

Temesgen, Tsedeniya Kinfе

Address: | baki1145@gmail.com | +4915510685617 | Karlstr. 31, 74072 Heilbronn, Germany.

☐ Education and training

PhD Linguistic Information in Chatbots, Technical University of Munich (TUM) —

15/10/2024 – Present, Heilbronn, Germany

Research Focus: Linguistic information in LLMs, multilinguality, and low-resource languages

PhD supervisor: Prof. Dr. Alexander M. Fraser

M.Sc in Information Science and Systems (Language Technology) Addis Ababa University

07/09/2016 – 11/07/2019 Addis Ababa, Ethiopia

Thesis title: “Grammar Checker for the Amharic Language Using Probabilistic and Rule-Based Approach”

Advisor: Martha Yifiru Tachbelie (Ph.D.)

B.Sc in Information Science Addis Ababa University

17/09/2012 – 07/07/2016 Addis Ababa, Ethiopia

Course Taken: Human-Computer Interaction (HCI), Multimedia Information System, Introduction to AI, Expert System,

Thesis project: “Motor Underwriting Management System” for the National Insurance Company of Ethiopia (NICE)

Training in R programming language Addis Ababa University

07/05/2018 – 12/05/2018, Addis Ababa, Ethiopia

Certification on Machine Learning University of Lisbon, jointly hosted by Instituto de

Telecomunicações (IT) – 11/07/2019 – 18/07/2019 Lisbon, Portugal

Certification on Deep Learning + Reinforcement Learning(DLRL) – Quebec Artificial Intelligence

Institute – 29/07/2020 – 06/08/2020 Montreal, Canada

☐ Professional Experience

Assistant Researcher(part-time) - Apple

01/03/2023 - 31/10/2023, Copenhagen, Denmark

-
- Study and examine tools(such as embeddings, and Language ID) for low-resource East African languages.
 - Design evaluation datasets that are representative of the language and the community.

Assistant Researcher - Technical University of Denmark

20/11/2023- 31/10/2023, Kongens Lyngby, Denmark

-
- Investigate data collection pipeline for low-resource African languages
 - The efficacy of Large Language Models (LLMs) in handling languages with distinctive scripts
 - Understand the impact of representation bias in pre-training.

Assistant Researcher - Apple

03/04/2023 - 30/09/2023, Paris, France

-
- Assess pre-training data quality and availability for low-resource African languages.
 - Examine the usability of a specific NLP task for these languages.
 - Evaluate the utility of the dataset used for the language in the study.

Lecturer - Hope University College

08/07/2019 – 23/10/2021, Addis Ababa, Ethiopia

Instructor - Department of Information Science

- Given lectures on Fundamentals of Programming(C++), Emerging technology, Multimedia Systems, Decision Support Systems, and Management Information Systems.
- Prepare lectures and class activities for undergraduate students focusing on the current technology (AI, Data Science, Big Data),

Undergraduate Mentor, Industrial Project, Department of Information Science

- Mentor six students on their Industrial Projects

- Guide students on planning, scheduling, and organizing their documentation and projects.

University Service - Hope University College

08/07/2019 – 23/10/2021, Addis Ababa, Ethiopia

Course Coordinator, Exam Committee, and Director of Research Committee in the Department of Information Science

- Organize course materials and arrange schedules for course chapters.
- Plan and lead required training sessions.
- Arrange, organize, and prepare examinations in the required content and format.
- Organize events for students to be exposed to the newest technologies and share research among instructors and students.
- Organize the 3rd national conference held by Hope University College on The impact of COVID-19 on higher education

☐ Research experience

Investigate Low Resource Language Data Quality and Availability for East African Languages

-
- Conduct a comprehensive analysis involving: data quality assessment, data availability review, usability testing for a specific task, and dataset utility evaluation
 - Address challenges in language technology development for underrepresented languages.

Syntax Analysis for URL using a Probabilistic with a Rule-based approach (Master's thesis)

-
- Examine the morpho-syntax nature of the language and apply manually crafted rules with the help of linguists.
 - Build a higher-order language model ($n=3-5$) to handle the correct word order of language.
 - Design, experiment, and communicate the experiment's results via a poster presentation.

Syntax Analysis for URL using a Statistical Approach (Master's thesis)

-
- Build a new labeling technique to tag each word with their appropriate detailed syntax tagger to be trained in higher-order N-gram language models
 - The higher-order language model is evaluated for handling longer-distance agreements.
 - Communicate the results of the experiment via a poster presentation.

Syntax Analysis for URL using a Dependency Parser (Master's thesis)

-
- Collect, and annotate data
 - Examine how well the approach works in checking grammatical agreement within a sentence.
 - The approach was examined for short and long-distance agreement Maltparser, Malt Optimizer, and evaluator
 - Communicate the results of the experiment via a poster presentation.

Assistant Researcher: Speech Processing

-
- Assist in gathering data from different resources, Describe the data collection tool to the group of participants, and Prepare a report on the collected corpus and data collection methods.

Assistant Researcher: Morphological Segmentation

-
- Facilitate and Coordinate data collection methods, Ensure the cleaned data are of quality, and Prepare a report on corpus collection and preprocessing Techniques.

Research on Speech Translation for Under-resourced Language (BlackAIR and Stanford University)

-
- Investigate Unsupervised speech-to-text translation without the need for parallel data. K-means segmental and LSTM algorithms are investigated to translate speech to text at the word level.
 - Communicate the findings on the poster presentation.

☐ Project Experience

Image Processing: Face Detection

-
- Collect and preprocess image data., Conduct and implement face detection algorithms using Matlab. Prepare a project report, and present and Communicate the result of the project through demonstration.

Prediction Model using Probabilistic Context-Free Grammar (PCFG)

-
- Collect and preprocess text corpus, Develop a predictive model using NLTK + Python as a tool, Prepare a report, and present the output of the project.

- Study the morphological behavior of Amharic language, Conduct a morphological analyzer using the Stuttgart Finite State Transducer (SFST) tool for the generation, manipulation, and processing of finite-state automata and transducers, Prepare a project report describing the limitation of Stuttgart Finite State Transducer.

❑ Skills

-
- Machine learning package: Tensorflow, Keras, PyTorch, Matplot, Scikit-Learn
 - Documentation: Mendeley and Latex
 - Programming languages and mathematical packages: Matlab, C ++, Python, Java, SWI-Prolog

❑ Conference participation and poster presentation

-
- NeurIPS 2018: Thirty-second Conference on Neural Information Processing Systems held in Montreal, Canada between 2nd to 8th December.
 - Poster presented at Women's in Machine Learning (WiML) and Black in Artificial Intelligence (BAI) Workshop titled "*Grammar Checker for the Amharic Language Using Probabilistic and Rule-Based Approach*".
 - Indaba X Ethiopia 2019: The 1st workshop was held at WolaitaSodo University, Ethiopia from May 2nd to 4th.
 - Poster presented at Indaba X Ethiopia workshop titled "*Automatic Grammar Correction for Under-Resourced Languages using Hybrid Approach*".
 - ACI 2019: the 1st AI conference was held from October 28 to October 30, at AAIT- Amist Kilo Campus, Addis Ababa, Ethiopia.
 - NeurIPS 2019: The 33rd Neural Information Processing Systems Annual Conference which takes place Dec 8th through 14th, at Vancouver Convention Center, Vancouver, Canada.
 - Joint Poster session presented at Women's in Machine Learning (WiML) and Black in Artificial Intelligent (BAI) Workshop, titled "*Syntax Analysis for the Amharic Language*".
 - Computer Science Summer School Research program (CURIS): Online presentation on August 26, 2021.
 - Poster session presented at CURIS titled "Speech translation for Under-resourced languages".

❑ References

Natalie Schluter(Prof.)
Machine Learning Researcher, Apple
Associate Professor, Department of Applied Mathematics and
Computer Science, Technical University of Denmark
(45)26338373, natschluter@apple.com

Martha Yifiru (Ph.D.), Associate Professor
School of Information Sciences, College of Natural and
Computational Sciences.
Addis Ababa University, Addis Ababa, Ethiopia.
(251) 923518241, martha.yifiru@aaau.edu.et