

ASHTONE ONYANGO OMONDI

Senior AI Engineer

Nairobi, Kenya | Tel: +254740497975

ashtoneonyango@gmail.com | [LinkedIn](#) | [GitHub](#) | [My Portfolio](#)

PROFESSIONAL SUMMARY

Results-driven **Senior AI Software Engineer** with 5+ years of experience delivering scalable, production-grade AI systems that drive measurable business impact. Proficient in **Python, LLMs, semantic search, MLOps**, and **cloud AI deployment** using AWS SageMaker and Lambda. Skilled in **GraphQL API integration** and **vector database optimization**, achieving up to **85% improvement in customer satisfaction**, **70% reduction in operational backlog**, and **over KES 2M in cost savings** through intelligent automation across healthcare, real estate, and fintech sectors.

SKILLS & COMPETENCIES

- **Programming:** Python, JavaScript (Node.js, React, Next.js), SQL.
 - **AI/ML Frameworks:** PyTorch, TensorFlow, LangChain, HuggingFace, CrewAI.
 - **NLP & LLMs:** OpenAI GPT-4/o1, Claude, Gemini, RAG Systems, Fine-tuning.
 - **Cloud & MLOps:** AWS SageMaker, Lambda, Azure OpenAI, GCP Vertex AI, CI/CD, Docker, Kubernetes.
 - **Databases:** PostgreSQL, MongoDB, Pinecone, FAISS, Redis.
 - **Data Pipelines:** Apache Spark, Kafka, Airbyte, BigQuery.
 - **API Development:** GraphQL (Graphene, Apollo), FastAPI, Express.js.
 - **Versioning & Deployment:** GitHub Actions, AWS CodePipeline, Jenkins.
-

WORK EXPERIENCE

Wan AI Labs, Nairobi, Kenya

Jun 2023 – Present

Senior AI Engineer

- Led the **AI division** of Wan AI Labs, an AI Automation Agency providing **24/7 intelligent business agents** across industries such as **healthcare, legal, finance, and enterprise operations**.
- Designed and deployed **20+ production-grade AI systems** using **LLMs (GPT-4, Claude, Gemini), LangChain and CrewAI**, automating support, scheduling, and data processing workflows - resulting in **40% higher productivity** and a **60% reduction in manual tasks**.
- Led cross-functional teams of engineers and designers to deliver tailored AI workflows for **government and enterprise clients**, contributing to an estimated **KES 2M+ in cost savings** through intelligent automation.
- Oversaw **integration with legacy enterprise systems** via **GraphQL APIs, Kafka pipelines, and REST services**, ensuring smooth adoption of AI across complex IT environments.

Rightsify Group LLC, Pasadena, California.

June 2023 – Aug 2024

AI Software Engineer

Architected scalable AI music generation systems serving 3M+ daily users, leveraging LLMs and transformer models for copyright-cleared content creation.

- Designed and deployed distributed training pipelines on GCP that reduced model training time by 60% while processing 1M+ songs through custom data preprocessing workflows
- Fine-tuned transformer models using PyTorch and HuggingFace, achieving 20% performance improvement through systematic hyperparameter optimization and validation
- Built production GraphQL APIs with React frontends, implementing semantic search capabilities that improved content discovery accuracy by 35%
- Orchestrated automated AWS S3 to GCP migrations for 10TB+ datasets, establishing scalable MLOps infrastructure for model serving across Paperspace and Digital Ocean environments
- Implemented CI/CD pipelines for ML models using Docker and Kubernetes, reducing deployment time from 4 hours to 15 minutes

ZURI HEALTH, Nairobi, Kenya.

Jan 2023 – April 2023

Software Developer

Developed AI-powered healthcare triage system serving 5,000+ users across Sub-Saharan Africa through intelligent symptom analysis and conversational AI.

- Built conversational AI chatbot using NLP and RAG architecture, reducing query resolution time by 30% while maintaining 92% diagnostic accuracy
- Implemented Graphene-based GraphQL server with React/Apollo Client frontend, processing 10,000+ patient interactions monthly
- Enhanced chatbot intelligence through semantic search integration with medical knowledge bases, improving response relevance by 40%
- Collaborated with platform engineers to design analytics pipelines that extracted actionable insights from multi-country user interaction data

Google Crowdsourcing, Nairobi, Kenya.

Aug 2021 – Present

AI Facilitator/Trainer - Part-time

Community leadership role promoting responsible AI development and ML education across Africa.

- Facilitated 11+ hands-on ML workshops for 33+ university students across African countries, covering LLM fine-tuning, NLP, and cloud ML deployment
- Managed community of 5,000+ AI contributors, coordinating open-source initiatives focused on ethical AI development and accessibility
- Trained participants on AWS SageMaker and GCP Vertex AI for deploying production ML models

ANALYTICS VIDHYA, Gurgaon, India

Sep 2022 – Dec 2022

Technical Writer – Data (Seasonal)

- Wrote and published 3 tutorial articles for the monthly [data science blogathon](#) on creating ETL pipelines for over 2000 rows of data extracted from OLAP and OLTP databases, which attracted over 50 readers and users to the platform.
- Implemented 2 application backends using Flask to integrate database APIs that enabled interaction with remote Postgres database servers hosted both locally and on the cloud.

Upwork, Nairobi, Kenya Feb 2021 – Sep 2022

Machine Learning Engineer - Freelance

- Delivered 33+ custom AI solutions leveraging AWS and Azure ML services, implementing semantic search and NLP systems for diverse client needs
- Built production-ready GraphQL APIs with Express and React frontends, integrating LLM capabilities for intelligent data processing
- Developed distributed data processing pipelines using Apache Spark that improved data ingestion speed by 65%
- Created automated reporting systems in Tableau that saved clients 10+ hours weekly through intelligent data visualization

TEENS IN AI, London, UK.

Machine Learning Trainer

Aug 2019 – Sep 2019

- Coached 5 teen girls out of 10 teams in Descriptive and Predictive Analytics using Azure Machine Learning visual tools.
- Guided approximately 10 students on using Azure Machine Learning to analyze access to affordable healthcare by low-income households in 5 countries across Sub Sahara Africa.

EDUCATION

KENYATTA UNIVERSITY, Nairobi, Kenya

Bachelor of Science, Biomedical Engineering

Relevant Coursework:

- Health Information Systems, Bioinformatics, Programming; Electronics; Database Systems.

PROJECTS

Pneumonia Detection & Monitoring in Infants

[Huawei Global Innovator Award](#) - Johannesburg, South Africa - April 2022

- Collaborated with a team of 3 to design an IoT wearable device to collect data from 3 body sensors and communicate the data, then trained a Neural Network model on over 5000 rows of data and predicted the risk of a child getting pneumonia, with an accuracy of 80%.

Kujia Jobs Mtandao - [AI-Powered Job Matching & Career Platform](#)

- Architected semantic search engine using vector embeddings and LLMs to match 5,000+ job seekers with 10,000+ opportunities across 100+ employers, achieving 90% resume parsing accuracy through custom NLP models
- Deployed AWS Lambda-based matching service processing 1,000+ daily queries with collaborative filtering algorithms that increased application success rates by 40%

Chesapeake Stays – [AI-Powered Real Estate Lead Qualification](#)

- Implemented RAG system processing 10,000+ property documents daily with 94% extraction accuracy using fine-tuned LLMs and vector databases
- Deployed AWS Lambda-based semantic search infrastructure that reduced client acquisition time by 3x and generated 5.2x ROI increase while saving 40+ hours monthly

Ruphids Autotech Solutions – [Computer Vision Biometric Security](#)

- Deployed production computer vision system for biometric verification serving 25+ schools with 99.8% accuracy and 1-second processing time, handling 1,000+ daily verifications
- Built real-time AI inference pipeline that reduced manual attendance processing by 80% with integrated SMS notifications for parents

ACTIVITIES

Google Developers Student Club, Nairobi, Kenya

Machine Learning Lead

Sep 2020 – Aug 2021

- Organized and advertised over 15 Machine Learning workshops that saw over 100 students enroll to learn data science.
- Led 3 Machine Learning competitions for over 40 students, increasing their Cloud Analytics skills in GCP and TensorFlow.

KamiLimu, Nairobi, Kenya

Mentee

Aug 2019 – June 2020

- Selected among 36 mentees from 18 Kenyan universities in a structured tech mentorship program that grew my ICT, public speaking, and Innovation skills after 8 months.