

# CHIBUZOR OKOCHA JOSEPH

Gainesville, Florida   ◇   +1 352-888-9432   ◇   Email   ◇   GitHub   ◇   Linkedin   ◇   Portfolio

## SUMMARY

---

My research focuses on using Natural language processing, machine translation, and Automatic speech recognition for low-resource languages. I have also researched ASR and accented speech for low-resource languages, especially African languages. I have experience in uni-modal and multi-modal large language models

## EDUCATION

---

**University of Florida, Gainesville, Florida**

*August 2023 - Expected summer 2025*

Masters in Engineering Education- concentration in AI

**University of Benin, Benin, Edo**

*Sept 2015 - Nov 2021*

BEng Computer Engineering

## PUBLICATION

---

### Papers

- Mardhiyah Sanni, Tassallah Abdullahi, Devendra Deepak Kayande, Emmanuel Ayodele, Naome A Etori, **Chibuzor Okocha**, Lukman Enegi Ismaila, Tobi Olatunji; Probing Multilingual and Accent robustness of Speech LLMs. Anonymous submission to Interspeech 2025 ([Interspeech 2025](#)).
- Mardhiyah Sanni, Tassallah Abdullahi, Devendra Deepak Kayande, Emmanuel Ayodele, Naome A Etori, Michael Samwel Mollel, Moshood O. Yekini, **Chibuzor Okocha**, Lukman Enegi Ismaila, FolaFunmi Omofoye, Boluwatife A. Adewale, Tobi Olatunji; Afrispeech-Dialog: A Benchmark Dataset for Spontaneous English Conversations in Healthcare and Beyond. 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics ([NAACL 2025 accepted](#)).
- **Chibuzor Okocha**, Mariko Adachi, Jeremiah J Blanchard, and Gloria J Kim (2024, October), Navigating Unfamiliar Waters: Enhancing Intercultural Understanding and Academic Self-Efficacy Among American Computer Science Students in Japan at 2024 IEEE Frontiers in Education Conference [[FIE'24](#)].
- **Chibuzor Okocha**, & Kim, G. J., & Choi, J. W., & Yoon, Y. K. (2024, June), Microelectronics Research and Global Competencies: Unpacking Research Abroad Experiences of Engineering Students Paper presented at 2024 ASEE Annual Conference & Exposition, Portland, Oregon. [[ASEE'24](#)].

### Workshops

- Tobi Olatunji, Chinemelu Aka, Tassallah Abdullahi, **Chibuzor Okocha**, Gloria A. Katuka, Naome A. Etori, Lukman E. Ismaila We lost 2 decades: a large-scale multi-country study on healthcare documentation speeds in LMICs. HCI and Health Workshop [[ACM CHI 2025](#)].
- **Chibuzor Okocha** Ethical Challenges Facing Data-Driven Policing: A Review, poster presented at the 2022 Black in AI Affinity workshop at NeurIPS [[NeurIPS 2022](#)].

## RESEARCH EXPERIENCE

---

**Researcher assistant, Bio-RAMP labs.**

*August 2023 - Present*

**Masakhane NLP**

- Collected and curated a comprehensive dataset of African-accented conversations, focusing on medical and non-medical contexts.
- Analyzed the accuracy and reliability of ASR outputs in understanding and transcribing African-accented speech, particularly in medical contexts.

- Worked on open and closed diarization models for accented medical data, focusing on identifying speaker turns and evaluating the performance of these models.
- Evaluated diarization models, assessing their effectiveness and reliability in distinguishing between multiple speakers in African-accented medical conversations.
- Synthesized research findings into a comprehensive report, highlighting key insights and implications for developing speech recognition and language models for African-accented conversations. The findings were submitted to NAACL 2024.

#### **Researcher Intern, Intron Health**

*August 2024 - August 2025*

- Exploring multimodal “voice-native” large language models (LLMs) for low resource languages and african accented speech.
- Researched different SOTA Multimodal LLms
- Organized and managed data collection workflows, ensuring timely submission and processing of diverse African-accented speech data from multiple contributors while maintaining high data quality standards through effective coordination and automated quality checks.
- Fine-tuned state-of-the-art TTS models, such as VITS and XTTS, using Intron custom dataset and available public, to improve synthesized African-accented voices’ naturalness, intelligibility, and accentedness.
- Contributed to the writing and presenting findings at Various venues like Interspeech and ACM CHI.

#### **Graduate Research Assistant, University Of Florida**

*June 2023 - May 2025*

##### **Department of Engineering Education**

- Successfully collected and analyzed qualitative data from diverse participant groups, contributing to a deeper understanding of the IRES IRIKA program’s impact, a global engineering initiative funded by NSF.
- Presented research findings at international conferences (ASEE), raising awareness of the project’s significance and outcomes.
- Developed research and interview protocols and instruments tailored to the project’s objectives and conducted qualitative interviews to achieve the research goals.
- Collaborated with local partners in South Korea to ensure smooth project execution and culturally sensitive data collection practices.

##### **Projects**

- NSF IRES: Interdisciplinary Research in Korea on Applied Smart Systems (IRiKA)
- NSF IRSC: Interdisciplinary Research in Smart City
- University of Florida at Kyoto University

#### **NLP Researcher, Part-time, Masakhane NLP**

*August 2023 - Present*

Masakhane is a grassroots organization that aims to strengthen and spur NLP research in African languages for Africans. Although 2000 of the world’s languages are African, African languages are barely represented in technology.

- Collaborated with diverse teams to develop innovative solutions for language processing.
- Implemented cutting-edge techniques to bridge the gap in African language technology.

## **PROJECTS**

---

- **Afrispeech Dialogue** - GitHub - Property of Intron
- **AfriSpeech-TTS** - GitHub - Property of Intron
- **Multi-Agent Systems** - **GitHub**
- **Real-Time Feedback System for Student Presentations** - GitHub
- **ASR-Fine-Tuning-with-Nvidia-NeMo** - GitHub
- **Building-RAG-Agents-with-LLms** - GitHub
- **Diarization with Nigerian Accented Speech** - GitHub
- **AfriSpeech-200** - GitHub - Property of Intron

## TEACHING EXPERIENCE

---

### **Instructor, Visiola Foundation (2021 - 2022)**

- Led a team of staff going to High schools to teach STEM to high school girls in public schools. It was called The after-school STEM Clubs for Girls ASSCG.
- I was responsible for creating the curriculum for the STEM club, Boot camp, and summer camp.
- supported 50+ girls for coding boot camps across Africa through planning and organizing.
- Designed the curriculum for the stem summer camp and the after-school stem club for girls.
- Taught robotic education and Virtual reality to 50+ teenage girls in a 1 week summer camp.

## LEADERSHIP EXPERIENCE

---

### **Engineering Graduate Student Council, UF - Vice president (2023 - present)**

- Acted as a liaison between the graduate student body and the engineering department's administration, advocating for student needs and interests.
- Implemented initiatives and programs to enhance graduate students' mental, physical, and emotional well-being within the engineering department.
- Coordinated and planned the annual spring visit for incoming PhD students, ensuring a smooth and welcoming experience.
- Served as a point of contact for graduate students to voice their concerns, providing support and directing them to appropriate resources when needed.

### **Graduate Student Council, UF Department representative (2023 - present)**

- Assisted committee chairs in planning and executing campus-wide events such as annual graduate symposiums and research workshops.
- Represented the Department of Engineering Education at the weekly general body meeting of the graduate student council

### **Policy Advocacy in Science and Engineering PASE UF (2023 - present)**

- Planned, organized, and executed educational events such as invited speaker series, panel discussions, and workshops on science communication.
- Coordinated logistics, including venue selection, scheduling, speaker coordination, and event promotion.

## SKILLS

---

- **Research Skills:** Qualitative research, Designing interview protocols, data analysis, quantitative research, interviewing.
- **Technical Skills:** Proficient in Python, JavaScript, HTML, CSS, and SQL.
- **IDEs:** Visual Studio Code, Spyder, IntelliJ, Netbeans, Eclipse, Atom, R studio, PyCharm
- **Visualizations:** Tableau, Microsoft PowerBI, Visio, LucidChart
- **Tools:** Google Colab, Jupyter Notebook, JMP, SPSS, GitHub.
- **Framework:** Pytorch, Tensorflow, Keras, Jax, Scikit learn
- **Specializations:** Prompt Engineering, Requirements Engineering, Project Management, Process documentation, Apps/ tools support, Presentation skills

## ADVANCED COURSES

---

Fundamentals of Machine learning & Quantitative and Qualitative Research methods & Deep Learning,