

Wilson Gichuhi

Systems & Applied AI Engineer

✉ winchygichu@gmail.com ☎ +254700652437 📍 Nairobi, Kenya 🔗 Portfolio

🌐 LinkedIn profile 🐙 GitHub profile 📅 12/12/2000

Professional Summary

Systems & Applied AI Engineer specializing in enterprise AI integration. I build production systems that connect AI with existing business software across healthcare, commerce, logistics, and enterprise operations. My focus is designing compliant, reliable systems that bring AI into production.

Professional Experience

AI Evaluation Engineer (Contractor), *Bespoke Labs*

03/2026

- Contributed task authorship to OpenThoughts, Bespoke Labs' open reasoning dataset initiative that produced state-of-the-art small reasoning models trained on curated chain-of-thought data in collaboration with Stanford, Berkeley, and UT Austin.
- Designed and authored Terminal Bench style agentic coding tasks in the Harbor format, covering concurrency bugs, semaphore leaks, and write-ordering failures across real-world TypeScript and healthcare system scenarios, with full pipeline ownership from environment setup to oracle verification.
- Applied Bespoke-MiniCheck for claim-level factual grounding in evaluation pipelines, using NLI-based verification over LLM-as-judge alone to improve signal reliability in automated test suites.

Builder & AI Systems Engineer, *Promco*(Founder)

03/2026

Building **Promco**, the commercial decision layer for beverage distributors. Promco captures the operational judgment that forms between customer conversations and business commitments, transforming WhatsApp orders into structured, commercially reviewable decisions before execution in ERP, inventory, compliance, or payment systems.

- Designed the commercial decision workflow that captures evidence, applies business constraints, coordinates human review, and records explainable commitments rather than simply processing orders.
- Architected durable workflow orchestration using explicit state machines and recoverable execution, replacing free-form agent loops with traceable, compliance-safe commercial processes.
- Built the AI decision layer for extracting commercial intent from noisy WhatsApp conversations, combining retrieval, ranking, structured validation, and human review to resolve ambiguous real-world product requests.
- Designed operational safety boundaries using typed tools, versioned prompts, approval checkpoints, and complete execution tracing to prevent irreversible business actions without explicit authorization.
- Built integrations and runtime services over downstream ERP, inventory, compliance, and payment systems, allowing AI to reason over live operational context while preserving clear ownership boundaries between Promco and systems of record.

AI Training & Model Evaluation Engineer, *AfterQuery*

08/2025

- Frontier Model Benchmarking: Conducted systematic evaluations of frontier LLMs for logic correctness and security; engineered automated test suites to identify vulnerabilities across domains.
- SWE-Bench Style Engineering (Project Anvil): Built end-to-end evaluation environments using base snapshots and gold-patch verification; automated oracle verification pipelines for complex domains including HL7 health-tech integrations and secure build systems.
- Terminal Bench v2: Built real-world agentic coding tasks targeting concurrency bugs, security vulnerabilities, and infra debugging — with full pipeline ownership from gold patches to oracle verification and pytest suites in containerized environments.
- Adversarial Engineering: Designed stress-tests for LLM-driven API design, moving beyond naive prompting to automated red-teaming of model-generated system architectures.

Backend Engineer, *Alcora (FMCG Operations Platform)*

02/2025 – 09/2025

- Architected Headless ERP as a multi-tenant backend with real-time stock and demand intelligence across distributed FMCG actors.
- Built core payment rails with automated debt reconciliation across brands, distributors, and retailers.
- Embedded geospatial intelligence for delivery tracking, route batching, and dispatch automation.
- Enabled promo cost vs incremental gross profit tracking (North Star Metric ≤ 0.6), reducing promotional wastage by 40%.

Founding Engineer, *AfyaTelemed*

05/2024 – 07/2025

- Architected a HIPAA-compliant telemedicine platform from scratch, covering triage, consultations, pharmacy, and lab integration.
- Built secure video consultation with E2E encryption (WebRTC) and payment processing across local and international gateways.
- Stack: Python/FastAPI, Next.js, WebRTC, AWS

Software Engineer, *McSystems & Medical Inc.*

04/2024 – 10/2024

- Architected a CRM with voice-to-CRM logging, reducing manual data entry by 70%.
- Rebuilt a legacy PHP medical sales marketplace into an offline-first Next.js platform.
- Designed a unified API layer integrating CRM and marketplace systems.

Fullstack Developer, *JHUB Innovation Africa*

10/2023 – 05/2024

- Built digital trade and climate-smart agriculture platforms within a Microsoft-partnered innovation ecosystem
- Delivered enterprise-grade authentication and workflow integrations across multi-stakeholder systems.

Selected Projects

CareWeave : AI-Native Healthcare Coordination Platform

Designed an AI-native platform that transforms fragmented healthcare documents into a structured knowledge graph, enabling grounded clinical reasoning, institutional memory, and automated provider-payer coordination through human-in-the-loop validation and low-latency voice agents.

Technologies: Neo4j, Vector Embeddings, FastAPI, LiveKit, Knowledge Graphs, Human-in-the-Loop AI

Taji : Real-Time AI Product Demo Agent

Built a multilingual AI agent that joins live video calls to deliver interactive product demonstrations, navigate SaaS applications through visual agentic automation, qualify prospects, and synchronize outcomes with CRM and scheduling platforms.

Technologies: Python, FastAPI, LiveKit, Deepgram, Cartesia, VLM-based Agentic Automation

CVCS : AI Cognition Version Control System

Designed a version control system for enterprise AI cognition that versions agent context, decisions, and corrections, preserving organizational knowledge while enabling portable and auditable AI behavior across runtimes.

Technologies: Go, PostgreSQL, Event-Driven Architecture

Technical Skills

Enterprise AI Integration

- Model Context Protocol (MCP), OAuth/API integrations
- Browser automation (Playwright, Stagehand, CDP)
- Event-driven architectures
- Workflow orchestration and state machines

AI Evaluation & Alignment

- RLHF & SFT data engineering
- LLM evaluation (DeepEval, RAGAS, NLI, LLM-as-a-Judge)
- Adversarial testing

UX & Decision Interfaces

- Review dashboards
- HITL approval workflows
- Generative UI, Shared Autonomy Interfaces, and Headless UI primitives

Applied AI Systems

- Agentic workflows and tool use
- Retrieval-Augmented Generation (RAG)
- Vector embeddings, reranking, hybrid search
- Knowledge graphs
- Human-in-the-loop systems
- Voice AI

Backend & Platform Engineering

- Python, Go, TypeScript
- AWS, Docker, Kubernetes (K8s), IaC, Cloudflare Workers, Ollama
- FastAPI, Next.js, Effect-TS
- PostgreSQL, pgvector, DuckDB, Neo4j
- Cloud-native architectures

Education

Bsc. Computer Technology, JKUAT University

2018 – 2023