

EMMANUEL NWANGUMA

Data Scientist | ML Engineer | MLOps Engineer | AI Engineer

Lagos, Nigeria | nwangumaemmanuel29@gmail.com

Portfolio emart29.vercel.app | GitHub [Emart29](https://github.com/emart29) | LinkedIn [nwangumaemmanuel](https://www.linkedin.com/in/nwangumaemmanuel) | Medium [@Emar7](https://medium.com/@Emar7) | X [@1guma_](https://twitter.com/1guma_)

PROFESSIONAL SUMMARY

ML and MLOps engineer who builds production AI systems end to end: data pipelines, model training, fine-tuning, deployment, monitoring and evaluation. Microsoft Certified Machine Learning Operations Engineer Associate. Independently built and maintain Emart AI, an open-source AI infrastructure project with six shipped tools across agent memory, evaluation, observability and security, five published to PyPI and one to npm. #13 of 200+ contributors to aden-hive/hive (OpenHive, 11,000+ GitHub stars) with 14 merged pull requests. Built a production MLOps platform covering feature store, pipeline lineage tracking and automated canary deployment with rollback, and fine-tuned Microsoft Phi-4 for financial question answering at a 69% ROUGE-L improvement over base. Author of technical articles in Towards AI, DataDrivenInvestor and NextGenAI.

TECHNICAL SKILLS

Languages: Python, Go, TypeScript, SQL, Bash

ML and AI: PyTorch, Hugging Face Transformers, PEFT/LoRA/QLoRA, scikit-learn, FAISS, ChromaDB, LangChain, LlamaIndex, Evidently AI, Model Context Protocol (MCP)

LLM APIs: OpenAI, Anthropic, Groq, Gemini, Ollama — multi-provider production experience

MLOps and Deployment: Docker, MLflow, FastAPI, GitHub Actions CI/CD, Prometheus, Grafana, Redis, MinIO, Bicep, Azure Machine Learning, Microsoft Foundry

Cloud and Data: AWS (SageMaker, EC2, S3), GCP (BigQuery), Modal, Vercel, PostgreSQL, Redis, Kafka/Redpanda, TimescaleDB

EXPERIENCE

Independent ML/MLOps Engineer | Emart AI

Jul 2024 to Present | Remote, Lagos

- Built and shipped six production open-source tools across memory, evaluation and observability/security for AI agents, all MIT-licensed and self-hostable; five published to PyPI, one to npm.
- Built Remembr and Mindr (memory layer): persistent cross-session memory for multi-agent systems and codebase memory for AI coding agents including Claude Code, Cursor and Aider. Dual Python and TypeScript SDKs, hybrid semantic, keyword and recency-weighted retrieval, multi-tenant isolation, 8 framework adapters.
- Built evalflow and PlayAgent (evaluation layer): CI-integrated regression testing for LLM prompts across four providers, and pre-production agent testing covering six real failure modes including tool call failures, context drift, hallucinated results and prompt injection.
- Built Spendline and Scopeform (observability and security layer): per-run cost and token tracking, and scoped, individually revocable agent credentials with full access auditing.
- Built a production MLOps platform: feature store (PostgreSQL/Redis dual-store, point-in-time correct joins, sub-10ms serving), pipeline lineage tracker (SHA-256 dataset versioning, five-hop prediction traceability), and canary deployment with Redis traffic split, Prometheus monitoring and automated rollback within 60 seconds.
- Benchmarked structured output reliability across 833 tests on 6 models, finding that 87.6% single-call success collapses to 26.7% over ten chained calls, and that enforcement tier never beat the prompt-only baseline.
- Fine-tuned Microsoft Phi-4 Mini (3.8B) on SEC 10-K financial Q&A using QLoRA with 4-bit quantisation, achieving a 69% ROUGE-L improvement over base.
- Built a real-time anomaly detection pipeline (Kafka/Redpanda, Faust, TimescaleDB) processing fraud, server-metric and IoT streams from event to deduplicated alert in under two seconds, validated against a real Kaggle fraud dataset after simulated data returned implausible results.
- Built a causal inference engine designed to decline questions the data cannot support, an online-learning recommender that updates within the request rather than on a nightly retrain, and a RAG system that verifies cited sources actually support the claims made.
- Deployed a production NLP classifier (DistilBERT, 19 categories) at approximately 42ms inference with real-time drift detection via Evidently AI.
- Ran a 90-day build-in-public campaign documenting production ML failure modes and their fixes.

Open Source Contributor (#13 of 200+) | aden-hive/hive (OpenHive), 11,000+ GitHub Stars

Jan 2026 to Present | Remote

- Merged 14 pull requests into an 11,000-star, 200+ contributor repository, ranking #13 by contribution volume.

- Added 458 tests across 7 security scanning tools and 3 API integration tools, closing coverage gaps flagged by maintainers.
- Shipped a BigQuery Model Context Protocol tool integration, and fixed credential exception handling, an MCP resource leak and storage path validation, all merged into production.
- Wrote 33 tool READMEs, the largest batch covering 22 tools, reducing onboarding friction for new contributors.

Health Data Specialist | Mido Health Diagnostic

Nov 2022 to Mar 2025 | Abia State, Nigeria

- Owned data quality for clinical datasets of 50,000+ records; built automated Python and SQL validation pipelines maintaining above 99% accuracy.
- Cut manual data corrections by roughly 30% through automated quality checks, while maintaining strict healthcare data privacy compliance.

Data Analyst Intern | ACME Software Lab

Jan 2024 to Apr 2024 | Remote

- Built a reusable ETL pipeline handling cleaning, transformation and validation across multiple sources; adopted as the team's standard template.
- Applied statistical analysis and BI techniques to translate raw data into business insights used in stakeholder decisions.

Data & Records Assistant | Ministry of Health, Osogbo, HIV/AIDS Unit (NYSC)

Oct 2021 to Oct 2022 | Osogbo, Nigeria

- Managed data collection and analysis for 150+ HIV/AIDS patients, producing monthly statistical reports used in public health planning.
- Redesigned filing systems, improving record retrieval efficiency by 40% while maintaining strict patient confidentiality.

PROJECTS

Emart AI: Six-Tool Open Source AI Infrastructure | github.com/emartai

Remembr and Mindr (agent memory), evalflow and PlayAgent (evaluation), Spendline and Scopeform (observability and security). All MIT-licensed; five published to PyPI, one to npm.

Structured Output Reliability Benchmark | github.com/Emart29/so-agent

833 tests across 6 models with confidence intervals on every rate. 87.6% single-call success collapses to 26.7% over ten chained calls; 100% structurally valid outputs were only 68.9% semantically accurate.

RAG with Verified Citation Grounding | github.com/Emart29/ragwell

Citations checkable by string containment, with InsufficientEvidence as a first-class answer. Benchmarked on four regulatory annual reports: 126x token cost difference between retrieval and long context. 381 tests.

ML Canary Deployment with Auto-Rollback | github.com/Emart29/ml-canary-deploy

Redis traffic split, Prometheus and Grafana observability, automated rollback on error rate or latency degradation. A deliberately underfit model at 54% against an 85% baseline was caught and rolled back in under 60 seconds with zero human involvement.

ML Pipeline Lineage Tracker | github.com/Emart29/pipeline-lineage-tracker

Traces any production prediction back through five hops to the exact raw data row, pipeline run and git commit. SHA-256 dataset versioning, NetworkX DAG lineage, MLflow integration.

Production ML Feature Store | github.com/Emart29/ml-feature-store

Dual-store architecture (PostgreSQL offline, Redis online) solving training-serving skew. Point-in-time correct joins, SHA-256 feature versioning, Evidently AI drift detection, sub-10ms serving.

Causal Inference Engine | github.com/Emart29/causal-inference-engine

Five estimators compared side by side, graded on synthetic data with a planted true effect (10 of 10 rows as expected). On the LaLonde benchmark it declines to quote an effect rather than returning the closest wrong answer.

Real-Time Anomaly Detection Pipeline | github.com/Emart29/realtime-anomaly-detection

Kafka/Redpanda and Faust streaming across fraud, server-metric and IoT sources. Sub-two-second detection to deduplicated alert, validated against a real Kaggle fraud dataset.

Phi-4 Mini Fine-Tuned for SEC Financial Q&A | github.com/Emart29/phi4-finance-finetuning

QLoRA fine-tuning with 4-bit quantisation on consumer hardware. 69% ROUGE-L improvement over base model.

Online Learning Recommender | github.com/Emart29/personalization-feedback-engine

Per-interaction learning with River. Hit rate improved 0.045 to 0.390 per user with zero retraining runs; published the population-level result where it came last.

locksmith: Dangerous Postgres Migration Checker (Go) | github.com/Emart29/locksmith

CLI tool flagging table rewrites, exclusive locks and blocking ALTER TABLE before they reach production. Distributed via Homebrew, used in CI pipelines.

Production E-Commerce Product Classifier | github.com/Emart29/ecommerce-product-classifier

DistilBERT across 19 categories at approximately 42ms inference. Real-time drift detection via Evidently AI, FastAPI serving, React analytics dashboard.

Heart Disease Risk Prediction | github.com/Emart29/Heart-Disease-Prediction

End-to-end production ML system at 88.5% accuracy. FastAPI REST API, SHAP explainability, MLflow experiment tracking, Docker deployment.

CERTIFICATIONS

- Microsoft Certified: Machine Learning Operations Engineer Associate (AI-300), Aug 2026 to Aug 2027. Credential ID 57A7E1CF27219B68
- AI Engineer for Data Scientists Associate, DataCamp, Jul 2026 to Jul 2028
- ISO/IEC 42001 AI Management Systems – Awareness eLearning, UKAS / AIQI, Jun 2026
- Machine Learning Engineering Career Track, AI Community Africa x DataCamp, Jul 2026
- Google Cloud Data Analytics Certificate, Google Cloud, Jun 2026 to Jun 2029
- Google Advanced Data Analytics Professional Certificate, Coursera, Jun 2026
- Data Scientist Associate, DataCamp, 2026
- IT Specialist, Artificial Intelligence, Certiport / Pearson VUE, 2025
- NSQ Level 4 in Artificial Intelligence, Computer Professionals Registration Council of Nigeria, 2025. Cert no. CPN/NSQ/24177273
- Google Data Analytics Professional Certificate, Google, 2025
- AWS Educate Machine Learning Foundations, AWS Training and Certification, 2025
- Artificial Intelligence Fundamentals, IBM SkillsBuild, 2025

EDUCATION

MSc Financial Engineering (in progress), WorldQuant University | Resuming Oct 2026

BSc Biochemistry (Second Class Upper), Clifford University, Abia State, Nigeria | Nov 2016 to Mar 2021

ACADEMIC PUBLICATION

[A Systematic Review of Generative AI in Education](#), Journal of Computer Sciences and Applications, 2024, 12(1), 25–30. DOI: 10.12691/jcsa-12-1-4 (open access, co-authored)

SELECTED TECHNICAL WRITING

- Every AI Tool Shows You Sources Now. I Checked Whether They Actually Say What the AI Claims, Towards AI, Aug 2026
- The AI Feature That Works in the Demo and Breaks in Production, Towards AI, Aug 2026 — 833 tests across 6 models
- I Built a Recommender That Learns Per Click. Then It Lost the Benchmark, Towards AI, Aug 2026
- I Built a Causal Inference Engine. The Best Thing it Does is Refuse to Answer, Towards AI, Jul 2026
- Real-Time Anomaly Detection With Kafka and Faust, Towards AI, Jul 2026
- I Built a System that Automatically Rolls Back ML Models Before They Ruin Your Production Data, Towards AI, Jun 2026
- Can You Trace Any Prediction Back to the Exact Data That Caused It? Towards AI, Jun 2026
- The MLOps Component Nobody Builds in Their Portfolio, Towards AI, Jun 2026
- I Tested 5 RAG Search Strategies on the Same Dataset. Here Are the Real Latency Numbers, NextGenAI, Mar 2026
- I Trained My Own Financial AI From Scratch. Here's What a 69% Improvement Actually Looks Like, Mar 2026
- I Added Real-Time Drift Detection to My ML Classifier, DataDrivenInvestor, Mar 2026
- I Added CI/CD and Monitoring to My ML Project, DataDrivenInvestor, Feb 2026
- Everyone's Building AI Wrappers. Few Are Solving Real Data Infrastructure Problems, NextGenAI, Feb 2026