Adioz D Eshitemi

Chief Technology Officer | Principal Software Engineer | Java, Spring Boot, Rust, Go, Python | Production AI
Nairobi, Kenya (Remote) | contact@adioz.dev | +254720804060
linkedin.com/in/adioz-eshitemi-5a594473 | adioz.dev | github.com/adiozeshitemi | dev.to/adioz

PROFESSIONAL SUMMARY

Chief Technology Officer at BBR (Big Bold Red Agency) with 8 years of experience in backend systems, distributed platforms and production AI. Leads engineering for Baze (Safaricom platform), a video and music streaming service built on Java 21 and Spring Boot microservices, and built BBRaiN, the Rust AI engine that powers its search, recommendations and personalisation. Owns every system from design to support: payments that cannot double-charge, auditable revenue sharing, and models that are trained, evaluated and monitored in production. Believes AI should be as efficient as it is intelligent, so designs resource-aware AI: small, quantized models that meet tight latency budgets on CPU without wasted compute. Treats agents as production software, with typed tools, fail-closed guardrails and a named operator behind every change.

KEY ACHIEVEMENTS

BBRaiN | AI Engine powering Baze (Safaricom platform) | Rust, Axum, Burn
- **GPU-free inference:** Removed the GPU from production. Vision, embedding and ranking models run from a single Rust service on CPU, after distilling large video and audio teacher models into compact INT8 students.
- **Replaced paid services:** Took over search from Algolia and recommendations from Recombee on Baze, built to a p95 page latency target of 250 ms across stateless replicas.
- **Engineering quality:** About 200,000 lines of Rust backed by about 4,000 automated tests.

Baze Platform Modernisation | Safaricom streaming service | Java 21, Spring Boot 4
- **Platform upgrade:** Inherited Baze's Java 17 services with open bugs and security gaps, and led the move to Java 21 and Spring Boot 4, resolving breaking changes across every service.
- **Java 25 migration:** Leading the move to Java 25 (LTS) to remove race conditions and unpredictable concurrency issues, close security gaps, cut intermittent latency and stabilise the quality gates that run on deployment change requests.
- **Security:** Remediated more than 150 dependency CVEs across payment, catalog, revenue, notification and playback services.
- **Delivery volume:** More than 2,300 commits across Baze's Java services in one year.

Carrier Rail | Baze Payments | M-Pesa, Safaricom carrier billing
- **No double charges:** Designed idempotent charging with double-charge detection, built so a charge is never lost or applied twice.
- **Resilience:** Added a Redis Lua global rate limiter, per-MSISDN request coalescing and a durable Postgres reconciliation queue.

Baze Revenue Engine | Revenue-sharing service | Java 21, WebFlux, ClickHouse, Kafka
- **Built from scratch:** Splits subscription revenue between content providers, Safaricom and the platform, using play-minute usage shares.
- **Auditable payouts:** A provisional, finalized, paid and disputed workflow with a full audit trail.

Zone01 and Upwork | Go and Java backends
- **Service performance:** Cut inter-service overhead by 30% by moving services to Go and gRPC.
- **Reliability:** Held 99% uptime on Dockerized Go and Java services for international clients.

TECHNICAL SKILLS

Backend Engineering | Java, Rust, Go, Python
- **Java and Spring:** Java 25, Spring Boot 4, Spring WebFlux (reactive), R2DBC, Spring Security, JWT, OAuth2, Spring Cloud Config, Spring Cloud Stream, Flyway, Resilience4j, JUnit 5, Testcontainers.
- **Languages:** Rust (Tokio, Axum), Go (gRPC), Python, TypeScript, SQL.
- **Architecture:** Microservices, event-driven systems, REST, GraphQL, WebSockets.

Data and Streaming | Storage, search and real-time pipelines
- **Databases:** PostgreSQL, ClickHouse, MongoDB, SQLite, Redis.
- **Streaming:** Kafka, Spring Cloud Stream, reactive Kafka consumers.
- **Search:** Tantivy (full-text search), Qdrant (vector search), hybrid BM25 and semantic ranking.

Artificial Intelligence | Machine learning, MLOps and agents
- **ML engineering:** Burn, Candle, PyTorch, CLIP, Hugging Face (Transformers, tokenizers, Hub), embeddings; GPU-accelerated training, multimodal (video, audio, vision) models.
- **LLM engineering:** Fine-tuning with LoRA, QLoRA and PEFT; LLM serving with vLLM, TGI, NVIDIA Triton and Ollama; retrieval-augmented generation (RAG); local small language models (SLMs).
- **Recommender systems and search:** Personalised ranking, recommendations, user clustering, hybrid BM25 and semantic retrieval, content moderation models.
- **MLOps:** Feature engineering, model training, hyperparameter tuning (HPO), experiment tracking (MLflow, Weights & Biases), accuracy and F1 evaluation, fairness checks, model registry, A/B testing, inference, data drift and latency monitoring, signed model releases.
- **Resource-aware AI:** Multi-teacher distillation, quantization-aware training (QAT), INT8 quantization, truncatable embeddings (Matryoshka/MRL), CPU-only and on-device inference, zero-copy memory-mapped weights, SIMD scoring, latency and memory budgets.
- **Agentic AI:** Tool-using and ReAct agents, LangGraph, LlamaIndex, Model Context Protocol (MCP), OpenAI and Anthropic tool calling, agent evaluation, typed audited tools.
- **Working knowledge:** Cloud TPU, JAX, PyTorch DDP/FSDP, DeepSpeed, SageMaker, Vertex AI, Azure ML, CUDA.

Frontend | Web and WebAssembly
- **Frameworks:** Angular 22, Next.js, React, Astro, Leptos.
- **Languages and tools:** TypeScript, RxJS, Rust/WebAssembly, HTML5, SCSS.

Cloud, DevOps and Security | Delivery and protection
- **Cloud and infrastructure:** AWS, Terraform, Docker, Kubernetes.
- **CI/CD and observability:** Jenkins, GitLab CI, GitHub Actions, CI Gates, Prometheus, Grafana.
- **Security:** SonarQube, Snyk, Trivy, CVE remediation, RBAC, Ed25519 signed releases, fail-closed controls, Offensive Rust (exploit proof of concept, security tooling, low-level memory security).

Engineering Leadership | Architecture, delivery and people
- **Architecture:** Design reviews, technical roadmaps, build-versus-buy decisions.
- **Delivery:** Work planning & delegation, runbooks, deployment change management, quality gates.
- **Operations and people:** Incident response and root-cause analysis, technical communications, mentoring.

PROFESSIONAL EXPERIENCE

Chief Technology Officer | BBR (Big Bold Red Agency) | Nairobi, Kenya | Apr 2026 - Present
Accountable for technology across Baze (Safaricom platform) and BBR's AI product, BBRaiN.
- **Engineering leadership:** Set architecture and review designs across Baze and BBRaiN; plan and assign technical work to engineers; own all technical communications; work daily with Safaricom engineering teams on integration, releases and production support.
- **Platform modernisation:** Led the upgrade of Baze's Spring services to Spring Boot 4 and Java 21, resolving breaking changes across services, and remediated more than 150 dependency CVEs found by Snyk, Trivy and SonarQube.
- **Payments architecture:** Designed the Carrier Rail for M-Pesa (Daraja) and Safaricom carrier billing, with idempotent charging, rate limiting and a durable reconciliation queue.
- **Operational excellence:** Own runbooks, deployment change requests and incident response, including root-cause analyses; maintain the quality gates of performance, smoke and end-to-end test suites, SonarQube checks and CI/CD pipelines.
- **MLOps:** Run BBRaiN's model lifecycle in production: feature engineering, training and HPO, experiment tracking, accuracy, F1 and fairness evaluation, model registration, A/B tests, and monitoring of inference, data drift and latency.
- **Security and guardrails:** Maintain fail-closed content safety controls, per-feature licensing and custom Rust security tooling (Offensive Rust methods) that block adversarial manipulation and unauthorized feature execution.

Lead Engineer (AI and Backend) | BBR (Big Bold Red Agency) | Nairobi, Kenya | 2025 - Apr 2026
BBRaiN | AI engine for Baze (Safaricom platform) | Rust, Axum, Kafka, Tantivy, Burn

- **Sole engineer:** Designed and built the Rust AI engine that powers search, recommendations and personalisation on Baze, deployed inside each client's own cloud (VPC).
- **AI in the binary:** Serve vision, embedding and ranking models from one Rust service on CPU, with no GPU runtime, within a tight latency budget.
- **Model training and distillation:** Built a GPU training sidecar that distils large video and audio teachers into a compact student and trains six model types: a content encoder, content-profile heads, a behavioural projector, a small reasoning core, a text encoder and a CLIP vision model. Each is evaluated on held-out data, tracking how data balance sets the accuracy ceiling.
- **Model release pipeline:** Cache teacher outputs once on GPU so training never reloads the teachers; quantize models to INT8 and sign them with Ed25519 so production loads only verified bundles, distributed to client VPCs through a signed model hub.
- **Search and runtime agility:** Blend BM25 keyword and semantic search with personalised ranking; run business rules as Rhai scripts that swap at runtime without a deployment.
- **Agentic AI in production:** Built an admin assistant agent that turns plain-language requests into audited configuration changes through typed write tools, and rolls back when it is unsure or a change fails its gates; evaluated on scripted operator conversations.

***Baze** | Safaricom streaming platform | Java 21, Spring WebFlux, R2DBC, Kafka, ClickHouse*

- **Revenue engine:** Built ms-baze-revenue-calculation from scratch: daily proration, play-minute usage shares, a gross-to-net tax and cost waterfall, configurable content-provider, Safaricom and platform splits, and a PROVISIONAL, FINALIZED, PAID and DISPUTED workflow with a full audit trail, tested with Testcontainers.
- **Analytics and attribution:** Built ms-baze-analytics: shortlinks, link previews, event ingestion and campaign attribution and revenue (ClickHouse, Kafka, Redis, GeoIP, 92 test files).
- **Core services:** Top contributor to Baze payment, catalog, auth, Skiza and playback services, using reactive Spring, Spring Security, Flyway, Redis and Resilience4j; playback events flow through Kafka into ClickHouse and on to recommendations.
- **Web platforms:** Delivered features on Baze Web and the VybStreamz CMS (Next.js, React, TypeScript, Tailwind CSS, Safaricom single sign-on).
- **Testing and delivery:** Built smoke, load and end-to-end test suites with simulated M-Pesa, carrier billing, DPO and USSD providers; ran back-pressure, spike and resource tests; profiled release builds under load and fixed latency regressions at the cause; kept Postman collections and client apps in step with every API change; wrote Terraform for the AWS estate.

Junior AI Engineer | Zone01 (Berrijam AI projects) | 2023 - 2025

- **AI integration:** Integrated ML models and LLM APIs into Go, Java and Python microservices.
- **LLM features:** Built prompt, retrieval (RAG) and structured-output features for client-facing tools.
- **Data preparation:** Cleaned, transformed and validated client datasets for modelling and analytics.
- **Prototype to production:** Turned validated AI prototypes into services with MLOps pipelines.
- **Evaluation and documentation:** Tested model outputs against baselines and documented APIs.

Fullstack Engineer (Java and Go) | Zone01 | Kisumu, Kenya | 2023 - 2025

- **Service performance:** Cut inter-service overhead by 30% by moving services to Go and gRPC.
- **Streaming pipelines:** Ran Kafka pipelines with sub-second ingestion for real-time monitoring.
- **Team standards and mentoring:** Set code-quality and DevSecOps standards, reviewed code, and mentored apprentice engineering teams.

Backend Engineer (Go and Java) | Upwork, international clients | Remote | 2018 - 2024

- **Real-time systems:** Built Go (REST, WebSockets) and Java Spring Boot backends for startups.
- **Full-stack delivery:** Integrated Angular frontends with Go and Java backends.
- **Reliability:** Deployed Dockerized services with auto-scaling and monitoring, holding 99% uptime, and resolved production incidents for clients.

KEY PROJECTS

Baze | Safaricom's video-on-demand and music streaming platform | Java 21, Spring Boot 4, Kafka, ClickHouse, Next.js | baze.co.ke

- **Platform:** Web, Android, iOS and TV.
- **Contribution:** Built the revenue engine and analytics services; top contributor to payments, catalog, auth, Skiza and playback; led the Spring Boot 4 upgrade; designed the Carrier Rail for M-Pesa and carrier billing.
- **AI customisation and integration:** Customised BBRaiN for Baze, adding page layouts and rows, hero slides, watch-next, kids and maturity filtering, and a catalogue and account sync from Baze's services. Then integrated it behind Baze's own login tokens to replace Algolia (search) and Recombee (personalised recommendations).

VybStreamz | entertainment hub powered by Safaricom | Next.js | vybstreamz.safaricom.co.ke

- **Dashboard:** Built content moderation using BBRaiN AI for the VybStreamz operations team.
- **Single sign-on:** Integrated Safaricom SSO (authorization-code flow) into the CMS.
- **AI customisation and integration:** Configured BBRaiN as a separate VybStreamz tenant with its own catalogue and rules, then integrated it to power search and recommendations.

BBRaiN | AI search and recommendations | Rust, Axum, Burn, CLIP, Tantivy, Kafka | thebbrain.io

- **Product:** One AI engine for search, recommendations and personalisation across all media, in production on Baze (Safaricom platform).
- **Platform:** A Rust inference engine, a GPU training factory, a signed model hub, a Leptos/WebAssembly operator console and an Angular viewer app.
- **Live demos:** Viewer app (bbrain.co.ke) and CMS studio (bbrain.co.ke/studio); access key available on request.

Hisiya | AI sentiment intelligence for African markets | hisiya.io

- **Local languages:** Reads brand sentiment in Swahili, Sheng, Pidgin and Amharic.
- **Brand intelligence:** Turns conversations into a brand-health score, trends and benchmarks.

EDUCATION AND CERTIFICATIONS

Software Engineering | ALX Africa | 2021 - 2024

Apprentice Mentorship | Zone01 Kisumu | 2024 - 2025

- **Programme:** Intensive hands-on software engineering in an enterprise setup. Credential Z01P-301.

HACKATHONS AND COMMUNITY

Agentic AI Hackathon | 48 hours | Jan 2026

- **Swahili farmer agent:** Built a ReAct agent for remote farmers that takes spoken requests in Swahili, diagnoses plant disease from a photo and alerts agricultural officers.

Community

- **Zone01 Kisumu:** Mentor to apprentice engineering teams.
- **Sui Foundation Community:** Developer Event (Dec 2025), a hands-on Sui blockchain workshop.
- **Kisumu Tech Community:** Member (2024 - Present).
- **Nairobi Gophers:** Member (2024 - Present).

REFEREES

- **Felix Mokua** | https://www.linkedin.com/in/felixmokua/
- **Paul Oguda** | https://www.linkedin.com/in/oguda/
- **Constantine Watitwa** | https://www.linkedin.com/in/constantine-watitwa-02120697/