

# SHARON IBEJIH

Lagos, Nigeria

Email: [sharonibejih@gmail.com](mailto:sharonibejih@gmail.com) Web: <https://sharonibejih.github.io/>

## EDUCATION

---

2013 – 2017      B.Sc. Physics  
Michael Okpara University of Agriculture, Umudike, Nigeria  
GPA: 4.07 / 5.0

## RESEARCH EXPERIENCE

---

2022      **Co-Lead Investigator**  
Project Title: A Probability-Based Optimiser for Boosting Breast Cancer Classification in Ultrasound Images.  
Collaborators: Olawale Abimbola.  
Project Status: In review at Black in AI Workshop, NeurIPS 2022  
Synopsis: The paper presents a much faster and more accurate technique that outperforms other state-of-the-art approaches for classifying cancer in ultrasound breast images. Our work was developed using Xception for feature extraction, Bayesian optimisation for hyper-parameter tuning and the LightGBM classification algorithm for modelling.  
Duties:  
○ Reviewed literature.  
○ Writing.

2022      **Co-Lead Investigator**  
Project Title: TCNSpeech: A Community-Curated Speech Corpus for Sermons  
Collaborators: Wuraola Oyewusi  
Project Status: Accepted at AfricanNLP workshop at ICLR 2022  
Synopsis: We presented TCNSpeech, a community-curated multispeaker sermon corpus for speech recognition tasks. It contains a total of 24 hours of English audio data recording, chunked and transcribed. The context of the dataset is domain-specific for sermons in a Nigerian English accent.  
Duties:  
○ Provided a simplified guide for the annotators including how to handle moments of prayers, singing, speaking in tongues, etc., in the audio samples.  
○ Experimented with the curated data by developing an ASR model by fine-tuning NeMo QuartzNet pre-trained ASR model.  
○ Writing.

2021 - 2022      **Lead Investigator**  
Project Title: EDUSTT: In-Domain Speech Recognition for Nigerian Accented Educational Contents in English  
Project Status: Accepted at AfricanNLP workshop at ICLR 2022

Collaborators: Wuraola Oyewusi, Olubayo Adekanmbi, Opeyemi Osakuade

Synopsis: We propose an ASR system for education in Nigerian accents.

Our experiment leveraged fine-tuning NeMo's QuartzNet15x5 English model on our accented education data.

Duties:

- Conceptualised the project idea.
- Facilitated the gathering and annotating of Nigerian education speech data in English.
- Reviewed existing literature; experimented with the NeMo QuartzNet15x5 English model.
- Wrote the research paper draft.

2021

### **Co-Investigator**

Research Interest: AFRIFASHION1600: A Contemporary African Fashion Dataset for Computer Vision.

Project Status: Published at the 35th Conference on Neural Information Processing Systems (NeurIPS).

Collaborator: Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, Opeyemi Osakuade, Ifeoma Okoh, Mary Salami

Synopsis: This work presents AFRIFASHION1600, an openly accessible contemporary African fashion image dataset containing 1600 samples labelled into 8 classes representing some African fashion styles. Each sample is coloured and has an image size of 128 x 128.

Duties:

- Reviews literature
- Developed computer vision models for the benchmark dataset using CNN and pre-trained models.
- Writing.

2021

### **Co-Investigator**

Research Interest: NaijaNER: Comprehensive Named Entity Recognition for 5 Nigerian Languages

Project Status: Accepted at the AfricaNLP Workshop, EACL 2021

Collaborator: Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, Opeyemi Osakuade, Ifeoma Okoh, Vitus Onuigwe, Mary Salami, Usman Abdullahi Musa

Synopsis: In this work, we developed named entity recognition (NER) models in 5 Nigerian languages by fine-tuning spaCy's NER model. We also worked on a combined model that can handle NER for any of the five languages. The combined model works well for identifying entities in each of the languages and with better performance compared to individual NER models trained specifically on annotated data for the specific language.

Duties:

- Reviews literature.
- Scrapes online Pidgin news articles, which is one of the languages used in the study, developed an individual NER model for the Pidgin language and the combined NER model, using spaCy's blank model.
- Writing.

## **PUBLICATIONS**

---

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

**Ibejih, 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 [[link](#)]

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

Oyewusi, W. F., Adekanmbi, O., **Ibejih, S.**, Osakuade, O., Okoh, I., & Salami, M. (2021). AFRIFASH-ION1600: A Contemporary African Fashion Dataset for Computer Vision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3968-3972) [[link](#)]

## **TECHNICAL STRENGTHS**

---

Machine Learning: Scikit-Learn, TensorFlow, Pytorch, OpenCV, NLTK, spaCy, HuggingFace  
Data Science: NumPy, Pandas, Matplotlib, Seaborn, MLflow, Streamlit  
Software: Flask, Docker, GitHub Actions, Linux, Git  
Others: Open-Source, Web Scraping, AWS (EC2, S3, RDS, SageMaker)

## **WORK EXPERIENCE**

---

2021 - Present

### **Co-Founder**

Organisation: Kadatemy Data School

Duties:

- Manage the operations of the school
- Python tutor.

2022 - Present

### **Machine Learning Software Engineer (Contract)**

Organisation: Insidatics AI Inc., Delaware (Remote)

Duties:

- 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.

2022

### **AI Engineer**

Organisation: RectLabs Inc., Lagos

Duties:

- 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.

2020 - 2022

### **Research and Solution Analyst**

Organisation: Data Scientists Network (DSN), Lagos

Duties:

- Developed a Pidgin Speech Recognition model by fine-tuning pre-trained 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 a community-curated local dataset on 5 Nigerian languages.
- Developed a complete audio analytics pipeline for call-centre 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 prototype

## **HONOURS, AWARDS & RECOGNITION**

---

2021

### **Best Poster Award**

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

## **TALKS**

---

2022

### **DatafestAfrica 2022**

Gave a talk on *Who is a Researcher?* This topic targets machine learning beginners who are curious about research. [[Slide](#)]

2022

### **AI Saturdays Lagos**

Gave a talk on *Opportunities in Machine Learning from a Community Manager Perspective*. The objective of this topic is to enlighten the audience on the broad opportunities ML can provide for community people, such as research collaborations, DevRel, etc. [[Slide](#)]

2022

### **AfricanNLP, ICLR'22**

Presented our papers on *EDUSTT: In-Domain Speech Recognition for Nigerian Accented Educational Contents in English*. In the 3rd Workshop on African Natural Language Processing and *TCNSpeech: A Community-Curated Speech Corpus for Sermons*. In the 3rd Workshop on African Natural Language Processing.

2021

### **Deep Learning IndabaX Nigeria**

Presented our paper on *Artificial Intelligence for Pharmacovigilance in Nigerian Social Media Texts*.

2021

### **Computational Research and Open Science Community Mentorship Indaba**

Shared at the event how we manage the She Code Africa Mentorship Programmes and Projects. [[Slide](#)]

## **VOLUNTEERING & COMMUNITY EVENTS**

---

2022 - Present

### **Python Mentor**

Organisation: DatafestAfrica

Duties:

- Provide learning materials to the community members
- Meeting with them weekly to address their possible difficulties

2022

### **ICLR'22 Conference Volunteer**

Organisation: ICLR 2022

Duties:

- Worked with other volunteers to help attendees navigate their way in GatherTown.
- Supported during the live sessions to transfer attendees' questions in the rocket chat to the Zoom chat.

2021 - Present

### **Co-organiser and Mentor**

Organisation: AI Saturdays Lagos

Duties:

- Assisting in overseeing the program operations
- Mentored a group of students during the Cohort 7
- Handled the Data Analysis Practical Class in the Cohort 7 [[Notebook](#)]

2020 - 2022

### **Data Science Channel Lead**

Organisation: She Code Africa

Duties:

- Was responsible for managing the community Data Science Members.
- Increased the number of engagements and members in the channel, from 600 to 1,200 women.
- Introduced events like Hackathons and Project Sprints.