

# FAITH WAVINYA MUTINDA

wavinya.faith@gmail.com | linkedin.com/in/faithwmutinda | github.com/faith-wm | Philadelphia, PA

## Summary

Innovative Data Scientist with extensive expertise in NLP, specializing in fine-tuning language models for diverse tasks such as information extraction, summarization, and question answering. Skilled in end-to-end model development, including data preparation, distributed training, and benchmarking. Looking to leverage my technical skills in a data-driven role within a forward-thinking company.

## Skills

Natural Language Processing	Machine Learning	Communication
Retrieval-Augmented Generation	Large Language Models	Team Collaboration
Distribute Training	Prompt Engineering	Problem Solving

## Experience

NLP Postdoctoral Research Fellow

Children’s Hospital of Philadelphia, U.S.A.

July 2024 - Present

- Developed a **RAG-based LLM pipeline** using Llama and DeepSeek-Distill-Llama models to **extract and normalize** dysmorphic findings from unstructured genetics notes, improving model performance by **13%**.
- Fine-tuned **clinical QA LLMs** using **HPC GPU clusters** and **Cloud TPUs**; optimized performance, built benchmarking frameworks, and applied **RLHF** to improve alignment and accuracy.
- Mentored undergraduate students, providing **technical guidance** and overseeing **research projects**.

Python Data Scientist/Analyst

Turing, Remote

May 2024 - July 2024

- Developed and curated robust datasets in data science and Python coding domains for LLM training and evaluation, ensuring quality and consistency.

Data Scientist

Biofourmis Singapore PTE. Limited, Singapore

Apr 2023 - Nov 2023

- Developed **predictive machine learning models** using electronic health records, achieving high accuracy in **identifying patients at risk of chronic conditions**, bolstering precision medicine initiatives.
- Enhanced **hospital readmission prediction models, increasing accuracy by 10%** through extensive exploratory data analysis, feature engineering, and hyperparameter optimization.
- Collaborated with **cross-functional teams** to transform complex business requirements into effective and scalable data-science solutions.

Graduate Researcher (Data Scientist)

Nara Institute of Science and Technology, Japan

Apr 2019 - Mar 2023

- Developed a **Longformer-based transformer model** that streamlined the meta-analysis process, achieving an **F1 score above 80%** by efficiently **extracting core concepts** from biomedical research articles.
- Trained a **transformer-based deep learning model** to compute the degree of **semantic similarity** in clinical texts, achieving a **correlation of 90%** with human scores.
- Fine-tuned **BERT-based model** for **medication and context extraction** from unstructured clinical notes, attaining a **94% F1 score** and earning a **top 10 ranking** in a shared task.

## Education

Ph.D. in Information Science & Engineering

Nara Institute of Science and Technology, Japan

2023

Master’s in Information Science & Engineering

Nara Institute of Science and Technology, Japan

2020