

Sharon Ibejih

☎ +2348128253909 • 🌐 <https://www.linkedin.com/in/sharonibejih/>
🌐 <https://bit.ly/sharonibejih-google-scholar-profile>
✉ sharonibejih@gmail.com

Research Interest

Low-resourced NLP, Multilingual Language Models and Low-resourced Data Collection/Generation

Education

B.Sc. (Hons), Physics

Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria.

2013 - 2017

CGPA: 4.07/5.00

Skills

Python, NumPy, Pandas, Seaborn, Scikit-Learn, TensorFlow, PyTorch, NLTK, spaCy, HuggingFace, MLflow, Streamlit, Flask, Docker, Git, Linux, AWS (EC2, S3, RDS, SageMaker)

Work Experience

Insidatics AI Inc., Delaware – ML Software Engineer

Feb. 2022 – Present

- Collaborated with the team to develop training pipelines for the finance industry's credit scoring and fraud detection machine learning model.
- Partook in the DevOps processes for deploying the software solutions on the cloud.
- Contributed to developing python APIs for model explainability charts using shap.
- Tested the API endpoints using Postman and fixed detected bugs.
- Developed and deployed building materials price prediction model for a client.

RectLabs Inc., Lagos – AI Engineer

July – Sept. 2022

- Developed an optical character recognition system for converting English and Yoruba written PDFs to TXT files, using pytesseract.
- Deployed sports analysis and object auto-annotation AI models.

Data Scientists Network (DSN), Lagos – Research and Solution Analyst

Oct. 2020 – April 2022

- Developed a Pidgin Speech Recognition model by fine-tuning pretrained NeMo QuartzNet 15x5 ASR model. This project attracted an international contract.
- Experimented with KerasOCR and TrOCR for medical handwriting recognition using synthetic data. The outcome showed that these fine-tuned models performed well on the synthetic data and not on a real prescription, therefore leading to further research.
- Co-wrote a proposal for the 2021/2022 Lacuna Funding for community-curated local dataset.
- Developed a complete audio analytics pipeline for call-center data. This entailed building a speaker diarizer (who spoke when) for a mono channel to separate the customers speech from the call agents; transcribe each speech chunk to analyse the output text. The success of this project, led to a bigger contract for the organisation.
- Led the collection and annotation of speech data in the team.
- Carried out machine learning research experiments for publications.
- Co-wrote the AI Ethics document for the organisation.
- Managed the development process of an EdTech product.
- Led the product and solution briefings, ran live product demos, and prototypes.

Selected Research Experience

Co-Investigator , *Oct. 2021 – Feb. 2022*

Project Title: TCNSpeech: A Community-Curated Speech Corpus for Sermons.

Project Status: Completed

Advisors: Wuraola Oyewusi

Synopsis: To curate publicly accessible multispeaker sermon corpus for speech recognition tasks.

Duties: Ideated on good data sources; provided a well simplified guide for the annotators including how to handle moments of prayers, singing, speaking in tongues, etc., in the audio samples. Experimented the curated data by developing an ASR model. Co-wrote the paper.

Investigator , *Feb. – Aug. 2021*

Project Title: EDUSTT: In-Domain Speech Recognition for Nigerian Accented Educational Contents in English.

Project Status: Completed

Advisors: Wuraola Oyewusi, Dr. Bayo Adekanmbi

Synopsis: To improve the speech recognition of educators who have Nigerian English accent.

Duties: Conceptualized the idea, sourced for related audio data and created annotation guide for the team. Reviewed existing literature, documented approach and results of the experiments.

Co-Investigator , *May – Sept. 2021*

Project Title: AFRIFASHION1600: A Contemporary African Fashion Dataset for Computer Vision.

Project Status: Completed

Advisors: Wuraola Oyewusi

Synopsis: To generate benchmark African Fashion dataset.

Duties: Supported in scraping fashion wears from the web. Developed various computer vision models such as CNN, ResNet and VGG16 to benchmark the dataset performance. Co-wrote the research paper.

Selected Research Publications

Journal Paper

Oyewusi, W. F., Adekanmbi, O., **Ibejeh, S.**, Osakuade, O., Okoh, I., & Salami, M. (2021). AFRIFASHION1600: A Contemporary African Fashion Dataset for Computer Vision. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3968-3972)*.

Conference Papers

Oyewusi, W.F., Adekanmbi, O., Okoh, I., Onuigwe, V., Salami, M.I., Osakuade, O., **Ibejeh, S.** & Musa, U.A. (2021). NaijaNER: Comprehensive Named Entity Recognition for 5 Nigerian Languages. *arXiv preprint arXiv:2105.00810*.

Ibejeh, S., Oyewusi, W. F., Adekanmbi, O., & Osakuade, O. (2022, March). EDUSTT: In-Domain Speech Recognition for Nigerian Accented Educational Contents in English. In 3rd Workshop on African Natural Language Processing

Oyewusi, W. F., **Ibejeh, S.**, Uzomah, S., Joseph, E. M., Cynthia, J., Ojemuyiwa, F., ... Akinjobi, S. (2022, March). TCNSpeech: A Community-Curated Speech Corpus for Sermons. In 3rd Workshop on African Natural Language Processing

Speaking Engagements

Computational Research and Open Science Community Mentorship Indaba, Apr. 2021

Topic: She Code Africa Community Mentorship Project

Deep Learning IndabaX Nigeria, Sep. 2021

Topic: Poster Presentation for "Artificial Intelligence for Pharmacovigilance in Nigerian Social Media Text"

AI Saturdays Lagos, Sept. 2022

Topic: Opportunities in Machine Learning

DatafestAfrica 2022, Oct. 2022

Topic: Who is a Researcher?

Honours, Awards and Recognition

Awarded the **Best Poster** for our research project *AFRIFASHION1600: A Contemporary African Fashion Dataset for Computer Vision* at **Deep Learning IndabaX Nigeria 2021**.

Volunteering and Community Events

AI Saturdays Lagos

Event Coordinator of Lagos NeurIPS Meetup 2019 Dec. 2019

She Code Africa

DS/ML/AI Channel Lead Dec. 2020 - Jan. 2022

AI Saturdays Lagos

Cohort 7 Co-organiser and Mentor Sept. 2021 - Jan. 2022

ICLR 2022

ICLR Conference Volunteer April 2022

DatafestAfrica

Python Mentor Oct. 2022 - Present

Professional Affiliations and Membership

AI Saturdays Lagos

Data Scientists Network (DSN)

Masakhane Community

Women in Machine Learning

MLOps Africa Community