Bernard Abuto Otieno

Machine Learning Research Engineer | AI Systems Engineer
Kenya · 0116270944 · benardabuto081@gmail.com · bernardabuto.vercel.app
github.com/benardabuto081 · linkedin.com/in/bernard-abuto-888207343

SUMMARY

Machine Learning Research Engineer and AI Systems Engineer working across applied machine learning, research engineering, and software systems development. Experience spans dataset engineering, experimental design, model evaluation, speech recognition, natural language processing, geospatial analysis, and full-stack application development. Builds research-driven systems that connect data, machine learning technologies, software architecture, and practical applications. Particular research interests include low-resource machine learning, multilingual AI, African speech and language technologies, model generalisation, and reliable AI systems.

EXPERIENCE

Founder, CEO & Chief Research Officer | Machine Learning Research Engineer | AI Systems Engineer, Sauti Labs (sautilabs.co.ke) Jun 2026 – Present

- Founded and lead an AI research and engineering initiative focused on African speech and language technologies, multilingual AI, and low-resource machine learning.
- Direct technical strategy and research priorities across speech intelligence, language intelligence, language resources, model intelligence, and AI infrastructure.
- Design a framework-first architecture for managing African language datasets, speech resources, metadata, provenance, validation, and cross-language resource discovery.
- Develop reusable infrastructure components, registries, SDK concepts, and engineering conventions connecting language resources with downstream machine learning workflows.
- Engineer dataset ingestion and registration workflows, including integration and validation of Common Voice Swahili resources.
- Design licensing and provenance mechanisms that preserve third-party dataset licensing conditions while maintaining independent licensing for Sauti Labs infrastructure.
- Establish automated repository, licensing, and structural validation within CI/CD workflows.
- Lead research and engineering programmes involving dataset preparation, preprocessing, experimental design, benchmarking, model evaluation, and reproducibility.
- Coordinate technical development across machine learning, speech technology, language resources, software infrastructure, and product engineering.

SELECTED PROJECTS

QueleaGuard — Machine Learning Research & Environmental Modelling 2026 – Present
Python, pandas, GeoPandas, Rasterio, xarray, scikit-learn · github.com/benardabuto081/QueleaGuard

- Designed a geospatial research workflow covering a 50 km study extent, divided into a 5.5 km grid comprising 328 spatial cells.
- Prepared a spatial dataset containing 133 presence records and 133 pseudo-absence records.
- Integrated environmental and geospatial variables from climate, vegetation, elevation, and hydrological datasets.
- Processed remote-sensing and environmental data, addressing issues including GBIF pagination, temporal sampling bias, and satellite NDVI fill values.
- Developed a baseline modelling and validation workflow, with attention to spatial cross-validation and the risk of spatial data leakage.
- Investigate model performance, spatial generalisation, and interpretability for ecological occurrence estimation.

Scribe AI — AI Systems & Desktop Application Engineering 2026
Python, Faster-Whisper, SQLite, Tauri 2, React, TypeScript, Rust, Tailwind CSS, Zustand, Vitest · github.com/benardabuto081/-scribe-ai

- Developed a desktop application foundation using Tauri 2, React 18, TypeScript, Tailwind CSS, Zustand, and Rust.
- Implemented a typed command-bus architecture supporting application actions, validation, permissions, undo operations, audit logging, keyboard shortcuts, and AI action gating.
- Designed independent Audience and Stage output windows, with on-air state managed in the Rust layer to improve output resilience.
- Built a local Python speech-processing prototype using Faster-Whisper for offline transcription and rule-based detection of Bible references.
- Integrated SQLite-based Bible lookup containing approximately 31,100 verses and a human approval workflow before detected scripture is displayed.
- Developed lyric-matching and service-state detection components, with integration into the live audio workflow remaining in progress.
- Established automated frontend and Rust testing, including 95 frontend tests and 5 Rust tests reported by the project documentation.

Rentra — Full-Stack Product Engineering 2026
Flutter, Dart, NestJS, TypeScript, PostgreSQL, Supabase, TypeORM, JWT, Africa's Talking, Resend · github.com/benardabuto081/Rentra

- Develop a multi-layer application architecture comprising a Flutter mobile client and a NestJS backend connected to PostgreSQL through Supabase and TypeORM.

- Implement backend domain modules covering authentication, users, property portfolios, buildings, rental units, tenancy records, payments, maintenance, dashboards, and independent tenant relationships.
- Build authentication workflows supporting landlord registration, tenant registration, phone- and email-based login, OTP verification, and landlord-issued tenant passkeys.
- Implement JWT-based authentication and role-based access controls using NestJS guards and role decorators.
- Develop mobile application onboarding and tenant workflows using Flutter and Dart, including account creation, verification, role selection, setup, navigation, and account management.
- Integrate external communication services for OTP delivery through Africa's Talking and Resend.
- Structure the product around managed and independent tenants, allowing rental relationships to be represented even when a landlord is not yet registered on the platform.

EDUCATION & TRAINING

Bachelor of Science in Computer Science, The Open University of Kenya — Sep 2026 – Expected Apr 2030
Software Engineering Programme, Zone01 Kisumu — Sep 2026 – Expected Feb 2028
Data Science & AI Bootcamp — Distinction, Ngao Labs — Jun 2026 – Aug 2026
AI & Software Engineering Scholarship Programme, Power Learn Project — May 2026 – Present
Bachelor of Science in Computer Science — Completed First Academic Year, Chuka University — Sep 2025 – Apr 2026

TECHNICAL SKILLS

Machine Learning & Research: Applied Machine Learning, Supervised Learning, Classification, Regression, Feature Engineering, Data Preprocessing, Model Evaluation, Experimental Design, Dataset Engineering, Research Methodology, Reproducible Experimentation
AI & Language Technologies: Natural Language Processing, Automatic Speech Recognition, Speech Processing, Faster-Whisper, AI Integration, Offline AI Inference, Multilingual AI
Programming Languages: Python, Go, TypeScript, JavaScript, Dart, SQL
Software & Systems Engineering: React, Tauri, Rust, Flutter, NestJS, REST API Development, Backend Architecture, Application Architecture, JWT Authentication, Role-Based Access Control
Data & Infrastructure: PostgreSQL, SQLite, Supabase, TypeORM, Git, GitHub, Linux, CI/CD, Automated Testing, Data Validation
Geospatial & Environmental Analysis: Geospatial Data Processing, Spatial Analysis, Remote Sensing Data, Environmental Data Integration, Spatial Validation