

# Biniyam Belayneh Demisse

 biniyambelayneh376@gmail.com

 +251939890540

 Addis Ababa, Ethiopia

 Website

 Linkedin

 Github

## Profile

---

AI/ML Engineer focused on MLOps: builds reliable, compliant, and cost-efficient ML platforms and pipelines on AWS and GCP using MLflow, Airflow, Docker, Kubernetes/EKS, Spark and Vertex AI. Ships models to production with CI/CD, model registry/versioning, and observability, enabling faster releases and safer rollouts. Builds also end to end streaming and batch processing ELT/ETL pipelines working with s3 (or MinIO), RDS, Postgres, Kinesis Firehose, DBT, Bigquery, etc... Delivered measurable impact: 40% cloud cost reduction, 1-week automated releases (from 2 weeks), +23% CTR via audience modeling, 92% creative classification accuracy, and 3.5% downtime reduction

## Professional Experience

---

2024/12 – present

London, England

**Data Scientist / MLOps**

*Ozoneproject* ↗

- **Automated ML pipeline:** Designed training & inference pipeline for *Audience & Brand Safety* model using **Airflow, MLflow, and Docker**, cutting model deployment cycle from **2 weeks (manual) → 1 week (automatic)** while enabling full **visibility** of performance metrics through mlflow dashboard.
- **Creative classifier:** Built end-to-end pipeline combining **OpenCV + Tesseract OCR** for visual/text extraction and **fine-tuned RoBERTa / EfficientNet** models for classification, deployed on **Vertex AI**. Automated creative tagging by type and content with **92% accuracy**.
- **Audience segmentation:** Implemented **two-tower model** with automated training & inference pipeline on **TensorFlow and Spark Clusters** generating custom audience segments from demographic & interaction data, boosting **campaign CTR by 23%**.
- Orchestrated ELT pipelines using Airflow with MinIO as staging storage and BigQuery as the warehouse; built subject-area data marts to power audience segmentation and delivered automated reporting via Superset dashboards.

2024/08 – 2024/11

**AI Engineer**

*RizzBuzz*

- Built a no-code platform enabling end users to design and deploy multi-agent researcher chatbots using personal LLM API keys. Orchestrated workflows with LangChain, retrieved external information via SerpAPI, queried internal datasets with a SQL Agent, and generated user-friendly summaries with LLMs, allowing non-technical users to create complex agentic applications.
- Developed a robust monitoring system and Superset dashboard to track agent activity and overall platform performance, leveraging MinIO as a data lake for raw agent data and a PostgreSQL-based data mart for reporting and scheduled Airflow scripts periodically update the marts from the data lake every 6 hours.

2022/09 – 2024/08

London, England

**AI and MLOps Engineer**

*Adludio*

- Designed and implemented end-to-end machine learning pipeline for **Lorenzo Project** using mlflow for model registry, EKS for background long running ml tasks, API Gateway + lambda for API request handling, Cognito for auth services.
- Optimized cloud infrastructure, reducing costs by 50% through serverless transitions and efficient resource allocation.
- Automated a bid optimization bid list generation and upload.

- Created CI/CD pipelines and development environment on EC2 instances for the DS team to streamline deployment and accelerate production cycles.
- Built ETL pipeline for ad interaction tracking server for streaming data using Autoscaling EC2 servers for the tracker, Amazon Firehose for handling streaming and on the fly transformation and s3 for final destination.
- Managed secure cloud environments, ensuring compliance with best practices and passing third-party security audits.

2020/11 – 2022/09

Addis Ababa, Ethiopia

### Data Analyst

*Ethio Telecom* ↗

- **Automated data workflows:** Designed **Python + SQL** pipelines for ETL and visualization, reducing manual reporting significantly
- Predictive maintenance: Developed an ML model (scikit-learn, XGBoost) that reduced network equipment downtime by 3.5% in a pilot district.
- **Customer segmentation:** Built behavioral clustering model with **K-means (scikit-learn)**, uncovering **3 high-value segments** that informed marketing strategy.

## Education

---

2024/10 – Present

Liverpool,

UK (remote)

### MSc in Artificial Intelligence and Machine Learning

*Liverpool John Moores University + IIITB and Upgrad*

2022/05 – 2022/07

Addis Ababa, Ethiopia

### Machine Learning Engineering

*10 Academy (Bootcamp)* ↗

2021/07 – 2022/07

Addis Ababa, Ethiopia

### Software Engineering

*ALX Africa* ↗

2015/09 – 2021/01

Addis Ababa, Ethiopia

### BSc in Electrical Engineering focus Electronic Communication

*Addis Ababa Science and Technology University* ↗

## Projects

---

### End-to-End ELT and ML Pipeline

*Ad Unit Data classification*

- An ELT pipeline that consumes data from csv files uses MinIO/postgres as data lake airflow for orchestration. dbt for transformation, postgres as warehouse, Redash for dashboard, mlflow for model tracking and serving.

### Real-Time Bidding Optimization ↗

- Designed and deployed a bid optimization algorithm to improve ad performance and maximize engagement.

### RAG Pipeline & Chatbot ↗

- Built an interactive Retrieval-Augmented Generation chatbot leveraging FastAPI, Elasticsearch, and Google Generative AI.

### MultiPDF Chat App ↗

- The MultiPDF Chat App is a Python tool that lets you ask questions in natural language and receive answers based on the content of multiple PDF documents.

## Skills

---

**Programming:**

Python (AI/ML pipelines, backend development), GO, JavaScript, React, Bash

**Cloud Platforms:**

AWS (S3, EC2, EKS, API Gateway, Lambda), Google Cloud Platform (GCP) (Vertex AI, VM, BigQuery)

**Data Tools:**

SQL, NoSQL, Pandas, bigquery, dbt

**AI & Machine Learning Tools:**

Scikit-learn, TensorFlow and PyTorch, Hugging Face, LangChain

**Orchestration & Automation:**

Docker, Kubernetes, GitHub Actions

**Additional Skills:**

LLM finetuning, Generative AI system design, AI ethics and security practices, RLHF frameworks, LanGraph

## Languages

---

**English:**

C1 Proficiency (IELTS Band 8.0)

**Amharic:**

Native

## Certificates

---

**AWS Solutions Architect Associate** 

Demonstrated expertise in designing and implementing scalable, secure, and cost-efficient cloud architectures using AWS services.

**ISC2 Certified in Cybersecurity** 

Validated proficiency in securing systems, networks, and applications, focusing on industry-standard cybersecurity best practices.

**Google Data Analytics** 

Developed skills in data cleaning, visualization, and analysis to extract actionable insights and support data-driven decision-making.