

Amina Mardiyah Rufai

Machine Learning Engineer

✉ amrufai@ebi.ac.uk

🌐 <https://www.linkedin.com/in/aminah-mardiyah-rufa-i>

🌐 <https://www.github.com/amina-mardiyah>

🏠 Google-Scholar

📍 Cambridge, United Kingdom

🔗 Amina's personal website

🗣️ HuggingFace Profile

📄 Medium Articles

Skills

Programming/Scripting Language:

Python, Shell

Packages/Libraries/ToolKits:

Pandas, Numpy, Matplotlib, Seaborn, Plotly, Pytorch, Scikitlearn, Git, HuggingFace Transformers etc

Machine learning, Deep Learning

Methods and Architectures:

Supervised & Unsupervised Modelling, EDA, Deep Learning, NLP, Automatic Speech Recognition(ASR), Xgboost, Transformer models, Computer Vision, SAM Model, Wav2vec2, Whisper, Pattern Recognition etc

Soft Skills:

Excellent Verbal and Written Communication Skills in English, Great Research Skills, Team player, Attention to details, Analytical Skills, Self-Motivated to Learn

Education

Masters in Machine Intelligence,

African Institute of Mathematical Sciences

Jan 2021 – Oct 2022 | Mbour, Senegal

Coursework: Foundations of Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, GNNs and Geometric Deep Learning, Reinforcement Learning, Research Paper implementation/Open Source Contribution etc.

Work Experience

Machine Learning Engineer,

European Bioinformatics Institute

May 2024 – Present | Cambridge, United Kingdom

- **Designed and implemented** end-to-end, scalable NLP pipelines for biomedical Named Entity Recognition (NER), enabling efficient extraction of entities such as cell lines and cell types from scientific literature
- **Led the integration** of state-of-the-art machine learning techniques for dataset curation and model development, significantly improving entity coverage and model performance.
- Explored and incorporated cutting-edge open-source LLMs to modernise existing architectures, with a focus on improving generalisation and reducing annotation dependency.
- **Developed** modular, reusable training pipelines using tools such as Weights & Biases and Hugging Face, enabling rapid experimentation, reproducibility, and streamlined model deployment.
- **Presented and published** research outcomes through peer-reviewed articles top-tier journals and conferences, contributing to the broader scientific and machine learning communities.

Core Technologies used: Huggingface toolkit, Git/Gitlab, Weight and Bias, Pytorch, Linux, Jira

Research Intern, Idiap Research Institute

Mar 2023 – present | Martigny, Switzerland

Tasks:

- Analysed text and speech dataset on storytelling to derive insights on storytelling abilities of interviewees and predict job performance.
- Developed a model by finetuning a language model able to classify storytelling versus non-storytelling features in a behavioural interview.

Core Technologies used: Huggingface toolkit, Git, XLMRoberta, Pytorch, Whisper, Linux CLI, Longformer etc

Research Summer Intern, École polytechnique

fédérale de Lausanne (EPFL) | MLO-IGH Lab

Jun 2022 – Aug 2022 | Lausanne, Switzerland

Tasks:

- Analysed a large time series data set from Twitter to derive insights from social media and predict the public's response to the pandemic health measures during COVID-19.
- Developed using state-of-the-art methodologies, a Multilingual Covid Sentiment Analyzer to uncover insights in socio-epidemiological behaviours during COVID-19.

Thesis: Statistical Machine Translation from English to Wolof


Masters of Science in Biomedical Engineering, *University of Lagos*
Jan 2018 – Mar 2019 | Lagos, Nigeria

- Top 5% in class
- Participated in a hackathon and Awarded with the Best Display of 21st Century Competencies and Best Female participant

Thesis: Classification of Retinal Fundus images with application to Glaucoma diagnosis.

Bachelor of Engineering in Chemical Engineering, *University of Benin*
Apr 2010 – Nov 2014 | Benin City, Nigeria


Publications

Towards an End-to-End Automatic Speech Recognition System for Nigerian Pidgin 
Aug 2025

Identifying storytelling in job interviews using deep learning,
Computers in Human Behavior Reports
May 2025

Natural language processing in drug discovery: bridging the gap between text and therapeutics with artificial intelligence, *Taylor & Francis*
Nov 2024

AfriWOZ: Corpus for Exploiting Cross-Lingual Transferability for Generation of Dialogues in Low-Resource, African Languages, *Arxiv | African NLP Workshop at ICLR 2022 | Published as a conference paper at IJCNN* 
2023

AI-Powered Understanding of Family Planning Using Fogg Behaviourial Model,
Published as a conference paper at ICLR 2022 
Apr 2021

Core Technologies used: AWS-CLI, huggingface toolkit, Git, Plotly, BERT, RoBERTa, mbert, Text2Blob, VaderAnalyzer, Pytorch

Solutions Analyst,
Data Scientists Network (Formerly Data Science Nigeria)
Nov 2021 – Jun 2022 | Lagos, Nigeria

- Designed, analysed, and engineered workflows for real-world problems using a combination of Deep Learning, Natural Language Processing, Automatic Speech Recognition, and Ontological Knowledge Graphs.
- Designed and developed a health-focused FAQ-Bot using DialogFlow and data insights from Audio data.
- Published and presented research paper accepted in ICLR 2022

Core Technologies used: Nvidia Nemo, Google DialogFlow, wav2vec, Audacity, Spacy, Librosa, PyAudio, Pytorch

Volunteer Experience

ACL, *Program Committee*
2023

Black in AI,
Volunteer and BlackAIR, Workshop Reviewer | 2021, 2022

Data Scientists Network (DSN),
Lead Co-ordinator of Ladies in AI community
2020 – 2022

African NLP Research Group (Masakhane),
Research Projects Collaborations: AfriWoz

Referees

Melissa Harrison, *Team Leader, EMBL-EBI*
mharisson@ebi.ac.uk

Phil Garner, *Senior Research Scientist,*
Idiap Research Institute
pgarner@idiap.ch

Martin Jaggi, *Professor, EPFL*
jaggi.martin@epfl.ch

Languages

English : — Fluent | **French :** — Basic