CHIBUZOR OKOCHA JOSEPH

Gainsville, 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

- **Research Skills:** AI research, Qualitative research, Designing interview protocols, data analysis, quantitative research, interviewing.
- **Technical Skills:** Proficient in Python, Finetuning LLMs, RAG, Agent systems, Typescript, Next.Js, JavaScript, HTML, CSS, SQL.
- **Visualizations:** Tableau, Microsoft PowerBI, Visio, LucidChart
- **Tools:** Google Colab, Jupyter Notebook, JMP, SPSS, GitHub, Vercel, Streamlit.
- **Framework:** Pytorch, Tensorflow, Keras, Jax, Scikit learn
- **Specializations:** Prompt Engineering, Requirements Engineering, Project Management, Process documentation, Apps/ tools support, Presentation skills

WORK EXPERIENCE

Researcher *August 2023 - Present*
Bio-RAMP labs, Masakhane NLP

- Curate and maintain a comprehensive dataset of African-accented conversations across medical and non-medical contexts to enhance speech and language model performance.
- Analyze ASR outputs to assess accuracy and reliability in transcribing African-accented speech, with a focus on improving performance in clinical and health-related conversations.
- Develop and fine-tune 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.

Research Intern *August 2024 - January 2025*
Intron Health

- Explored multimodal "voice-native" large language models (LLMs) to assess their potential for supporting low-resource languages and African-accented speech, informing future model development.
- Researched state-of-the-art multimodal LLMs, including their architecture and training strategies, to evaluate their applicability in speech-based African language tasks.
- Managed end-to-end data collection workflows from multiple contributors, ensuring high-quality and timely submission of African-accented speech through streamlined coordination and automated quality assurance processes.
- Fine-tuned advanced TTS models such as VITS and XTTS on a custom Intron dataset and public data, resulting in improved naturalness, intelligibility, and accentedness of synthesized African voices.
- Co-authored and presenting research findings at leading venues including Interspeech and ACM CHI, contributing to the broader discourse on speech technology inclusion for African language users.

Graduate Research Assistant, *June 2023 - May 2025*
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.

AfriSpeech-200 - GitHub - Property of Intron

- Released AfriSpeech, a 196-hour Pan-African speech dataset covering 120 indigenous accents, to address the lack of publicly available benchmarks for African-accented clinical and general domain ASR.
- Highlighted critical performance gaps in commercial ASR systems, especially racial and accent bias, by benchmarking African-accented clinical speech against state-of-the-art ASR models.

Multi-Agent Systems - GitHub

- Designed and orchestrated multi-agent workflows using CrewAI and OpenAI models to demonstrate autonomous collaboration on complex, multi-step tasks.

Real-Time Feedback System - GitHub

- Designed a smart feedback tool that analyzes students' oral presentations in real time, evaluating clarity, pacing, filler words, tone, and emotional expression to enhance communication effectiveness.

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

Papers

- **M**ardhiyah 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).
- **M**ardhiyah 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

- **T**obi 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].

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.