

# Samuel Wanyoro Wanyua

Nairobi, Kenya | (+254) 745 74 34 74 | samuelwanyua@gmail.com

[linkedin.com/in/samwanyua](https://www.linkedin.com/in/samwanyua) | [github.com/samwanyua](https://github.com/samwanyua) | [behance.net/wanyuasam](https://www.behance.net/wanyuasam)

## PROFESSIONAL SUMMARY

---

AI Engineer and Software Engineer with hands-on experience building inclusive speech recognition systems, predictive analytics solutions, and cloud-deployed machine learning applications. Proficient in deep learning, NLP, ASR, and modern data engineering workflows. Strong track record of working across multidisciplinary teams to deliver scalable ML pipelines and accessible, user-centered AI products. Currently building Sauticare — an ASR platform for learners with speech impairments — while maintaining freelance work on bilingual ASR tutors and real estate ML systems deployed on Hugging Face Spaces.

## WORK EXPERIENCE

---

**Sauticare** — Nairobi, Kenya | July 2025 – Present

### Founder

- Tuned the ASR pipeline to respond in under 500ms, with solid support for Kenyan English.
- Built a pronunciation scoring system using WER and CER so learners get real, measurable feedback on their progress.
- Wired up a FastAPI + Supabase backend handling lesson delivery, progress tracking, and automated milestone rewards.
- Shipped a Next.js 15 frontend with gamified achievements that pushed simulated learner engagement up by 25%.

**NW Realite** — Nairobi, Kenya | Nov 2025 – Feb 2026

### Data Analytics Engineer

- Designed OCR and RAG-based pipelines to extract structured data from unstructured property valuation reports, enabling automated ingestion of 100s of documents previously handled manually.
- Built scalable ETL workflows using Python and loaded transformed data into Supabase databases, reducing data preparation time and improving downstream query performance.
- Developed Power BI dashboards supporting predictive and prescriptive analytics, giving executives real-time visibility into property market trends.
- Translated complex analytics outputs into executive-level insights and business strategy recommendations, bridging the gap between technical findings and decision-making.

**Freelance** — Remote | Jan 2025 – Present

### Software Engineer & AI Engineer

- Designed and developed an end-to-end bilingual ASR speech tutor for Kenyan English and Swahili using Whisper and Wav2Vec2, targeting accessibility for learners with speech impairments.
- Optimized WER/CER metrics and inference latency for offline AI deployment in low-resource, low-connectivity environments suitable for rural Kenyan classrooms.
- Built and deployed HemaHaus, a real estate price prediction system achieving 91.9% accuracy across 190K+ listings using Scikit-learn, FastAPI, Streamlit, and Hugging Face Spaces.
- Engineered early-stage RAG pipelines using embeddings, vector databases, and chunking strategies to enable retrieval-based insights from unstructured property documents.
- Managed full ML project lifecycle — problem scoping, data acquisition, experimentation, deployment, and user feedback integration — across multiple concurrent client projects.

**Innovex Solution Limited** — Nairobi, Kenya | Aug 2024 – Jan 2025

### Software Developer

- Built insurance management software using Next.js, React, TypeScript, Redux, and Tailwind CSS in an agile team environment.
- Improved client operational efficiency by 10% through frontend architecture optimization and reusable component design.
- Refactored legacy codebases, resolved complex UI/UX issues, and improved cross-device responsiveness.

[CodePerfect Solutions](#) — Remote (Part-time) | May 2025 – Present

### Technical Writer — Data & AI

- Write in-depth technical articles on data engineering and AI engineering topics including ETL workflows, model deployment, and MLOps.

## PROJECTS

---

### HemaHaus — Real Estate Price Prediction System

Technologies: Scikit-learn, FastAPI, Streamlit, Hugging Face Spaces

- Achieved 91.9% prediction accuracy across a dataset of 190K+ property listings through rigorous feature engineering and model selection.
- Built full ML infrastructure: data cleaning, feature engineering, model evaluation, REST API serving via FastAPI, and interactive Streamlit dashboards.
- Deployed live on Hugging Face Spaces — accessible at [datascienceportfol.io/samuelwanyua](https://datascienceportfol.io/samuelwanyua).
- Led data mining of unstructured real-estate documents and engineered extraction workflows for downstream valuation analysis.

### Inclusive ASR Speech Tutor (Offline, Bilingual)

Technologies: Whisper, Wav2Vec2, PyTorch, Transformers, Hugging Face

- Fine-tuned ASR models for impaired and non-standard speech patterns common in Kenyan learners, significantly improving transcription accuracy over baseline models.
- Built real-time STT–TTS feedback loop with latency optimization for offline deployment on low-end devices in low-connectivity classrooms.
- Implemented adaptive pronunciation scoring using custom WER/CER evaluation metrics and personalized lesson difficulty adjustment.
- Applied human-centered co-design methodology with educators and caregivers to improve usability, cultural relevance, and accessibility.
- Designed modular ASR–TTS–UI architecture enabling future model updates and scalable data expansion.

## EDUCATION

---

### Software Engineering — Distinction | Nov 2023 – June 2024

Moringa School, Nairobi, Kenya

### Bachelor of Construction Management — Second Class Honours (Upper Division) | Sept 2019 – Sept 2023

University of Nairobi

## TECHNICAL SKILLS

---

**Programming:** Python, JavaScript, TypeScript, SQL

**Machine Learning & AI:** PyTorch, TensorFlow, Scikit-learn, XGBoost, Transformers, Whisper, Wav2Vec2, NLP, Deep Learning, LLMs, ASR

**MLOps & Deployment:** Docker, FastAPI, CI/CD, GitHub Actions, Hugging Face Spaces, GCP, Vercel, Railway, Heroku

**Analytics Engineering:** ETL Pipelines, OCR, Data Modeling, Supabase, Feature Engineering, WER/CER, Fine-tuning

**Data Visualization:** Power BI, Looker Studio, Plotly, Matplotlib, Seaborn, Streamlit

**RAG & Data Mining:** Vector Databases, Embeddings, Chunking Strategies, Retrieval Pipelines

**Developer Tools:** Git, GitHub, Linux CLI, Jupyter, VS Code, PyCharm, Postman

## LANGUAGES

---

Kikuyu (Native) | English (C2 — Proficient) | Swahili (C2 — Proficient)