

MALIK SHOLADOYE GARUBA

Senior Software, AI & Data Engineer

+44 7424 630315 | garubamalik@gmail.com | linkedin.com/in/garubamalik

PROFESSIONAL SUMMARY

Senior Software, AI and Data Engineer with 9+ years of experience designing and productionising enterprise-grade systems across banking, capital markets, higher education and automotive. Currently building AI transformation, retrieval and PII-redaction pipelines at NatWest, with prior delivery of RAG systems, LLM-powered agents and large-scale ETL/ELT platforms on AWS and Azure. AWS Certified Developer and Data Engineer with deep Python expertise, combining Generative AI engineering (LLMs, agents, embeddings, vector databases, prompt and context engineering, evaluation) with hands-on data engineering (PySpark, Databricks, Kafka, Azure Data Factory, event-driven architectures). Track record of turning disparate enterprise data into governed, searchable, AI-consumable knowledge assets, and of taking solutions from proof of concept to secure production deployment while reducing processing time by 20% and infrastructure cost by 10%.

CORE COMPETENCIES

Generative AI & LLMs: OpenAI GPT-4 APIs, Llama 3, DALL-E, LangChain, LangChain Agents, OpenAI Agents SDK, prompt engineering, context engineering, structured outputs, function calling and tool-augmented agents, multi-step agent workflows, NLP

Retrieval & Knowledge: Retrieval-Augmented Generation (RAG), embedding pipelines, semantic search, vector databases (FAISS, Pinecone, Weaviate), document chunking and enrichment, retrieval caching strategies, enterprise search and knowledge discovery

LLMOps & Evaluation: Automated and human-in-the-loop evaluation frameworks, LLM and agent testing, prompt and model versioning, AI observability and tracing, hallucination mitigation, guardrails and policy enforcement, token optimisation and inference cost management, CI/CD for AI pipelines

Responsible AI & Governance: PII detection and redaction, content safety and moderation, bias- and fairness-aware evaluation, explainability-focused response design, data privacy and compliance-aligned AI workflows (financial services regulated environments)

Data Engineering: ETL/ELT pipeline design, PySpark, Databricks, Polars, Pandas, Dask, Kafka, Azure Data Factory, AWS Glue, event-driven and real-time streaming architectures, data modelling, metadata management, data quality, lineage and governance

Data Platforms & Storage: Redshift, Azure Lakehouse, Azure Data Lake, AWS S3, PostgreSQL, SQL Server, MongoDB, MariaDB, Redis, advanced SQL

Software Engineering: Python (Expert), Java, JavaScript, FastAPI, Django, Spring Boot, Netty, React.js, REST API design, microservices, SOLID, TDD (Pytest, JUnit, Mockito)

Cloud & DevOps: AWS (Lambda, S3, SQS, EKS, RDS, Aurora, DocumentDB, Redshift), Azure (Data Factory, Key Vault, Storage, DevOps), Docker, Kubernetes, Terraform, Ansible, GitLab CI, Jenkins, Azure DevOps Pipelines, Git

Leadership: Cross-functional team leadership, stakeholder collaboration, technical design reviews, engineering standards, mentoring

CERTIFICATIONS

- AWS Certified Data Engineer – Associate (2024)
- AWS Certified Developer – Associate (2023)
- International Project Management Professional (IPMP) (2016)

PROFESSIONAL EXPERIENCE

NatWest Group

AI Engineer

April 2026 – Present

- Design and maintain AI rewriter and transformer pipelines that improve retrieval-as-a-service (RAAS) relevance and response quality for enterprise knowledge retrieval, applying prompt and context engineering to raise accuracy and consistency.
- Built and integrated an NLP-based PII scanner that detects names, emails, phone numbers, identifiers and financial data in enterprise documents before they reach any LLM, with secure redaction and masking workflows supporting privacy, compliance and safe model usage.
- Developed transformation and enrichment pipelines that convert Blueworks process artefacts into normalised, metadata-rich, AI-ready structured formats for indexing and downstream retrieval.

- Embedded LLM inference, evaluation and regression testing into GitLab CI/CD (GitLab Runner), enabling repeatable validation of prompt and model changes before deployment.
- Collaborate with data, platform and governance teams to define engineering standards for trustworthy, traceable AI workflows in a regulated banking environment.

University of Salford

Data Developer

August 2025 – April 2026

- Designed, developed and optimised ETL/ELT pipelines in Python and PySpark on Databricks to ingest, cleanse, model and enrich large-scale institutional datasets from multiple source systems.
- Orchestrated and monitored pipelines with Azure Data Factory, consolidating disparate data sources into a centralised, governed lakehouse architecture with clear lineage and data quality checks.
- Wrote robust SQL transformation and validation scripts; designed data models supporting analytics and reporting use cases.
- Automated build, test and release through Azure DevOps YAML pipelines and Git-based workflows, delivering reliable CI/CD across environments and enforcing peer code review.
- Integrated Azure Storage and Azure Key Vault for secure storage and secrets management; partnered with cross-departmental stakeholders on secure system integrations.

London Stock Exchange Group (LSEG)

Senior Software Engineer

March 2025 – August 2025

- Architected secure, scalable cloud-native services on AWS with strong emphasis on data protection and regulatory compliance for World-Check on Demand, a real-time risk-intelligence API embedded in payment and onboarding workflows.
- Developed Python ETL pipelines with complex data mapping and enrichment across diverse sources, delivering high-quality, integrated analytics data.
- Implemented infrastructure-as-code and CI/CD with Terraform and GitLab CI, including a centralised AWS backup platform for S3, DocumentDB, Aurora and RDS, improving deployment efficiency and resilience.

Freelance

Software Engineer (AI Applications)

September 2024 – December 2024

- Built and optimised scalable FastAPI web applications integrating multiple LLM providers, with structured outputs and tool-calling for intelligent decision support.
- Improved throughput and cost through response caching, request queuing and database optimisation.

EPAM Systems

Senior Software Engineer

May 2022 – September 2024

- Implemented ETL pipelines processing high-volume vehicle driving data to update in-vehicle HAD map systems over the air (ADAS Remote Services), running on Kubernetes and AWS (S3, SQS, Lambda, EKS, Redis).
- Designed and delivered LLM-powered web applications and an agent-based Generative AI solution (GPT-4, DALL-E, OpenAI SDK, FastAPI) for automated content generation and ideation, with automated testing and CI/CD on AWS.
- Built a RAG proof of concept for legal analysis using GPT-4, embeddings and FAISS to enable semantic search, document retrieval and summarisation, with agent workflows that cut research time by 40%.
- Developed real-time data streaming and forecast-monitoring applications on Redshift (Polars, PySpark, Plotly Dash) supporting forecasting and strategic planning.
- Improved processing performance and cost by selecting data libraries appropriate to workload and budget; resolved memory leaks using Valgrind and cProfile.
- Built FastAPI and Django backends and designed APIs integrating internal and external systems in cloud-native environments; enforced data integrity through RDBMS triggers and validation.
- Led a cross-functional team to deliver a complex programme on time and within budget.

Swifta Services Limited

Back-End Developer

November 2020 – March 2022

- Designed, developed and supported SwiftaHR, a comprehensive ERP for employee management, payroll and finance (Django, DRF, Frappe, PostgreSQL, Redis).
- Built configurable workflow engines allowing each client to model its own internal processes; enforced data integrity across employee lifecycle events.
- Developed real-time streaming applications providing live performance tracking, simplified reporting and data-driven decision making; wrote clean, SOLID-compliant APIs.

MOMAS (Memmmcol)

Software Developer

January 2019 – November 2020

- Developed and deployed scalable server-side applications in Python and Java (Netty) for meter data management, implementing DLMS, MODBUS and MQTT protocols.
- Delivered rigorous automated and load testing (Pytest, Locust) across MSSQL and PostgreSQL backends on AWS.

MOMAS (Memmmcol)

Design Engineer

November 2016 – December 2018

- Developed memory-efficient firmware for metering hardware and automated reliability testing using Raspberry Pi and Python.

Earlier career: System Integrator / Support Personnel, MOMAS System Nigeria Limited; Tax Officer, Cross River Internal Revenue Service.

SELECTED PROJECTS

- **Enterprise PII Scanning & Redaction (NatWest):** NLP-based detection and secure redaction of sensitive data in documents ahead of LLM processing; integrated into GitLab CI/CD.
- **RAG System for Legal Analysis:** GPT-4 + FAISS retrieval pipeline with embeddings, semantic search, summarisation and automated research agents; 40% reduction in research time.
- **IoT Generative AI Agent:** Agent-based solution orchestrating GPT-4 and DALL·E via the OpenAI SDK and FastAPI for automated article generation and IoT ideation; tested and deployed through CI/CD on AWS.
- **World-Check on Demand (LSEG):** Cloud-native, real-time risk-intelligence API delivering secure compliance checks inside payment and onboarding flows (AWS, Terraform, Python).
- **Centralised Backup Platform (LSEG):** Terraform-provisioned AWS service centralising backups for S3, DocumentDB, Aurora and RDS.
- **ADAS Remote Services (EPAM):** Over-the-air HAD map management for autonomous driving on Kubernetes and AWS (S3, SQS, Lambda, EKS, Redis).
- **FOMO & QA Toolkit (EPAM):** Forecast-monitoring and QA applications on Redshift using Polars, PySpark, Pandas and Plotly Dash.
- **BadgerDoc (EPAM):** ML delivery platform visualising model delivery cycles (Python, Plotly Dash, Docker, PostgreSQL).
- **SwiftaHR ERP & OManager:** Enterprise HR/payroll ERP and community management platform (Django, Frappe, React.js, PostgreSQL, AWS).
- **Meter Data Management & Token Generation Server (MOMAS):** High-throughput Java/Netty services for energy meters and secure token generation using DLMS, MODBUS, MQTT and Thrift.

EDUCATION

- B.Sc. Computer Science (Second Class Upper)
- National Diploma in Computer Science