Olusegun Ekoh

AI/ML Engineering | Agentic AI & Multi-Agent Systems

segun.ekoh@gmail.com • 07471900437 LinkedIn • Portfolio: olusegunekoh.com • Hatfield, Hertfordshire

Innovative AI/ML Engineer with 6+ years of experience designing and deploying production-grade intelligent systems that integrate Generative AI, Agentic AI, and cloud technologies to drive enterprise transformation. Proven expertise in building scalable AI agent architectures, multi-agent systems, and RAG pipelines deployed across AWS and GCP infrastructure. Distinguished track record architecting full-stack solutions for financial services (serving 2M+ users), implementing MLOps practices, and delivering enterprise-grade systems with 99.99% uptime. Skilled in leveraging LLMs, transformers, vector databases, and modern frameworks to build autonomous agents, semantic search systems, and intelligent automation platforms. Adept at translating complex AI requirements into production-ready solutions while ensuring compliance, security, and operational excellence.

Areas of Expertise

- Generative AI & Agentic AI
- Multi-Agent Systems & Orchestration
- AI/ML Engineering & MLOps
- RAG Pipelines & Vector Databases
- LLM Integration & Prompt Engineering
- Production AI Deployment (HPC/GPU)
- Cloud Infrastructure (AWS)
- API Design & Microservices
- Full-Stack Engineering
- Performance Optimization
- CI/CD & DevOps

Technical Skills

AI/ML & Agentic AI

PyTorch | scikit-learn | Hugging Face Transformers | LangChain | CrewAI | SpaCy | Sentence-Transformers | OpenAI API | Anthropic API | Gemini API | Prompt Engineering

Vector Databases & Search

FAISS | pgvector | Chroma | Semantic Search | Embeddings | RAG (Retrieval-Augmented Generation)

Programming Languages

Python | TypeScript | Java

Frontend Frameworks

React.js | Next.js

Backend Frameworks & APIs

FastAPI | Node.js | Express.js | NestJS | | RESTful API Design | GraphQL | Microservices

Cloud & Infrastructure

AWS (IAM, ECR, API Gateway, S3, Lambda, App Runner, Bedrock, SageMaker) | GCP (Cloud Run) | Docker | Terraform | Kubernetes

DevOps & MLOps

CI/CD (GitHub Actions) | Datadog | Prometheus | Grafana | Model Deployment | Cost Optimization

Professional Experience

AI Engineer (DevOps)

University of Hertfordshire – Hatfield, UK | 2024 — 2025

Engineered production-grade AI system to automate Simplified Technical English (STE) translation under Knowledge Transfer Partnership with Sono Vision Ltd, ensuring full alignment with ASD-STE100 compliance standards for aerospace and defence documentation. Architected advanced NLP pipeline leveraging *SpaCy* for deep linguistic parsing, OpenAI GPT models for intelligent correction, and statistical validation frameworks. Orchestrated enterprise-scale deployment on High-Performance Computing infrastructure with SLURM job scheduling and GPU acceleration, demonstrating complete AI lifecycle ownership from development through production monitoring.

- Optimised LLM integration with OpenAI GPT models through efficient token utilisation, exponential backoff retry mechanisms, and dynamic cost-control algorithms to enable scalable, cost-effective AI processing in production environments.
- Architected multi-phase rule-based correction pipeline incorporating 53 STE linguistic rules, comprising regex-driven rule detection, automated text normalisation, and LLM-powered adaptive correction with intelligent prioritisation and retry optimisation strategies.
- Elevated BLEU scores by approximately 41% and ROUGE-2 scores by 64% on complex semantic test cases, achieving successful correction of 78 instances within a 944-case evaluation dataset while maintaining 75.1% stability on non-LLM test segments.
- Designed and deployed interactive Streamlit application delivering real-time STE sentence correction with detailed compliance reporting, dynamically generated rule violation insights, and comprehensive audit trails for governance.
- Orchestrated large-scale production deployment on HPC infrastructure using SLURM job scheduling and GPU acceleration, implementing monitoring systems and optimising computational efficiency for enterprise workloads.

AI/ML Engineer (Portfolio)

Personal AI Engineering Projects – olusegunekoh.com | 2025 — Present

Built production-ready AI systems demonstrating expertise in Generative AI, Agentic AI, multi-agent orchestration, and MLOps practices. Developed intelligent applications leveraging LLMs, RAG pipelines, vector databases, and modern AI frameworks deployed on cloud infrastructure.

- Engineered multi-agent AI system using CrewAI and LangChain frameworks to automate job discovery, analysis, and personalized application generation, implementing agent orchestration with specialized agents for search, analysis, content writing, and outreach, integrated with third-party APIs and comprehensive resource monitoring for token consumption tracking across multiple LLM providers (OpenAI, Gemini).
- Developed production RAG system (Doc Insight) using FastAPI, PostgreSQL with pgvector extension, and OpenAI embeddings, implementing semantic vector search with 1536-dimensional embeddings, OCR support for scanned PDFs (PyMuPDF + Tesseract), background task processing for document chunking, and context-aware GPT-4o responses with source citations.
- Built semantic search engine processing 41,000+ ML research papers using PyTorch, Sentence-Transformers (*MPNet* model), and GPU-accelerated FAISS indexing, achieving 768-dimensional embeddings with 41-second generation time (~30 iterations/second) and near real-time semantic retrieval based on meaning rather than keywords.
- Created intelligent PDF reader with LangChain and OpenAI GPT featuring conversational memory, semantic search using Chroma vector database, question-answering with source citations, and embedded PDF viewer for context verification.

Product Software Engineer

Oval Finance – Delaware, USA | 2022 — 2024

Architected complex payment integrations with three crypto-service providers, navigating intricate API documentation and security protocols to expand the platform's payment capabilities twofold. Delivered and maintained scalable microservices architecture aligned with business requirements and regulatory compliance. Designed high-performance APIs, implemented DevOps practices, and led cross-functional initiatives.

- Directed three cross-functional initiatives in close collaboration with product managers and UX designers, refining requirements-gathering frameworks that accelerated development cycles by 30% and improved feature delivery accuracy.
- Built high-performance mobile backend APIs capable of processing over 2,000 blockchain transactions per day, amplifying analytical data throughput and boosting processing efficiency by 50% through optimised database queries and caching strategies.
- Designed and developed microservice APIs for KYC engine to automate synchronisation and backup of 100+ business records daily from third-party providers, attaining 99.99% data accuracy and ensuring full regulatory compliance with financial services standards.
- Architected and deployed end-to-end internal admin dashboards using modern frameworks, empowering over 15 administrators with real-time financial intelligence and enhancing decision-making responsiveness by 40%.

Full-Stack Software Engineer

Guaranty Trust Bank – Lagos, Nigeria | 2020 — 2022

Delivered and supported enterprise technology solutions tailored to meet customer requirements and business objectives for financial institution serving 2M+ users. Investigated, diagnosed, and resolved software bugs to ensure system reliability and 99.99% uptime. Drove continuous improvements in mobile applications to enhance performance and user experience.

- Drove 35% increase in user engagement and digital adoption metrics by engineering and delivering two large-scale customer-facing mobile applications serving over 2M users across Nigeria's banking ecosystem.
- Revamped manual banking operations by developing and deploying six automated workflows and administrative modules, reducing staff processing time by 45% and enhancing overall operational efficiency through intelligent automation.
- Elevated customer satisfaction scores by 25% by implementing and maintaining three strategic technology solutions aligned with business objectives and client needs, ensuring seamless integration with core banking systems.
- Improved three legacy in-house systems performance through proactive diagnostics and resolution of eight critical defects, achieving 99.99% system uptime and enhancing application performance by 40% through code optimisation and database tuning.

Software Developer

Secure Networks Ltd – Lagos, Nigeria | 2018 — 2020

- Enhanced project delivery accuracy by 40% by translating complex business requirements into precise technical specifications in collaboration with cross-functional stakeholders.
- Ensured 99.9% uptime and maintained sub-300ms response latency by developing and optimising over 25 RESTful API endpoints with comprehensive error handling.
- Revitalised deployment workflow by implementing robust CI/CD pipelines, reducing deployment duration by 70% and elevating release frequency by 80%.
- Enforced code quality through comprehensive testing, attaining 85% test coverage and cutting production defects by 50%.

Education & Professional Development

MSc Artificial Intelligence & Robotics (Distinction)

University of Hertfordshire – Hatfield, UK | 2025

Thesis: Automated Translation from Standard English to Simplified Technical English using Prompt-Engineered LLMs

BSc Physics & Electronics (First Class Honors)

Redeemer's University – Osun, Nigeria

Licenses & Certifications

- Google Cloud Certified Professional Cloud Developer
- AWS Machine Learning Foundations
- Introducing Generative AI with AWS