HAMZAT TIAMIYU

Data Scientist and Machine Learning Engineer

@ hamzattiamiyu@gmail.com   📞 +2347040258976   📍 Abuja, Nigeria
💼 tiamiyuhamzat   🐙 AbooMardiiyah   ⓘD 0009-0001-4839-6927

SUMMARY

Machine learning and AI engineer with 5+ years of experience building and automating end-to-end machine learning systems that drive measurable business impact. Proven track record, including a 50% increase in talent recommendation accuracy and an 80% reduction in manual model retraining of a batch processing task. Skilled in deploying scalable pipelines using Docker, AWS, and GCP, with production-grade monitoring via Grafana.

Recent work includes fine-tuning a large language model (LLM) using the Unsloth library on a synthetic dataset generated using OpenAI and distilabel techniques and deploying it with FastAPI. Developed a prompt-to-podcast app using OpenAI for AI-generated content and ElevenLabs for generating audio based on the scripts from OpenAI. Also developed data pipelines to support the training of a text-to-SQL model. Currently working on revolutionizing the African financial market using AI.

TECHNICAL EXPERIENCE

Data and AI Engineer

Wamiri

📅 June 2025 – present   📍 Remote

- Developed a data collection pipeline to pull data from an ERP system (Acumatica) into Supabase for easy querying.
- Implemented a feedback endpoint for our SQL agent (text-to-SQL model) so expert users can enter correct SQL queries when the agent fails.
- Re-wrote the training script for the SQL agent to use updated data schema and user feedback as part of the training data for the next training cycle thereby improving the SQL agent's overall accuracy.
- Designed, Developed and deployed an AI reconciliation agent that reconciles bank data with General ledger from ERP systems.
- Implemented robust rule based logic for transaction matching and integrated AI to handle exceptions and give concise reasonsing why a match is considered, handle edge cases like many-to-one match (multiple bank accounts to one GL account).

Supabase | Acumatica | Langchain | Langgraph | GCS | Git | Python | SQL

AI solutions Consultant

Plugh

📅 Sept 2025 – present   📍 Remote

- Designed high-level architecture diagrams for the Tipsy AI system.
- Feedback was provided to AI engineers working on SLM fine-tuning and RAG design.

- Sync daily with the teams to discuss progress and suggest improvement methonds for scaling the AI system for Tipsy AI recommendation platform.

Machine learning Engineer

BigCodeGen

📅 Nov 2024 – Feb 2025   📍 Contract (Remote)

- Implemented a record linkage algorithm using sentence embedding to find similar records within tabular data.
- Integrated Vanna-AI (a text-to-SQL RAG system) to generate compex SQL queries from natural language.
- Developed the interface for the record linkage bot using streamlit.
- Co-authored a paper currently in review on probabilistic record linkage system.

SentenceTransformers   Spark   streamlit   Vanna AI   Python   SQL

Data Scientist

Toptal LLC (core team)

📅 Nov 2023 – Oct 2024   📍 USA (Remote)

- Co-led the project on talent recommendations by developing a catboost regressor model, which recommends favorable hiring rates for clients with an overall increased performance of 50% better than the baseline algorithm (heuristic model).
- Conducted research on designing a skill similarity ML system to increase the relevancy of Toptal talent profiles on search engines.
- Reduced manual model training efforts by 80% via a cloud-based orchestration framework (Apache Airflow/Composer) for the skill similarity ML system.
- Streamlined client onboarding by developing no-code Dify workflows to auto-generate interview scripts for matchers.

BigQuery   SentenceTransformers   FastText   Dify   Cloud Composer   Grafana   DVC   Git   Python   SQL

Data Scientist Intern

Toptal LLC (core team)

📅 May 2023 – Nov 2023   📍 USA (Remote)

- Completed a 6-month internship with a senior data scientist and coach, where I was taught best practices in writing Python codes for machine learning.
- Improved my understanding of Deep learning and traditional machine learning models, such as CatBoost, which is often used within the data science team.
- Learnt how to write proper unit tests for machine learning applications.
- Developed a CatBoost regressor model to predict the job lifetime value of Toptal clients for the products team.

GCP   Pytest   BigQuery   FastAPI   CatBoost   Pytorch   Python   SQL

AI | NLP Engineer
Uniccon Groups

📅 Feb 2023 – May 2023 📍 Abuja, Nigeria (On-site)

- Deployed cutting-edge speech-to-text, bi-directional translation (Hausa-English-Hausa), and conversational chatbots with FastAPI, Docker, and Kubernetes on an on-premise cloud.
- Collaborated with software and mobile engineers on requirements for consuming the ML endpoints on their applications.

`Pytorch` `Langchain` `Docker` `MiniKube` `Python` `Git` `FastAPI` `Jira`

Research AI Engineer
iQube Labs

📅 April 2021 – Jan 2023 📍 Lagos,Nigeria (Remote)

- Conducted research on speech recognition in low-resourced languages by training a QuartzNet and Wav2Vec model on Hausa, Yoruba, and Nigerian Pidgin languages.
- Led the speech-to-text project by developing an API endpoint using Flask.
- Conducted research on face aging applications by exploring groundbreaking papers on StyleGAN and CycleGAN.

`Python` `Pytorch` `NLP` `Computer Vision` `Machine Learning` `Flask`

EDUCATION

B.Sc
Electronic and Electrical Engineering

📅 May 2015 – Feb 2020 📍 Obafemi Awolowo University

GPA: 4.40/5.0 (Top 10%)

Coursework:
- Mathematical Methods 1 & 2 (Linear Algebra, Calculus)
- Engineering Analysis 1 & 2
- Introduction to Computer Packages
- Computational Structures
- Signals and Systems
- Statistics for Engineers
- Digital Circuit Analysis and Design
- Microelectronics Devices and Circuits

SELECTED PROJECTS

Talent Recommendation
Toptal LLC

- Built a talent-recommendation engine that recommends favorable hiring rates for clients.
- Deployed the application as a micro-service (FastAPI) to Google cloud
- Integrated monitoring and alerting system with Grafana.

Python | FastAPI | CatBoost | Google cloud | Grafana

Automated Skill Similarity
Toptal LLC

- Designed and developed a skill similarity algorithm that predicts secondary skills for primary skills, thereby increasing the relevancy of Toptal talent profiles .
- Deployed and orchestrated the data collection, training, and inference pipelines as a batch service that runs monthly with Cloud Composer.
- Implemented monitoring and alert rules when the model degrades in Grafana and alerting via Slack.

Python | Bigquery | Fasttext | OpenAI | Grafana | Airflow | DVC

Hausa-English-Hausa Translation service
Uniccon Groups

- Trained the Hausa-English-Hausa machine translation system.
- Containerized the FastAPI application and deployed it to Kubernetes on an on-premise cloud.
- Collaborated with mobile engineers to consume the endpoints in the mobile application.

Python | FastAPI | Transformers | Docker | minikube | Git

Hausa and Yoruba speech-to-text recognition
iQube Labs

- Researched and implemented speech recognition by training Quartznet and Wav2Vec models on Hausa, Yoruba, and Nigerian Pidgin languages.
- Deployed the applications using Flask.

Python | Flask | Wav2vec | Nvidia Quarznet | Bitbucket

Cybersecurity AI assistant
link-to-project
Personal Project

- I fine-tuned and deployed Qwen2.5-7B model on cybersecurity datasets from several GitHub repositories and tracked performance with Comet-ML.
- Implemented data collection, feature, training, and inference pipelines.
- Generated supervised fine-tuning/instruct dataset using OpenAI.

- Deployed the fine-tuned model with FastAPI and ngrok and saved the artifacts to huggingface.

Python   OpenAI   LLM   FastAPI   ngrok   Comet-ML

Podcast AI assistants
Personal Project

- Created chunked datasets from the Huberman Labs podcast transcripts.
- Built an RAG-based chatbot on top of the chunked datasets, integrated hybrid search, query re-routing, enhanced system performance with reranker models, and integrated LLM for dynamic question-answering.
- Built a front end with Streamlit for the application for early testing.

Python   Weaviate Vector DB   RAG   DeepEval   Streamlit

TECHNICAL SKILLS

Machine Learning & Data Science

- **Python**
- **Rust**
- **PyTorch**
- **Scikit-learn**
- **CatBoost**
- **Pandas**
- **HuggingFace**
- **DVC**

GenerativeAI

- **Langchain**
- **LlamaIndex**
- **RAG Systems**
- **LangGraph**

Model Deployment & Monitoring

- **Opik**
- **FastAPI**
- **Django**
- **Pytest**
- **ZenML**
- **Grafana**
- **Comet**
- **ML-flow**

Databases

- **MongoDB**
- **PostgreSQL**
- **Weaviate**
- **ChromaDB**

GCP Services

- **Vertex AI**
- **BigQuery**
- **Google Cloud**
- Storage

AWS Services

- **EC2**
- **S3**
- **Lambda**
- **Boto3**