Hailay Kidu Teklehaymanot

Location: Hannover, Germany
Email: teklehaymanot@l3s.de | hailaykidu.teklehaymanot@uzh.ch
Phone: +49 176 45174080
LinkedIn: linkedin.com/in/hailay-kidu-teklehaymanot
ORCID: 0000-0003-2360-3448
Website: SIGIR LiveRAG Challenge

Professional Summary

PhD candidate at **L3S Research Center, Leibniz Universität Hannover** (Supervisors: Prof. Wolfgang Nejdl, Prof. Niloy Ganguly) and Research Associate at **University of Zurich (UZH)**. Over ten years of experience in **AI, multilingual NLP, and low-resource language modeling**, specializing in **morphology- and structure-aware tokenization, retrieval-augmented generation (RAG) systems, and responsible AI**. Winner of the **SIGIR 2025 LiveRAG Challenge** and the **AI Grid Hackathon 2024**, with expertise in integrating linguistic structure into Transformer-based models, designing evaluation frameworks for compositionality, and deploying NLP systems in resource-constrained environments.

Research and Professional Experience

2022–Present **PhD Candidate**, L3S Research Center, Leibniz Universität Hannover, Germany
Multilingual NLP research on low-resource and morphologically rich languages. Developed linguistically aware subword tokenizers, embedding strategies, and RAG pipelines (OpenSearch, Pinecone). Designed compositionality-focused evaluation frameworks and investigated bias, fairness, and infrastructure-aware NLP. Contributed to EU Horizon proposals and interdisciplinary AI projects.
Research Associate, University of Zurich (UZH), Switzerland
Applied NLP methods to historical, scriptural, and digital humanities corpora. Integrated linguistic structure into language models for cross-disciplinary applications, contributing to reproducible and inclusive NLP research pipelines.

2019–2021 **PhD Student**, Beijing University of Technology, China(started)
Conducted research in Computational Linguistics, focusing on morphologically rich and underrepresented languages. Worked on subword tokenization, embedding representation, and cross-lingual language model adaptation.

2009–2019 **Assistant Professor**, Mekelle University, University of Gondar & Adigrat University, Ethiopia
Taught courses in Information Retrieval, NLP, Data Structures and Algorithms, and Information Systems. Supervised MSc theses on Amharic text mining and morphological analysis, established Ethiopia's early NLP research groups, and modernized AI curricula. Mentored students in linguistically informed NLP research.

2007–2009 **Data Scientist / ML Engineer**, Cyber Software, Ethiopia
Developed national-scale decision-support systems for Ethiopia's justice and public administration sectors, including analytics pipelines and reporting tools.

Education

2022–Present **PhD in Computer Science**, Leibniz Universität Hannover, Germany
Thesis: *Advancing LLM Infrastructure for Morphologically Rich Underrepresented Languages*
Honors: SIGIR 2025 LiveRAG; AI Grid Hackathon 2024

2010–2012 **MSc in Information Technology**, University of Madras, India

2004–2007 **B.Ed. in Information Technology**, Mekelle University, Ethiopia

Skills and Methods

AI & NLP	Multilingual NLP, low-resource languages, structure- and morphology-aware tokenization, retrieval-augmented generation (RAG), LLM evaluation, compositionality analysis
Programming & Systems	Python (Transformers, PyTorch), Scikit-learn, OpenSearch, Pinecone, SQL/NoSQL, Linux/HPC
Responsible AI	Bias and fairness analysis, equity-aware NLP, infrastructure-aware modeling, AI ethics and governance
Research Methods	Experimental design, statistical analysis, diagnostic evaluation, morphologically-aware benchmarks, mixed qualitative–quantitative methods
Linguistics & Modeling	Language modeling, subword and embedding representation, cross-lingual adaptation, linguistic theory for morphologically rich languages
Languages	Amharic (native), Tigrinya (native), English (fluent), German (B1)

Selected Publications

- MoVoC: Morphology-Aware Subword Construction for Ge'ez Script Languages (Teklehaymanot et al., Findings 2025)
- TIGQA: An Expert-Annotated Question-Answering Dataset in Tigrinya (Teklehaymanot et al., LREC-COLING 2024)
- Morphological Synthesizer for Ge'ez Language: Addressing Morphological Complexity and Resource Limitations (Gidey et al., RAIL 2024)
- LGSE: Lexically Grounded Subword Embedding Initialization (LREC 2026 Accepted)
- RAGtifier: Evaluating RAG Generation Approaches of State-of-the-Art RAG Systems for the SIGIR LiveRAG Competition (SIGIR 2026 LiveRAG)
- Low-Resource English-Tigrinya MT: Leveraging Multilingual Models, Custom Tokenizers, and Clean Evaluation Benchmarks (FLLM 2025, IEEE Xplore)
- Tokenization Disparities as Infrastructure Bias: How Subword Systems Create Inequities in LLM Access and Efficiency (FLLM 2025, IEEE Xplore)

References

Prof. Wolfgang Nejdl
Main Supervisor, Director L3S Research Center, Leibniz Universität Hannover, Germany
Email: nejdl@L3S.de | Phone: +49 511 762 4660

Prof. Niloy Ganguly
Professor, Department of Computer Science, Indian Institute of Technology, Kharagpur, India
Email: niloy@cse.iitkgp.ac.in