

# HELLINA HAILU NIGATU

hellina\_nigatu@berkeley.edu ◊ 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.** Present  
Advised by [Sarah Chasins](#) and [John Canny](#) CGPA: 3.78

**MSc in Computer Science, University of California at Berkeley.** May 2023  
Advised by [Sarah Chasins](#) and [John Canny](#) CGPA: 3.73

**BSc. Electrical and Computer Engineering, Addis Ababa University** Dec 2020  
Advised by [Biruk Tilahun](#) and [Feven Fisseha](#) CGPA: 3.89

## PUBLICATIONS

---

**The Zeno’s Paradox of Low-Resource Languages.** EMNLP 2024  
Hellina Hailu Nigatu, Atnafu Lambebo Tonja, Benjamin Rosman, Thamar Solorio, Monojit Choudhury  
In Proceedings of the Empirical Methods for Natural Language Processing. –**Outstanding Paper Award**

**Exploitation All the Way Down: Calling out the Root Cause of Bad Online Experiences for Users of the Majority World.** DC-DAIG 2024  
Hellina Hailu Nigatu, Zeerak Talat.  
Data and AI Governance Coalition (DAIG) at the 19th Annual Meeting of Internet Governance Forum.

**A Capabilities Approach to Studying Bias and Harm in Language Technologies.** WAI 2024  
Hellina Hailu Nigatu, Zeerak Talat.  
Extended Abstract accepted to Workshop on New Perspectives on Bias and Discrimination in Language Technology.

**Evaluating Gender Bias in Machine Translation for Low-Resource Languages** AfricaNLP 2024  
Walegn Tewabe Sewunetie, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Hellina Hailu Nigatu, Gashaw Kidanu, Zewdie Mossie, Hussien Seid, Eshete Derb, Seid Muhie Yimam  
Proceedings of the 5th Workshop on African Natural Language Processing

**“I Searched for a Religious Song in Amharic and Got Sexual Content Instead.”: Investigating Online Harm in Low-Resourced Languages on YouTube.** FAccT 2024  
Hellina Hailu Nigatu, Inioluwa Deborah Raji  
In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency

**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 Oy-erinde, Benjamin Ayoade Ajibade, Helper Zhou, Emeka Felix Onwuegbuzia, Handel Chiagozie Emezue, Ifeoluwatayo Adeseye Ige, Atnafu Lambebo Tonja, Chiamaka Ijeoma Chukwuneke, Shamsuddeen Hassan Muhammad, Olanrewaju Samuel.

In Proceedings of ICLR Workshop for African Natural Language Processing.

## **PROJECTS**

---

### **A Taxonomy of Errors for Machine Translation in Healthcare for Amharic and Tigrinya**

- Using linguistics and clinical research, build a taxonomy of common errors in translating between Amharic, Tigrinya, and English.
- Identifying common error types that lead to clinical risk by collaborating with clinicians.
- Investigating ways to use taxonomy for improving machine translation in healthcare.

### **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**

[Paper](#)

- 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.
- 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 low-resourced 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)

[GitHub Paper](#)

- 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

[GitHub Poster](#)

- 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

[Poster](#)

- 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

[GitHub](#)

- 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)

[GitHub](#)

- 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

---

- **Outstanding Paper Award**, EMNLP 2024
- **Fineuning 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*
- **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** June 2024 - August 2024  
Mohammed Bin Zayed University of Artificial Intelligence *Abu Dhabi, UAE*

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

**Data Scientist** Jan 2021 - May 2021  
Opian Information Communication Technology *Addis Ababa, Ethiopia*

- 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** Aug 2019 - April 2020  
University of Michigan *Addis Ababa, Ethiopia (Remote role)*

- 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** May 2020 - Oct 2020  
Opian Information Communication Technology *Addis Ababa, Ethiopia*

- 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

**Berkeley EECS PEERS, Peer support provider** Since 2023

**Berkeley STEM\*FYI, ambassador** Since 2023

**FAccT Volunteer Chair** 2023.

**Black Graduate Engineering and Science Students (BGESS), E-Board Member** Since 2021

**BAIR ReU, Committee Member** 2021-2024

Berkeley NLP Seminar, Organizer	2022-2023
JUAMI-Open Compute Facility, Communications Lead	2021-2022
NeurIPS ML4D, Reviewer	2021
Deep Learning Indaba, Reviewer	2022
HCMaT at ICML, PC member	2022

## TALKS AND PRESENTATIONS

---

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