

STEFAN O. JUMA

Nairobi, Kenya
+254 707 322 993
stephenjuma500@gmail.com
linkedin.com/in/stefan-juma-a13489150
github.com/steph500

SENIOR PRODUCT ENGINEER | FULL-STACK & APPLIED AI

7.5 yrs software engineering	10,000+ users supported	20+ production features	40% faster builds
--	-----------------------------------	-----------------------------------	-----------------------------

Senior product engineer with 7.5 years of experience building production software across Python, TypeScript, React/Angular, APIs, SQL/PostgreSQL, data-heavy systems and applied AI. Strong record of owning features end to end, debugging production issues, improving CI/CD and performance, and translating loosely defined product needs into maintainable systems. Recent work includes RAG, embeddings, pgvector, LLM-powered workflows, structured outputs and human approval controls in a multi-tenant AI operations platform.

CORE SKILLS

Backend & Data

Python, FastAPI, Node.js, Express/NestJS, Java/Spring Boot, REST, GraphQL, WebSockets, PostgreSQL, SQLAlchemy, Alembic, Redis, MongoDB, Supabase, pgvector

Frontend

TypeScript, React, Next.js, Angular, Ionic, React Native, Tailwind, RxJS, Redux, D3, Recharts

AI / LLM

OpenAI/GPT, Qwen, Ollama, RAG, embeddings, vector retrieval, structured outputs, prompt engineering, tool-oriented workflows, deterministic validation, local/cloud model routing

Cloud & Delivery

AWS, GCP, Microsoft Azure, Docker, Kubernetes, GitHub Actions, GitLab CI/CD, Linux, Vercel, Supabase, Firebase

Engineering Practice

Production debugging, code review, testing, migrations, RBAC/RLS, audit logging, idempotency, performance profiling, mentoring

EDUCATION

BSc Computer Science

Meru University of Science and Technology

EXPERIENCE

Full-Stack Software Engineer

Cell Collective / University of Nebraska-Lincoln | Jan 2023 - May 2026 | Remote

- Built and maintained production scientific software used by 10,000+ users in a distributed US research team.
- Delivered 20+ production features across React/TypeScript interfaces and Python/Node integrations, owning work from design through testing, release and maintenance.
- Worked on data-heavy biological modelling workflows, simulations, API integrations and interactive visualisation with D3/WebGL.
- Improved build performance by about 40% through frontend architecture and tooling changes, while maintaining automated test and review standards.
- Investigated production defects, performance issues and integration failures across frontend, backend and modelling services.

Founder & Independent Software Engineer

SelfAwareTech | Feb 2021 - Present | Part-time

- Own product discovery, architecture, engineering, deployment, analytics and support across multiple SaaS and mobile products.
- **Tasko:** built a multi-tenant AI operations platform with grounded department assistants, RAG/pgvector retrieval, persistent prepared work, OAuth/provider boundaries, RBAC/RLS, audit controls and local/cloud LLM routing.
- Tasko is being used by real trial users and has generated its first paid service revenue.
- **Nyumba Services:** built and launched a services marketplace that reached 1,000+ Play Store downloads and 120+ providers in its first month.
- Regularly turn loosely defined business problems into practical product decisions, ship quickly, measure usage and iterate from real user feedback.

Senior Software Developer

EnigmaScore | Feb 2019 - Nov 2022 | Nairobi

- Led full-stack delivery across Angular/TypeScript, Java/Spring Boot and PostgreSQL for production scoring and analytics products.
- Designed APIs, data models, authentication and reporting workflows; reviewed code, mentored engineers and supported deployment on AWS.
- Worked directly with business stakeholders to clarify requirements, break down ambiguous problems and deliver maintainable solutions.

SELECTED PRODUCT & AI WORK

Tasko - AI Operations Platform

Applied AI / SaaS | SelfAwareTech

- Designed an assistant-first workflow across Marketing, Sales, Support, Finance, Operations and HR, with scoped AI assistance tied to each department rather than a single generic chatbot.
- Built grounded retrieval across company knowledge and connected sources using PostgreSQL/pgvector, with tenant, role and department boundaries enforced through RLS and application permissions.
- Implemented generation controls, persistent prepared work, human approval boundaries, audit trails and provider-action restrictions so AI output could be useful without bypassing business controls.
- Integrated cloud and local model options, including OpenAI-compatible APIs and Qwen through Ollama, allowing different privacy/cost trade-offs.
- Built and validated the system with API, frontend and database test suites, production builds, migration checks and security controls.

Data / SQL / Production Engineering Highlights

- **Relational modelling:** designed PostgreSQL schemas and migrations for multi-tenant products with foreign keys, uniqueness constraints, role-scoped access, audit data and lifecycle state transitions.
- **Non-trivial SQL:** worked with joins, aggregations, ranked/filterable datasets, permission-aware queries, migration backfills and analytics/reporting queries rather than limiting database work to ORM CRUD.
- **Data integrity:** used idempotent workflows, transaction boundaries, validation and uniqueness constraints to avoid duplicated actions and inconsistent state during retries or integration failures.
- **Production support:** debugged distributed frontend/backend issues, API failures, stale state, build regressions and data inconsistencies across web, mobile and scientific systems.

ADDITIONAL PROJECTS

Haulio	Freight marketplace with Angular/Ionic, NestJS, PostgreSQL/PostGIS, route matching, pricing, quotes/bookings and realtime/GPS foundations.
Nyumba Services	React Native/Expo/Supabase marketplace with provider onboarding, approvals, messaging, search, analytics and growth tracking.
Stok	Offline-first inventory app using Angular/Ionic/Capacitor/Supabase with local persistence, sync and import/export workflows.

WORKING STYLE

- Comfortable working independently on ambiguous product problems, but escalate early when a decision affects architecture, security or product direction.
- Strong async communication from several years working remotely with distributed teams, including a US university research team.
- Use AI-assisted development heavily for implementation, tests, debugging and code review, while manually auditing generated code, running automated tests and checking security/data assumptions before merge.
- Available for remote collaboration with UK working hours from Kenya.