

# Ifeoma Doris Okoh

+2348139112379 | [odibest1893@gmail.com](mailto:odibest1893@gmail.com) | [Google Scholar](#) | [ifyokoh.github.io](https://github.com/ifyokoh)

## EDUCATION

---

### University of Ibadan

*Master of Science in Mathematics*

Oyo, Nigeria

2018 – 2020

### University of Nigeria

*Bachelor of Science in Mathematics*

Nsukka, Nigeria

2011 – 2015

## RESEARCH EXPERIENCE

---

### Pre-Doctoral AI Research Fellow | *Fatima Fellowship*

Sept 2024 – June 2025

- Created a parallel resource for multilingual LLM red teaming that enables cross-lingual and cross-cultural analyses with focus on tasks that have harmful societal impacts.
- Finetuned model to abstain even in hard cases of low-resource misinformation.

### NLP Researcher | *Masakhane*

Aug 2023 – Present

- Contributed to the IgboAPI dataset project as part of a team, leading the fine-tuning of machine translation models to improve performance on dialectal variations, resulting in enhanced representation of Igbo dialects in the model as demonstrated in this [paper](#).
- Applied Elastic Weight Consolidation (EWC) to improve NER model generalization across domains in multiple languages, enhancing F1 scores in MRL 2024 Shared Task on Multi-lingual Multi-task Information Retrieval.

### Research Intern | *Data Science Nigeria*

Oct 2020 – Apr 2021

- Led a collaborative project developing a CNN-based gesture analysis system to monitor and analyze student activities in classrooms, providing insights into academic performance and engagement.
- Pioneered a Named Entity Recognition (NER) model for five under-resourced Nigerian languages, covering the research lifecycle from data collection to model development, and co-authored a research paper, achieving a foundational F1 Score of 62.43% as demonstrated in this [paper](#).

## PUBLICATIONS

---

- Ifeoma Okoh\***, Chris Chinenye Emezue\*, Chinedu Mbonu, Chiamaka Chukwunke, Daisy Lal, Ignatius Ezeani, Paul Rayson, Ijemma Onwuzulike, et al. [The IgboAPI Dataset: Empowering Igbo Language Technologies through Multi-dialectal Enrichment](#) *LREC-COLING Conference 2024* [[Paper](#)]
- Shivalika Singh, Freddie Vargus, Daniel Dsouza, ..., **Ifeoma Okoh**, ..., Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee, Sara Hooker [Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning](#) *ACL Conference 2024* [[Paper](#)]
- Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, **Ifeoma Okoh**, Vitus Onuigwe, Mary Idera Salami, Opeyemi Osakuade, Sharon Ibejih, Usman Abdullahi Musa [Naijaner: Comprehensive named entity recognition for 5 Nigerian languages](#) *AfricanNLP workshop at EAACL 2021* [[Paper](#)]
- Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, **Ifeoma Okoh**, Mary Idera Salami, Opeyemi Osakuade, Sharon Ibejih, Vitus Onuigwe [Artificial intelligence for pharmacovigilance in Nigerian social media text](#) *AI for Public Health Workshop at ICLR 2021* [[Paper](#)]
- Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, Sharon Ibejih, Opeyemi Osakuade, **Ifeoma Okoh**, Mary Salami [AFRIFASHION1600: a contemporary African fashion dataset for computer vision](#) *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021* [[Paper](#)]

## PRESENTATION

---

- LREC-COLING Conference 2024, Italy. [The IgboAPI Dataset: Empowering Igbo Language Technologies through Multi-dialectal Enrichment](#) *Oral Presentation*
- Women in Data Science Conference 2019, Ibadan. Invited to speak following success in achieving second runner-up at the 2019 WiDS Hackathon *Panelist*.

## INDUSTRY EXPERIENCE

---

### Data Scientist | *Prunedge*

Jan 2022 – Aug 2023

- Developed a real-time sentiment analysis model using Twitter data and NLP techniques (Tweepy, NLTK) to assess product sentiment, providing insights for product adjustments via an interactive dashboard (Flask).
- Utilized GraphQL API for data extraction from GitHub, followed by analysis and cleansing, enabling real-time performance tracking of 100+ developers via ELK stack integration, showcasing its utility in software engineering research and project management.

### Data Scientist | *Monetha*

Jun 2021 – Nov 2021

- Formulated a seasonal segment model employing the decision tree algorithm to analyze purchasing patterns, facilitating data-driven decision-making for shop owners regarding seasonal product demand.
- Analyzed customer lifetime value, using statistical methods and data analytics to understand and predict long-term customer behavior.

## AWARDS AND ACHIEVEMENTS

---

- Selected among the top 500 in an international pool of 4,000 for the 2023 KaggleX BIPOC Mentorship Program.
- Secured second place at the 2023 Data Search Engine Hackathon hosted by a German company, Innoscripta.
- Achieved second runner-up position among 204 international participants in the 2019 Women in Data Science (WiDS) Hackathon.
- Awarded full sponsorship for travel, accommodations, and conference fees to attend the 2019 ACM COMPASS Conference on Computing and Sustainable Societies in Accra, Ghana.
- Received full sponsorship for travel, accommodations, and conference fees to attend the Web and Internet Economics (WINE) conference in Oxford, UK, 2018.

## TEACHING AND MENTORING

---

### Data Science Mentor | *She Code Africa*

Jan 2020 – Sep 2022

- Designed and implemented a beginner-friendly data science curriculum for women, mentoring learners through end-to-end project completion and achieving a 70% success rate in skill enhancement within three months.

### Python Instructor | *Computer Science Academy Africa*

Aug 2019 – Sep 2021

- Taught Python programming to university students, creating and delivering curriculum content that fostered computational thinking and coding skills.

## SERVICES

---

**Reviewer** | *AfricaNLP Workshop at ICLR 2024, Deep Learning Indaba 2024.*

**Language Ambassador and Researcher** | *Cohere for AI*

May 2023 – Present

- Contributed to an open-science NLP research project focused on instruction multilingualism for low-resource languages and co-authored a paper in the process.

**Women Techmakers Ambassador** | *Google's Women Techmakers*

Aug 2019 – Nov 2024

- Championing gender diversity in tech through community events, skill-building workshops, and advocacy at various forums.

## TECHNICAL SKILLS

---

**Programming Language:** Python

**Libraries:** PyTorch, Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn, Pyspark

**Frameworks:** Flask, FastAPI

**Developer Tools:** Git/Github, Docker, Airflow, ELK Stack, Google Cloud Platform, VS Code

**Others:** SQL, HTML/CSS, Latex

## LANGUAGE

---

**English:** Fluent , **Igbo:** Fluent