	203
<i>Neural Morphological Tagging for Nguni Languages</i>	
Cael Marquard, Simbarashe Mawere and Francois Meyer	210
<i>Multilingual NLP for African Healthcare: Bias, Translation, and Explainability Challenges</i>	
Ugochi Okafor	221
<i>Beyond Metrics: Evaluating LLMs’ Effectiveness in Culturally Nuanced, Low-Resource Real-World Scenarios</i>	
Millicent Ochieng, Varun Gumma, Sunayana Sitaram, Jindong Wang, Vishrav Chaudhary, Keshet Ronen, Kalika Bali and Jacki O’Neill	230
<i>Y-NQ: English-Yorùbá Evaluation dataset for Open-Book Reading Comprehension with Open-Ended Questions</i>	
Marta R. Costa-jussà, Joy Chen, Ife Adebara, Joe Chuang, Christophe Ropers and Eduardo Sán- chez	248