

GEVVERSON KIPKOECH

AI SOFTWARE ENGINEER · FULL-STACK ENGINEER · LLM & AUTOMATION

Nairobi, Kenya (Open to Global Remote) · +254 724 329 386 · gevversonkipkoech@gmail.com
linkedin.com/in/gevverson-kipkoech · github.com/gevverson · gevverson.vercel.app

PROFESSIONAL SUMMARY

AI & full-stack software engineer who ships production-grade products end-to-end — RAG/LLM platforms, mobile healthtech, and trading/fintech systems — from first commit to live deployment. Strong across the modern AI stack (RAG pipelines, vector retrieval, prompt engineering, structured output, and LLM evaluation) and full-stack delivery (Python · FastAPI/Flask, NestJS/Node, React/Next.js, Flutter, PostgreSQL, Docker). Comfortable owning a product solo and leading a small team, with a grounding in Computer Science and Statistics.

CORE COMPETENCIES

AI / LLM	RAG pipelines, vector retrieval (Pinecone, ChromaDB), prompt engineering, structured output, LLM evaluation & red-teaming, Google Gemini, OpenAI & Anthropic Claude APIs, LangChain
Backend / APIs	Python (FastAPI, Flask, Django), Node.js (NestJS, Express), REST API design, SQLAlchemy, Prisma, Celery, JWT / OAuth2
Frontend	React, Next.js, TypeScript, Tailwind CSS, Flutter / Dart
Data	PostgreSQL, TimescaleDB, MongoDB, Redis, Pandas / NumPy, time-series analysis
DevOps / Cloud	Docker, Docker Compose, GitHub Actions (CI/CD), Fly.io, Render, Vercel, Linux, Nginx
Security	Field-level encryption, rate limiting, CSRF, CSP / Talisman, audit logging, PCI-aligned checkout

PROFESSIONAL EXPERIENCE

Lead AI Engineer | [Helix](#) 2026 – Present · Remote

- Building a **genetic-compatibility dating platform** end-to-end: users store an encrypted genotype/blood profile (typed or OCR-scanned from a medical report) and match with a sickle-cell / rhesus **traffic-light verdict** that gates whether matched photos un-blur — framed as educational, not diagnostic.
- FastAPI + PostgreSQL backend (SQLAlchemy 2.0, 13 Alembic migrations, **27 API endpoints, ~150 tests**) with a **15-pair genotype engine** explaining per-child inheritance risk as GREEN/AMBER/RED.
- Privacy-by-design: field-level **Fernet (AES) encryption** of genetic data, pseudonymisation behind a random subject-ref, JWT + bcrypt + Google OAuth, per-route rate limiting, and a right-to-erasure cascade (GDPR / Kenya-DPA).
- Flutter app (16 screens, go_router auth guards, FCM push), Tesseract OCR, and server-enforced photo-blur via signed Cloudinary URLs.

Software Engineer (AI & Agents) — Team Lead | [IntelliVerse](#) Nov 2025 – Jan 2026 · Remote

- Built and deployed a **multi-user platform where users create custom AI chat agents** — each with its own persona and uploaded document knowledge (RAG) — with agent and chat sharing and forking.
- Flask backend (app-factory, 4 blueprints, **21 routes, 49 tests**), PostgreSQL via SQLAlchemy + Alembic, Dockerised and deployed on **Fly.io**.
- Document-RAG pipeline: PDF/DOCX/TXT → chunking → Google Gemini embeddings → **Pinecone** vector retrieval grounding each agent's answers.
- Hardened the platform: Google OAuth + password auth with account lockout, CSRF, Talisman CSP/HTTPS, per-route rate limiting, audit logging, and a **Gemini model-fallback chain** for resilience.
- Led a **small team (2–3 including me)**: set sprint goals, ran code reviews, and established the Docker-based dev environment.

LLM Trainer & Quality Evaluator (Contract) | [Turing](#) Dec 2025 – Feb 2026 · Remote

- Designed and executed **evaluation frameworks** for frontier LLMs on agentic / computer-using tasks — assessing reasoning quality, execution reliability, and edge-case robustness.
- Applied **RLHF-based evaluation and data annotation** to surface reasoning failures and instruction-following gaps that fed model-improvement iterations.
- Produced red-teaming reports and prompt-failure taxonomies across multiple experimental checkpoints.

Lead Developer | Constituency Tech Initiatives

2025 – Present · Nairobi, KE

- Built **Macho ya Mwananchi**, a civic-accountability platform surfacing ghost projects and misused public funds from official **Auditor-General** data (**Ksh 3.3B tracked, 144 facilities, 1,885 ghost workers**).
- RAG pipeline (ChromaDB + Gemini 2.5 Flash) for natural-language Q&A over audit documents in Swahili / English; GPS citizen-verification consensus engine; anonymous corruption reporting.
- React/Vite + Node/Express + MongoDB Atlas + FastAPI, Dockerised with a GitHub Actions pipeline; live deployment.

Data Annotation Specialist | iMerit (Ango Hub)

2023 – 2024 · Remote

- Annotated complex multi-modal datasets and applied RLHF protocols to improve reasoning accuracy, output safety, and instruction adherence of large language models.
- Specialised in edge cases for reasoning chains, code-generation quality, and factual grounding — producing annotation guidelines later adopted across the team.

Data & Software Engineering Intern | Kenya National Bureau of Statistics (KNBS)

2023 · Nairobi, KE

- Engineered automated Python (Pandas / NumPy) data pipelines to clean, validate, and transform large demographic datasets — reducing manual review time and data-inconsistency errors.
- Built internal Flask + JavaScript dashboards for interactive geographic data visualisation across county-level data.

KEY PROJECTS

AI Trading Bot — self-hosted quant research harness

OPEN SOURCE

Python · FastAPI · TimescaleDB · Redis · ChromaDB · Docker Compose · MT5 (demo)

- A **9-gate signal + risk pipeline** (news blackout, drawdown circuit-breaker, Kalman trend, multi-timeframe confluence, VWAP order-flow, HMM regime, conviction sizing) over TimescaleDB time-series.
- 7 quant-math modules** — Kalman filter, 3-state HMM regime detection, Bayesian win-probability, half-Kelly sizing with a 5% cap and negative-expectancy rejection.
- ~5.1k backend LOC across 4 Docker services, 23 API endpoints, an MT5 demo bridge, and Telegram alerts. **Paper-trading by default** — personal / educational, no live-money or performance claims.

PSP Platform — PayCore — multi-country payment orchestration

TEAM · CORE CONTRIBUTOR

TypeScript · NestJS · Prisma · PostgreSQL · pnpm workspaces · Vitest · Docker

- Routes mobile money (MTN MoMo, M-Pesa) and card payments through one provider-agnostic core. Built the **payment-rail adapters** and **hosted tokenised checkout** (PCI-aligned — card data never touches the server).
- Signature-verified provider webhooks and idempotent transaction handling; live sandbox demos in a pnpm monorepo with a Vitest suite.

EDUCATION

B.Sc. Computer Science | University of the People

May 2026 – Present · In progress

B.Sc. Statistics and Programming | Kenyatta University

2020 – 2024