

CHIBUZOR OKOCHA JOSEPH

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

EDUCATION

Masters in Engineering Education- (Concentration in AI)

August 2023 - August 2025

University of Florida, Gainesville, Florida

Bachelor of Engineering Computer Engineering

Sept 2015 - Nov 2021

University of Benin, Benin, Edo

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, Jupyter Notebook, JMP, SPSS, GitHub, Vercel, Streamlit.
- **Framework:** Pytorch, torchlightning, Tensorflow, Keras, Jax, Scikit learn
- **Specializations:** Prompt Engineering, Project Management, Process documentation, Apps/tools support, Presentation skills

WORK EXPERIENCE

Graduate Researcher

January 2025 - Present

UF Data Studio

- Spearheaded research and benchmarking of speaker diarization models on African-accented datasets; co-authored paper A Comparative Analysis of Speaker Diarization Models [ASRU 2025, in review]
- Designed and executed large-scale evaluation pipelines for accent and language robustness in speech foundation models; co-authored Evaluating the Multilingual and Accent Robustness of Speech Language Models [ASRU 2025, in review].
- Led reproducibility audits and cross-evaluation experiments to benchmark zero-shot and few-shot generalization of Audio Language Models (ALMs) across reasoning, translation, transcription, and other audio understanding tasks; included fine-tuning and evaluation of models such as Qwen-2 Audio and SALMONN.

Research Intern

August 2023 - Present

Bio-RAMP labs, Masakhane NLP

- Curated and maintained a comprehensive dataset of African-accented conversations across medical and non-medical contexts to enhance speech and language model performance.
- Analyzed ASR outputs to assess accuracy and reliability in transcribing African-accented speech, with a focus on improving performance in clinical and health-related conversations.
- Developed and fine-tuned diarization models (open and closed) to accurately identify speaker turns in accented medical data, optimizing for speaker attribution and segmentation quality.
- Evaluate diarization and ASR model performance using metrics like DER and WER, iterating on model architectures and embeddings to increase precision and robustness.
- Synthesize research findings into internal and external reports, including conference submissions (e.g., NAACL 2024), to inform development of inclusive speech and language technologies for African users.

Graduate Research Assistant,

June 2023 - Dec 2024

University Of Florida

- Collected and analyzed qualitative data from diverse participant groups to evaluate the impact of the NSF-funded IRES IRIKA program, contributing to a nuanced understanding of global engineering education outcomes.

- Presented key research findings at international conferences such as ASEE, increasing visibility and scholarly engagement with the program's cross-cultural and educational significance.
- Designed and implemented research protocols and interview instruments aligned with project objectives, enabling effective data collection and thematic analysis.
- Conducted in-depth interviews with program participants to capture rich narratives and insights related to cross-cultural learning and professional development.

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 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.

PROJECTS

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.

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.

Diarization with Nigerian Accented Speech - 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.

Speechbrain-LLaMA3-Story-Writer - GitHub

- Developed an end-to-end pipeline that converts voice prompts into creative short stories by integrating SpeechBrain for speech-to-text and LLaMA 3 for text generation.

PUBLICATION

Conference Papers

- *Afrispeech-Dialog: A Benchmark Dataset for Spontaneous English Conversations in Healthcare and Beyond*. 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 (2025). *Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics*. [Link]
- *Navigating Unfamiliar Waters: Enhancing Intercultural Understanding and Academic Self-Efficacy Among American Computer Science Students in Japan*. **Chibuzor Okocha**, Mariko Adachi, Jeremiah J. Blanchard, Gloria J. Kim (2024, October). *IEEE Frontiers in Education Conference*. [Link]

- *Microelectronics Research and Global Competencies: Unpacking Research Abroad Experiences of Engineering Students*. **Chibuzor Okocha**, Gloria J. Kim, Jiwon W. Choi, Yong-Kyu Yoon (2024, June). *ASEE Annual Conference & Exposition, Portland, Oregon*. [Link]

Journal Publications

- *A Multi-Country Study Comparing Typed to Automatic Speech Recognition-Based Medical Documentation Speeds Among Low-and Middle-Income Country Trained Clinicians*. Tobi Olatunji, Chinemelu Aka, **Chibuzor Okocha**, Gloria Katuka, Amina Tassallah, Naome Etori, Lukman Ismaila, Bilal A. Mateen, Rebecca Weintraub (2025). [Link]
- *Probing the Surgical Competence of LLMs: A Global Health Study Leveraging AfriMedQA Benchmarks*. Tobi Olatunji, Folafunmi Omofoye, Ezinwanne C. Aka, Gina Itzikowitz, Daniel Macaulay, Oyinoluwa G. Adaramola, Boluwatife A. Adewaledeleye, Chidi Asuzu, Simisola Popoola, Wendy Kinara, Emmanuel Ayodele, Ifeoluwa Yinusa, Oluwatoni Adekunle, Mardhiyah Sanni, **Chibuzor Okocha**, Tassallah Abdullahi, Abraham Owoduni, Charles Nimo, Mercy N. Asiedu, Bilal Mateen, Rebecca Weintraub (2025). [Link]
- *AI-Enabled Frontline Workers: A Multilingual Multimodal Point-of-Care LLM Evaluation Study in Nigeria*. Tobi Olatunji, Chinemelu Aka, **Chibuzor Okocha**, Emmanuel Ayodele, Jennifer Orisakwe, Toni Adekunle, Mardhiyah Sanni, Abdulameed Abiola, Oluwatomi Owopetu, Tolu Afolaranmi, Peter Suoyo Yougha, Mira Emmanuel-Fabula, Vaishnavi Menon, Alastair Denniston, Bilal Mateen (2025). [Link]

Workshop Papers

- *A Comparative Analysis of Speaker Diarization Models: Evaluating a Dataset for African Accented Speech*. **Chibuzor Okocha**, Kelechi Ezema, Christan Grant (2025). *IEEE Automatic Speech Recognition and Understanding Workshop* [ASRU 2025, in review].
- *Evaluating the Multilingual and Accent Robustness of Speech Language Models*. **Chibuzor Okocha**, Tobi Olatunji, Mardhiyah Sanni, Busayo Awobade, Devendra Kayande, Gloria Katuka, Naome Etori, Chinemelu Aka, Fola Omofoye (2025). *IEEE Automatic Speech Recognition and Understanding Workshop* [ASRU 2025, in review].
- *We Lost 2 Decades: A Large-Scale Multi-Country Study on Healthcare Documentation Speeds in LMICs*. Tobi Olatunji, Chinemelu Aka, Tassallah Abdullahi, **Chibuzor Okocha**, Gloria A. Katuka, Naome A. Etori, Lukman E. Ismaila (2025). *HCI and Health Workshop at ACM CHI*. [Link]
- *Ethical Challenges Facing Data-Driven Policing: A Review*. **Chibuzor Okocha** (2022). Poster presented at the *Black in AI Affinity Workshop at NeurIPS*. [Link]

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.