

CHIBUZOR OKOCHA JOSEPH

+1 352-888-9432 ◇ C.okocha@ufl.edu ◇ GitHub ◇ Linkedin ◇ Portfolio

Summary

Researcher with a background in Speech and Audio AI research. Experienced in building speech-native and multilingual AI applications with large language models (LLMs), doing research in speech LLMs and multimodal language models.

EDUCATION

PhD in computer science and Engineering Education

August 2023 - Present

University of Florida, Gainesville, Florida

Advisor - Dr Christan Grant

Bachelor of Engineering Computer Engineering

Sept 2015 - Nov 2021

University of Benin, Benin, Edo

RESEARCH EXPERIENCE

Graduate Researcher

January 2025 - Present

UF Data Studio

- Prototyped and evaluated speaker diarization systems on African-accented datasets, improving reliability of speech models in diverse contexts; co-authored Domain-Aware Speaker Diarization on African-Accented English (ICASSP 2025).
- Designed and scaled benchmarking pipelines to assess multilingual and accent robustness in speech foundation models, contributing to the development of the AfriVox benchmark (EACL under review).
- Led reproducibility audits and cross-task evaluation of large audio language models (ALMs) across reasoning, transcription, translation, and educational tasks; fine-tuning and pre-training audio models AfriVox and working on other publications (ASRU 2025).

Research Intern

Aug 2024 - Dec 2024

Bio-RAMP Labs

- Developed and optimized ASR and diarization pipelines for African-accented and low-resource speech, improving turn-taking accuracy in real-time applications.
- Benchmarked model performance across architectures using DER/WER, guiding integration choices for safe and effective Claude-style deployments.
- Authored technical reports and co-developed peer-reviewed submissions, advancing scalable speech LLM systems for healthcare in low resource setting (NAACL 2025).

Research Intern, Intron Voice AI (Remote)

Summer 2023

- Designed and evaluated multilingual ASR and speaker diarization models for African-accented English.
- Contributed to the development of **AfriSpeech-Dialog**, a benchmark of 50 simulated African-accented conversations.
- Benchmarked Whisper, XLS-R, and Conformer models, identifying accent-based performance gaps.
- Collaborated with global research teams to optimize robustness and documentation for open-source release.

NLP Researcher, Part-time, Masakhane NLP

August 2022 - 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.

PUBLICATION

Conference Papers

- *Can Large Audio Language Model understand Child Stuttering Speech.* **Chibuzor Okocha**, Maya Bakri, and Christan Grant. (ICASSP 2025 under review)
- *Domain-Aware Speaker Diarization on African-Accented English.* **Chibuzor Okocha**, Kelechi Ezema, Christan Grant (2025). *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP under review)*. [Link]
- *Afrispeech-Dialog: A Benchmark Dataset for Spontaneous English Conversations in Healthcare and Beyond.* Mardhiyah Sanni, Tassallah Abdullahi, Devendra D. Kayande, Emmanuel Ayodele, Naome Etori, Michael Mollel, Moshood Yekini, **Chibuzor Okocha**, et al. (2025). *NAACL*. [Link]
- *AfriVox: Probing Multilingual and Accent Robustness of Speech LLMs.* Tobi Olatunji, Mardhiyah Sanni, **Chibuzor Okocha**, et al. (2025). *EACL*. [Link]
- *Navigating Unfamiliar Waters: Enhancing Intercultural Understanding.* **Chibuzor Okocha**, Mariko Adachi, Jeremiah J. Blanchard, Gloria J. Kim (2024). *IEEE FIE*. [Link]
- *Microelectronics Research and Global Competencies.* **Chibuzor Okocha**, Gloria J. Kim, Jiwon W. Choi, Yong-Kyu Yoon (2024). *ASEE Annual Conference & Exposition*. [Link]
- *WIP: Navigating Reintegration Challenges After a Research Abroad Experience.* **Chibuzor Joseph Okocha**, Gloria J. Kim, Seri Park, Young-Jae Lee (2025). *IEEE FIE, Nashville*. [Link]

Journal Publications

- *A Multi-Country Study Comparing Typed to Automatic Speech Recognition-Based Medical Documentation Speeds.* Tobi Olatunji, Chinemelu Aka, **Chibuzor Okocha**, et al. (2025). [Link]
- *Probing the Surgical Competence of LLMs: A Global Health Study Leveraging AfriMedQA Benchmarks.* Tobi Olatunji, Folafunmi Omofoye, **Chibuzor Okocha**, et al. (2025). [Link]
- *AI-Enabled Frontline Workers: A Multilingual Multimodal Point-of-Care LLM Evaluation Study in Nigeria.* Tobi Olatunji, Chinemelu Aka, **Chibuzor Okocha**, et al. (2025). [Link]

Workshop Papers and Poster Presentations

- *Ethical Challenges Facing Data-Driven Policing: A Review.* **Chibuzor Okocha** (2022). Poster presented at the *Black in AI Affinity Workshop at NeurIPS*. [Link]

SKILLS

- **Technical Skills:** Proficient in Python, Finetuning and Inferencing LLMs, RAG, Agent systems, Typescript, Next.js, JavaScript, HTML, CSS, SQL.
- **Research skills:** AI research, qualitative research, design of interview protocols, data analysis, quantitative research, Research writing.
- **Visualizations:** Tableau, Microsoft PowerBI, Visio, LucidChart
- **Tools:** Google Colab, HuggingFace, JMP, SPSS, GitHub, Vercel, Streamlit.
- **Framework:** Pytorch, torchlightning, Tensorflow, Keras, Scikit learn

PROJECTS

CodecEval-Africa - GitHub A comprehensive benchmark for neural speech codecs on African-accented English and conversational audio.

Afrispeech Dialogue - GitHub - Property of Intron

- Developed AfriSpeech-Dialog, a benchmark dataset of 50 simulated long-form African-accented conversations—including code-switching—for evaluating ASR, speaker diarization, and multi-agent summarization in both medical and non-medical contexts.
- Benchmarked state-of-the-art multilingual ASR models (e.g., Whisper, Conformer, MMS, XLS-R) on African-accented speech, identifying key limitations and performance gaps relative to datasets from other regions.

Afrivox - GitHub

- Probing Unimodal ASR vs Multimodal LLMs for Multilingual Translation and African-accented English Transcription in healthcare and beyond.

ASR-Fine-Tuning-with-Nvidia-NeMo - GitHub

- Built and deployed an ASR system optimized for Nigerian English using NVIDIA's Riva and NeMo frameworks, addressing under-representation in existing speech technologies.

RAG Application with NVIDIA NIM + Streamlit - GitHub

- Built a Retrieval-Augmented Generation (RAG) pipeline using NVIDIA's NeMo Inference Microservices (NIM) and LangChain, enabling efficient document understanding and question answering.

Dialect Classification - GitHub

- Evaluated various open and closed speaker diarization models on Nigerian-accented speech to assess their effectiveness in distinguishing speaker turns in multi-speaker conversations.

LEADERSHIP EXPERIENCE

Human-Centered Data Science - Student Chair (Jan 2025 - Present)

The University of Florida Human-Centered Data Science group is focused on investigating novel applications that cross the boundaries of Databases, Machine Learning, Bioinformatics, and Spatial-Temporal Data Science.

- My duties as the student chair include getting speakers for our weekly seminar series.
- Helping to organize other HCDS events, such as hackathons and workshops.
- keeping an updated list of all the members of the groups and sending out reminders of meetings.
- Updating and maintaining the website.

Engineering Graduate Student Council, UF - Vice president (2023 - 2024)

- 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