# Samwel Waweru

✉ samuelwawerukamau01@gmail.com    📞 +254792022398    in samuelwaweru2001    ⌥ samkamau81

## PROFESSIONAL SUMMARY

Senior AI Engineer with over **4 years** of hands-on experience building production-grade, cloud-native AI and automation systems. Highly skilled in Python 3, REST API design, and distributed backend architectures, with a proven track record deploying GenAI and on-device ML solutions across Docker- and Kubernetes-based environments. Experienced in designing intelligent agents, optimizing model performance, and integrating automation workflows at scale. Published ML researcher with international conference presentations and a strong engineering foundation that bridges AI, systems design, and real-world product delivery.

## PROFESSIONAL EXPERIENCE

**Machine Learning Engineer (Kaggle BIPOC), Kaggle**                    **USA, California (Remote)**
Dec 2022 – Present

- Developed and maintained production-grade LLM integrations (Python, REST APIs) including a RAG-powered analytics engine using FAISS and Pinecone, ensuring scalable retrieval workflows, strong observability, and stable API contracts.
- Built and iterated on an AI agent development platform (Python, JavaScript) enabling secure, reliable deployment of multi-agent LLM workflows to embedded IoT systems (ESP8266). Designed modular backend components with a focus on maintainability and backward-compatible API evolution.
- Fine-tuned and optimized Gemma-7B using LoRA, PEFT, and quantization techniques to deliver a lightweight, latency-efficient LLM microservice for cybersecurity Q&A, applying performance-minded debugging, profiling, and SQL-backed data evaluation.
- Diagnosed and resolved production issues across model APIs, background jobs, and data pipelines, contributing to system reliability through root-cause analysis, test refinement, and iterative quality improvements.
- Collaborated asynchronously with globally distributed engineering teams, providing architectural guidance, conducting code reviews, and mentoring teams on best practices in LLMOps, API design, and secure model integration.

**AI Solutions Expert, EdTech Hub, Ministry of Education — Consultant**          **Nairobi, Kenya**
Oct 2025 – Present

- Engineered and shipped AI-powered backend services (Python, REST APIs) for education platforms, integrating NLP and computer vision models to improve data processing, system reliability, and user experience at national scale.
- Led the design and deployment of secure LLM integrations within Ministry systems, ensuring compliance with governance, data protection, and safe model interaction — echoing GitLab's secure AI model integration requirements.
- Worked cross-functionally with software engineers, product teams, and educators in an async-first, distributed environment to deliver end-to-end AI solutions, from scoping and architecture to testing and rollout.
- Built and validated multiple proof-of-concept AI services, optimizing model pipelines, refining data flows, and iterating on features using feedback from stakeholders and real-world usage data.

## EDUCATION

**Jomo Kenyatta University of Agriculture and Technology**                    **Nairobi, Kenya**
B.Sc. Mechatronic Engineering, Robotics - Honors (2:1)

- **Capstone Project (Rank 1/18):** Designed a Delta Laser-Weeding Robot with computer vision, real-time inference (Raspberry Pi + YOLOv11), and distributed microcontroller orchestration.
- Took courses on Calculus 1- 4, Advanced Robotics, Differential Equations, and Computer Science.

- Built and deployed NLP models using Hugging Face Transformers (BERT) for sentiment analysis of Twitter data via API integration (Python).
- **Hackathon Winner:** Authored award-winning report on "Effects of COVID-19 vis-à-vis Mental Health", applying data-driven modeling and analytics to a real-world challenge.

# PUBLISHED PAPERS

- **Samwel Waweru (Lead Author)**, Ray Njoroge. *Edge-Optimized YOLOv11-Based Real-Time Weed Detection for Autonomous Precision Agriculture in Resource-Constrained Environments.* Accepted to the **IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2025)**, Hangzhou, Zhejiang, China (OralPoster).
- Jackson Mutembei, **Samwel Waweru (Contributor)**. *A Biodiversity Conservation Protocol in Meru County for Fire Prevention using Drone Technology.* Accepted at the Forum for Global Challenges, University of Birmingham, May 2022 (Highlighted).

# ENGINEERING PROJECTS

- **FinGPT** — **Financial Language Model System** Implemented a backend service for a domain-specific financial LLM using **FastAPI** with well-typed **Pydantic** models and clean request/response schemas. Added input validation, rate management, and caching to improve reliability of real-time inference. Integrated the generation pipeline with financial text-processing modules and designed REST endpoints focused on maintainability and backward-compatible response formats. (*Key Tech: FastAPI, Pydantic, Python, REST API Design, Async I/O, Financial Text Processing.*) Link
- **Delta Laser Weeding Robot** (Computer Vision + Robotics Backend) Built a computer vision inference service around **YOLOv11** running on Raspberry Pi 4, optimizing it using INT8 quantization and I/O pipeline tuning to reduce processing latency. Designed Python modules that expose detection results through structured interfaces consumed by the robot's motion controllers. Debugged and maintained data flow between the CV service, G-code generator, and microcontroller firmware to ensure stable field operation. (*Key Tech: Python, YOLOv11, TensorFlow, OpenCV, REST-like Module Interfaces, Raspberry Pi, Microcontroller Integration.*) Link
- **EmbedMind Studio** Developed a lightweight platform for assembling and deploying LLM-powered agents on embedded devices. Built the Python backend that orchestrates agent logic, integrates with **Gemini and OpenAI APIs**, and manages device communication over serial/WiFi. Used **Docker** to create reproducible development and simulation environments, and implemented CI/CD workflows to ensure consistent builds across devices. Added logging and basic monitoring to simplify debugging distributed device deployments. (*Key Tech: Python, C++, Docker, CI/CD Workflows, Gemini/OpenAI APIs, Embedded Networking (ESP8266, Raspberry Pi).*) Link
- **MentalMuse** Created a secure NLP analysis pipeline for mental-health text. Implemented scalable data-processing services using Python and orchestrated batch/stream workloads using **Snowpark**. Built API integrations to route users to mental-health support services and added lightweight explainability features around NLP outputs. Focused on stable data handling, schema consistency, and secure text processing. (*Key Tech: Python, NLP Pipelines, Snowpark (Snowflake), Secure API Integration, Data Validation.*) Link
- **Kilimo-AI** Built a simple LLM-assisted irrigation control system for IoT hardware. Created Python–C++ modules to process user natural-language input, apply prompt-engineering logic, and translate decisions into device-level irrigation commands. Integrated cloud REST APIs for device state updates and alerts. Designed the system with low-power hardware constraints in mind to ensure stable operation in rural settings. (*Key Tech: Python, C++, LLM Prompt Engineering, IoT Controllers, REST APIs, Cloud Integration.*) Link

# PROFESSIONAL CERTIFICATION

**Oracle Cloud Infrastructure Certified Data Science Professional** (Credential ID OC4504520). Recognizes ability to design, train, deploy, and manage machine learning models using OCI's Data and AI services.

# COMMUNITY INVOLVEMENT

**County Coordinator/Master Trainer, Raspberry Pi Foundation UK, Young Scientist Kenya
United Kingdom, Kenya (Hybrid)**
Apr 2024 – Present

- Led the establishing 10 Code Clubs focused on Python and AI development training for over 250 students, while simultaneously training and evaluating mentors on pedagogy and project design.
- Acted as liaison between government, educators and nonprofits, ensuring alignment of digital learning initiatives with local education priorities.

# SKILLS AND INTERESTS

- **Programming & Tools:** Python, JavaScript, C++, FastAPI, REST API Design, GraphQL (basic), Pydantic, SQL, PostgreSQL, Docker, CI/CD Workflows, Git, Kubernetes (basic), TensorFlow, Hugging Face, LangChain, Snowflake/Snowpark, Embedded Systems (ESP8266/Raspberry Pi), Linux.
- **AI & LLM Engineering:** LLM integration (OpenAI, Gemini), prompt design, retrieval-augmented generation (RAG), model optimization (quantization, LoRA, PEFT), latency/performance tuning, secure text-processing pipelines.
- **Backend & Systems:** API schema design, background job orchestration, asynchronous I/O, debugging and profiling, containerized services, monitoring/logging, reliability engineering, distributed development workflows.
- **Core Strengths:** Proactive problem solver, strong communicator, collaborative team player, adaptable in fast-paced Agile environments, continuously learning.
- **Affiliations:** IEEE Robotics and Automation, Raspberry Pi Foundation UK, Kenya Karate Federation.
- **Language:** English (Fluent), Kiswahili (Native).