HELLINA HAILU NIGATU

hellina_nigatu@berkeley.edu \diamond Berkeley, CA

Website

I am interested in improving access and utility of language technologies by (1) centering linguistic characteristics and community collaboration in design and (2) investigating the impact of currently deployed language technologies on marginalized communities in the Global Majority. My research area lies in the intersection of NLP, HCI, and AI Ethics.

EDUCATION

PhD in Computer Science, University of California at Berkeley. Advised by Sarah Chasins and John Canny CGPA: 3.78	Present
MSc in Computer Science, University of California at Berkeley. Advised by Sarah Chasins and John Canny CGPA: 3.73	May 2023
BSc. Electrical and Computer Engineering, Addis Ababa University Advised by Biruk Tilahun and Feven Fisseha CGPA: 3.89	Dec 2020
PUBLICATIONS	
mRAKL: Multilingual Retrieval-Augmented Knowledge Graph Construction for Loguages. Hellina Hailu Nigatu, Min Li, Maartie ter Hoeve, Saloni Potdar, Sarah E, Chasins	ow-Resourced Lan- ACL (Findings) 2025
In Proceedings of Findings of the Association of Computational Linguistics.	
Exploring-Transliteration-Based-Zero-Shot-Transfer-for-Amharic-ASR Hellina Hailu Nigatu, Hanan Aldarmaki In Proceedings of the AfricaNLP Workshop.	AfricaNLP 2025
Cognate Detection for Historical Language Reconstruction of Proto-Sabean Lang Ge'ez, Tigrinya, and Amharic Elleni Sisay Temesgen, Hellina Hailu Nigatu, and Fitsum Assamnew Andargie. In Proceedings of the 31st International Conference on Computational Linguistics.	uages: the Case of COLING 2025
The Zeno's Paradox of Low-Resource Languages. Hellina Hailu Nigatu, Atnafu Lambebo Tonja, Benjamin Rosman, Thamar Solorio, Monojit Ch In Proceedings of the Empirical Methods for Natural Language Processing. –Outstanding Pa	EMNLP 2024 noudhury aper Award
Exploitation All the Way Down: Calling out the Root Cause of Bad Online Exper the Majority World. Hellina Hailu Nigatu, Zeerak Talat. Data and Al Governance Coalition (DAIG) at the 19th Annual Meeting of Internet Governance	e Forum
A Capabilities Approach to Studying Bias and Harm in Language Technologies. Hellina Hailu Nigatu, Zeerak Talat. Extended Abstract accepted to Workshop on New Perspectives on Bias and Discrimination in I	WAI 2024 Language Technology.
Evaluating Gender Bias in Machine Translation for Low-Resource Languages Walelign Tewabe Sewunetie, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Hellina Hailu Nig Zewdie Mossie, Hussien Seid, Eshete Derb, Seid Muhie Yimam Proceedings of the 5th Workshop on African Natural Language Processing	AfricaNLP 2024 atu, Gashaw Kidanu,
"I Searched for a Religious Song in Amharic and Got Sexual Content Instead.": In Harm in Low-Resourced Languages on YouTube. Hellina Hailu Nigatu, Inioluwa Deborah Raji In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency	FAccT 2024

Low-Resourced Languages and Online Knowledge Repositories: A Need Finding Study. CHI 2024 Hellina Hailu Nigatu, John Canny, Sarah Chasins In Proceedings of the CHI Conference on Human Factors in Computing Systems The Less the Merrier? Investigating Language Representation in Multilingual Models **EMNLP 2023** Hellina Hailu Nigatu, Atnafu Lambebo Tonja, Jugal Kalita In Proceedings of Empirical Methods in Natural Language Processing (EMNLP) 2023 Findings Co-Designing for Transparency: Lessons from Building a Document Organization Tool for the Criminal Justice Domain. FAccT 2023 Hellina Hailu Nigatu, Lisa Pickoff-White, John Canny, Sarah Chasins In Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) A Need Finding Study with Low-Resource Language Content Creators AfriCHI 2023 Hellina Hailu Nigatu, John Canny, Sarah Chasins In Proceedings of 4th ACM African Human-Computer Interaction Conference (AfriCHI 2023) Enhancing Translation for Indigenous Languages: Experiments with Multilingual Models. AmericasNLP 2023 Atnafu Lambebo Tonja, Hellina Hailu Nigatu, Olga Kolesnikova, Grigori Sidorov, Alexander Gelbukh, Jugal Kalita In AmericasNLP 2023 Shared Task on Machine Translation into Indigenous Languages DyGraph: A Dynamic Graph Generator and Benchmark Suite GRADES-NDA 2022 Andrew McCrabb, Hellina Nigatu, Absalat Getachew and Valeria Bertacco In Proceedings of ACM Workshop for Graph Data Management Experiences and Systems and Network Data Analytics The African Stopwords Project: Curating Stopwords for African Languages AfricaNLP 2022 Chris Chinenye Emezue, Hellina Hailu Nigatu, Cynthia Thinwa, Lerato Louis, Idris Abdulmumin, Samuel Gbenga Oyerinde, Benjamin Avoade Ajibade, Helper Zhou, Emeka Felix Onwuegbuzia, Handel Chiagozie Emezue, Ifeoluwatavo Adeseye Ige, Atnafu Lambebo Tonja, Chiamaka Ijeoma Chukwuneke, Shamsuddeen Hassan Muhammad, Olanrewaju Samuel.

In Proceedings of ICLR Workshop for African Natural Language Processing.

PROJECTS

Multilingual Knowledge Graph Completion

- Collaboration with Apple MLR and Product team.
- Focusing on Amharic, Tigrinya, and Arabic.

Improving Automatic Speech Recognition (ASR) for Amharic.

- Based on findings from need-finding work, working on improving ASR for Amharic.
- Investigating transfer learning approaches and impact of transfer language selection in system improvement.

Understanding Multilingual Language Models

- Conducted experiments to understand the differences in embedding space representations across languages in Multilingual Language Models.
- Compared community-based (models built for specific language families or geographic locations and by communities who speak those languages) to generic models (models that include several languages with no clear relation).
- Compared representation among language families, writing scripts, and dialects.
- Investigated how different language representations relate to performance in models being able to distinguish between languages downstream.

Auditing YouTube in Low-Resourced Languages

• Auditing YouTube for online harm in Low-Resourced Languages, particularly Amharic.

Paper

• Used HCI methods for quantitative and qualitative analysis of data.

Need-finding Study of Low-Resourced Language Content Creators

- Conducted two separate studies to understand the challenges—technological and societal—that prevent lowresourced language content creators interacting with online knowledge repositories.
- Used Contextual Inquiry and qualitative content analysis for the study.

AI Augmented Document Organization Tool (Masters Thesis)

- Building a tool for social scientists to organize and sort through large amounts of messy data based on specific entities within the data.
- Pooling techniques from NLP, ComVis, image processing, and HCI.

DyGraph: A Dynamic Graph Generator

- Implemented a dynamic graph generator in C++ that outputs graphs that mimic properties of real-life dynamic graphs such as social networks.
- Focused on capturing the power law nature of degree distributions observed in the naturally existing dynamic graphs.

Obfuscating Electric Circuits by Replicating Logic Gates

- Implemented a tool to replicate sets of connected gates by taking in a Verilog description of a hardware and building a data structure of the netlist in the description.
- Objective of the tool was to protect logical circuits from hardware trojans by adding obfuscation using replicated gates and crossbars.

Analyzing CNN news transcripts to identify frequency of interruptions

- A mini-project that was mainly to serve as an NLP learning experience.
- Used NLP techniques like Sentiment Analysis and Topic Modeling to analyze the characteristics of each of the selected anchors and shows.

HealthCare Services Repository (Undergrad Thesis)

- This project arose as the need to mitigate the challenges of access to healthcare facilities in Ethiopia.
- A Google Maps integrated mobile application that allows patients to search for healthcare services on their phones and locate the nearest facilities that provide those services.
- A web application where facilities enter the products and services they have so patients can access them.
- The metadata for the project is generated using a dynamic graph generator I built for my research project in order to match real-life demands of patients.

AWARDS AND HONORS

- Wikimedia Foundation Research of the Year Award 2025
- Outstanding Paper Award, EMNLP 2024
- Fineunning Meta Models on African Languages Hackaton(3rd place), Deep Learning Indaba 2024
- Best Poster Award, Deep Learning Indaba 2024
- Data Science for Health Ideathon (2nd place), Deep Learning Indaba 2024
- FAccT DEI Scholar, ACM FAccT
- ACM-W Scholarship for Attendance of Research Conferences program, ACM
- SIGHPC Computational and Data Science Fellowship, SIGHPC
- BAIR Research Ignition Award, UC Berkeley, Berkeley Artificial Intelligence Research Lab
- EECS Excellence Award, UC Berkeley, Department of Electrical Engineering and Computer Sciences

Poster

GitHub Paper

GitHub Poster

GitHub

GitHub

- Award for Highest Cumulative GPA, Women 1st Place, Addis Ababa Institute of Technology
- Award for Highest Cumulative GPA, Women 1st Place, Addis Ababa Institute of Technology, Department of Electrical and Computer Engineering
- BSc with Very Great Distinction, Addis Ababa Institute of Technology, Department of Electrical and Computer Engineering

EXPERIENCE

Research Assistant Mohammed Bin Zaved University of Artificial Intelligence

• Working as a summer intern on Machine Translation in Health Care in Prof.

Data Scientist

Opian Information Communication Technology

- Served as a team lead for the Data Science division of the company.
- Led projects related to utilizing police report data to assess reporting climate in relation to Gender Based Violence cases and analyzing tourist review data to extract insights about Ethiopia's tourism sector.
- Continually conduct requirement analysis for other projects in the company and suggest improvements and fixes for the existing tools.

Research Assistant

University of Michigan

• As an extension of prior working engagement with the University, I improved and added features to projects I had implemented during my position as an undergraduate research assistant in Summer 2019.

Data Scientist

Opian Information Communication Technology

- Initially worked as a researcher and student mentee under the student mentorship program. Was promoted to full time data science researcher and analyst.
- Undertook research for projects related to Machine Learning for Healthcare; specifically, carried out tasks to improve pre-existing prediction tools of the company.
- After proposing a new project, which was approved by my supervisor, I worked on adopting a machine learning prediction tool for Intensive Care Units (ICUs) to the Ethiopian context.

Undergrad Research Assistant	May 2019 - Aug 2019
University of Michigan	Ann Arbor, MI
• African Undergraduate Research Adventure (AURA) participant:	Website

- Executed research for projects related to graph computing and hardware security.
- Developed tools in support of the research projects.

SERVICE

COLING Reviewer	2024.
SUMEval organizer	2024.
CHI Reviewer	2024.
LREC Reviewer	2024.
FAccT Program Committee	2024.
WiNLP Organizing Committee	Since 2023

June 2024 - August 2024 Abu Dhabi, UAE

> Jan 2021 - May 2021 Addis Ababa, Ethiopia

> May 2020 - Oct 2020 Addis Ababa, Ethiopia

Aug 2019 - April 2020

Addis Ababa, Ethiopia (Remote role)

Since 2023
Since 2023
2023.
Since 2021
2021-2024
2022-2023
2021-2022
2021
2022
2022

TALKS AND PRESENTATIONS

Low-Resourced Languages and Online Knowledge Repositories: A Need Finding Study	7. Nov 2024.
Speaker for Wikimedia Research Showcase.	Virtual
Practical State of Language Technologies for Low-Resourced Languages	Oct 2024.
Spoke at the PhD Seminar at Cornell University	Virtual
Starting with the Need: How we can use HCI to inform technological design for low-	resourced lan-
guages.	Apr 2024.
Speaker for IoTDayWomen2024.	Virtual
DOT: Reflections from Building a Document Organisation Tool in the Criminal Justic 2023	e Domain Nov
Invited Anchor Speaker for Dathathon4Justice	Virtual
Challenges and Prospects: Building a Document Organisation Tool for Processing Poli	ce Misconduct
Data	Nov 2023
Invited Anchor Speaker for Ingram Olkin Forum on Police Use of Force	NorthField, MN
Auditing YouTube Recommendation in Low-Resourced Languages	July, 2023
The 7th Workshop on Online Abuse and Harms	Virtual
Human-Centered and Ethical NLP for Low-Resourced Languages.	July, 2022
Invited talk for Learning Session at WhoseKnowledge?	Virtual
Build Your Case: Using AI and HCI to aid Document Organization	April 2022
Invited talk at Stanford for Watchdog Reporting Class	Stanford, CA
Detecting and Exploring Replicated Files in Large Document Dumps	April 2022
A demonstration workshop for investigative journalists on using a document organizing tool I am building.	Stanford, CA