




JOY OLUSANYA

+234 7063842298 ✉ joynaomiolusanya@gmail.com  LinkedIn  GitHub  Website

Career Summary

Linguistics graduate with 4 years of experience across research and engineering roles in industry and academia, building datasets, evaluation pipelines, and multilingual NLP systems with a focus on low-resource and African languages and Cultures. Experienced in developing speech and language technologies, including Automatic Speech Recognition (ASR), text-to-speech (TTS), and Machine Translation (MT) systems, such as Yoruba and other underrepresented languages. Skilled in data preparation, model training, evaluation, and deployment using Python, PyTorch, and modern ML frameworks. Passionate about leveraging AI to preserve languages, enhance digital inclusion, and advance AI safety and governance through impactful research-driven solutions.

KEY SKILLS

- ✓ Responsible AI
- ✓ AI Safety Evaluation
- ✓ Project Management
- ✓ Cultural Awareness
- ✓ Natural Language Processing
- ✓ Low-resource Languages
- ✓ Endangered Languages
- ✓ Linguistic Data Analysis
- ✓ Machine Learning
- ✓ Community-Centered AI
- ✓ Dataset Curation

TOOLS & FRAMEWORK

• Python (pandas, NumPy, scikit-learn, matplotlib) • Julia • SQL • Deep Learning Frameworks (PyTorch, TensorFlow, Keras, JAX, Scikit-learn) • NLP Language Models (Transformers, Hugging Face, NLTK, spaCy, Machine Translation, ASR, LLM Evaluation) • Model Evaluation (BLEU, ROUGE, BERTScore, COMET, Human Annotation, Benchmarking) • MLOps Tools (Docker, MLflow, WB, GitHub, AWS, GCP) • Web Frameworks / APIs (Streamlit, FastAPI, Flask, LangChain, OpenAI API, Gemini API, Tavily API)

LINGUISTIC SKILLS

• Linguistics & Computational Linguistics (Phonetics, Morphology, Syntax, Semantics) • Linguistics Software (Praat, ELAN, FLEx, Audacity)

WORK EXPERIENCE

NLP Researcher & Yoruba Language Lead

August 2025 – Present

Tonative

Lagos, Nigeria

- **Dataset Creation:** Developed multilingual NLP datasets for low-resource African languages, ensuring linguistic diversity.
- **Machine Translation Contribution:** Translated English into Kikuyu, Kinyarwanda, and Yoruba, contributing to research paper work.
- **Benchmarking:** Performed benchmarking using OpenAI healthcare datasets to evaluate model performance and data quality.
- **Data Validation:** Directed data collection, annotation, and validation workflows to guarantee high-quality resources.
- **System Design:** Collaborated on multilingual NLP system design to improve accessibility, inclusivity, and

translation quality.

NLP / ML Engineer Lead

August 2025 – Present

YorLect

Remote

- **Model Development:** Led research to improve NLP models' performance for Yoruba and its dialects through theoretical and applied linguistics.
- **Dataset Creation:** Developed language resources and corpora, building high-quality evaluation benchmarks for low-resource NLP tasks.
- **Model Deployment:** Supervised model training, evaluation, and deployment pipelines to ensure accuracy and efficiency.

Reviewer

August 2025

Deep Learning Indaba Ideathon 2025

Remote

- **Evaluation:** Reviewed AI/ML project submissions, ensuring fairness, quality, and adherence to research standards.
- **Community Support:** Promoted collaboration and knowledge sharing in African AI research.
- **Certification:** View Certificate.

AI/LLM Developer

April 2025

Atom Group (Contract)

Lagos, Nigeria

- **AI Agent Development:** Developed a commercial AI agent with Instagram scraping pipelines for social media intelligence.
- **Data Systems:** Built modular, automated data collection and API integration systems to enhance efficiency and reliability.
- **Scraping Optimization:** Improved scraping accuracy and reduced failure rates through robust exception handling and reusable utilities.

AI Developer

January 2024 – March 2025

Brilla AI Project, Kwame AI Ghana

Remote

- **Speech-to-Text Models:** Developed deep learning ASR models for multilingual speech with accent robustness.
- **Data Validation:** Validated and curated diverse datasets to improve transcription accuracy and model reliability.
- **Evaluation Workflows:** Built reproducible evaluation workflows and aligned preprocessing pipelines with training for scalability.

ML/NLP Engineer

June 2024 – December 2024

Rendo AI Nigeria

Lagos, Nigeria

- **Web Scraping Pipelines:** Designed Python-based pipelines using Requests, Selenium, and BeautifulSoup.
- **Data Automation:** Automated large-scale data acquisition, improving efficiency and dataset completeness.
- **Reusable Tools:** Developed reusable scraping tools to handle structured data from diverse sources.

LEADERSHIP EXPERIENCE

Founder & CEO

2025 – Present

Center for Low-Resource Languages and Cultures (CLRLC)

Nigeria

- **Research & Leadership:** Leading initiatives focused on low-resource languages research, multilingual NLP, and culturally aware AI systems.
- **Community Building:** Building communities and collaborations around underrepresented languages through workshops, mentorship, and research initiatives.

Training Manager

January 2026 – Present

Tonative

Lagos, Nigeria

- **Training Design:** Designed and implemented curricula for participants across diverse linguistic and cultural contexts, improving learner engagement and outcomes.
- **Recruitment & Onboarding:** Led the recruitment, interview, and onboarding process for tutors, evaluating technical expertise and pedagogical alignment.
- **Quality Assurance:** Reviewed and validated tutors' teaching materials to ensure curriculum alignment and program quality.
- **Performance Monitoring:** Coordinated feedback collection and performance assessments to refine training methodologies.

Workshop Chair

September 2025

CLRLC-LLMs Workshop @ NeurIPS 2025

Mexico City, Mexico

- **Event Coordination:** Planned and coordinated an international research workshop, engaging stakeholders from academia and industry.
- **Program Management:** Managed program committee, speaker invitations, and community partnerships for smooth execution.
- **Technical Design:** Structured the workshop program, including keynotes, lightning sessions, and interactive activities.

AI Mentor

September 2024 – November 2024

LevelUp Techies Bootcamp

Remote

- **Mentorship:** Guided 50+ participants in Python programming and machine learning fundamentals, improving technical readiness and confidence.

COMMUNITY ENGAGEMENT

Deep Learning / AI Co-Lead

January 2025 – November 2025

Data Science Nigeria

Remote

- **Training:** Facilitated hands-on deep learning sessions for 20+ participants, improving skill development and collaboration.
- **Event Planning:** Organized national events and hackathons, increasing community engagement.
- **Mentorship:** Managed bootcamp programs and mentored emerging AI professionals.

AI / Machine Learning Lead

October 2024 – November 2025

Women in Data Science Nigeria (WiDS)

Remote

- **Workshop Facilitation:** Organized AI/ML workshops for women across skill levels.
- **Learning Design:** Developed structured learning paths and accessible educational resources.
- **Community Engagement:** Moderated panels and sessions, fostering mentorship and inclusivity.

LANGUAGES

- Yoruba (Native)
- English (Fluent)
- Nigerian Pidgin English (Fluent)
- German (Basic)

SELECTED PROJECTS

Evaluating LLM Performance on Nigerian Pidgin Sentiment

2026

[Link](#)

- *Zero-Shot Sentiment Classification · GPT-4o-mini · NLP · Python*
- Benchmarked GPT-4o-mini on the NaijaSenti corpus (Nigerian Pidgin subset of MTEB/NaijaSenti)
- Designed a zero-shot prompt leveraging Pidgin linguistics knowledge to classify sentiment into Positive, Neutral, and Negative classes.
- Evaluated on a stratified sample of 2,000 examples achieving **Accuracy: 63.20%**, **Macro F1: 54.92%**, and **Weighted F1: 69.10%**.
- Built a full evaluation pipeline including stratified sampling, metrics computation, and visualisations (confusion matrix, F1 bar chart).

Yorùbá Proverb Classification

July 2024 – September 2024

[Link](#)

- **Research Focus:** Investigated culturally grounded semantic classification in a low-resource setting.
- **Model Comparison:** Compared Naive Bayes with transformer-based models (BERT, AfroLM) for proverb interpretation.
- **Linguistic Analysis:** Analysed morpho-semantic patterns and ambiguity in culturally embedded texts.
- **Evaluation:** Assessed performance using precision, recall, and F1-score with structured error analysis.

Multilingual Translation Evaluation: mBART50

August 2025

[Link](#)

- **Evaluation Design:** Evaluated pretrained mBART50 for English–French translation using BLEU and ROUGE metrics.
- **Error Typology:** Conducted qualitative error analysis on syntactic divergence and lexical ambiguity.
- **Experimental Analysis:** Examined model behaviour across sentence length and structural complexity.

Neural Machine Translation: English Portuguese

June 2025

[Link](#)

- **Fine-Tuning:** Adapted Helsinki MarianMT transformer using domain-aware preprocessing strategies.
- **Benchmarking:** Evaluated translation quality using BLEU and complementary metrics.
- **Comparative Study:** Analysed baseline versus fine-tuned performance through structured evaluation.

Automated Essay Scoring with DistilBERT

June 2025

[Link](#)

- **Model Development:** Fine-tuned DistilBERT for regression-based essay scoring.
- **Human Alignment:** Evaluated correlation between automated predictions and human grading.
- **Comparative Analysis:** Benchmarked transformer performance against classical ML approaches.

EDUCATION

Bachelor of Arts in Linguistics; Obafemi Awolowo University, Osun State, Nigeria
September 2021 – September 2025

January 2021 –

AWARDS

NeurIPS 2025 Scholar Award <i>Neural Information Processing Systems (NeurIPS)</i>	2025 <i>Mexico City, Mexico</i>
WiML @ NeurIPS 2025 Travel Funding Award <i>Women in Machine Learning (WiML) Workshop at NeurIPS</i>	2025 <i>Mexico City, Mexico</i>
Deep Learning Indaba 2025 Travel Grant <i>Deep Learning Indaba Conference</i>	2025 <i>Kigali, Rwanda</i>

PUBLICATIONS

- Tone in Yoruba Automatic Speech Recognition: Evaluating the Impact of Tone Recognition on Transformer-Based ASR Models** ([link](#))
Accepted to LoResLM Workshop @ EACL 2026, European Chapter of the Association for Computational Linguistics, 2026
- ÒWE-YOR: Leveraging Transformer-Based Models for Yoruba Proverb Classification** ([link](#))
Accepted to WiML @ NeurIPS 2025, Women in Machine Learning Workshop, 2025