

# Tola Shobande

Based in Guildford, United Kingdom

+447547756583 | [tolashobande@gmail.com](mailto:tolashobande@gmail.com) | [tolajs](#) | [tolajs](#)

## Summary

---

Machine Learning Engineer and recent M.Sc. graduate with a background in Data Engineering and Data Science. Enthusiast for efficient workflows who enjoys exploring open-source tools and optimizing development environments to maximize productivity. Interested in devising a better problem-solving method for challenging tasks, and learning new technologies and tools if the need arises.

## Experience

---

### Data Scientist

Remote

Smart Data Warehouse Limited

Apr. 2023 - Present

Worked remotely for a Canadian start-up consultancy firm, delivering AI solutions and data mastery to enterprise clients.

- Built credit risk model for fintech client processing 2,000-3,000 weekly loan applications; developed features in Databricks using PySpark and trained XGBoost classifier achieving 0.76 precision in identifying loan defaults.
- Created feature pipelines reading from Snowflake, processing in Databricks, and writing results back to Snowflake data warehouse; scheduled weekly retraining jobs using Databricks workflows.
- Deployed churn prediction model for B2B SaaS client with 12K users; API processed 2-3M events monthly on AWS Lambda and helped reduce churn from 8% to 5.5% over six months.
- Built demand forecasting system for e-commerce client with 1,200 products across 3 warehouses; weekly forecasts reduced stockouts by 15% and overstock by 12%.
- Set up model monitoring tracking prediction drift and performance; caught precision drop from 0.72 to 0.58 before sales team noticed, enabling rapid retraining with recent customer data.
- Resolved production bottlenecks by migrating high-volume jobs from AWS Lambda to EC2 batch processing, and curbed Databricks cost overruns by implementing auto-termination policies.
- Automated monthly reporting for 8-10 clients using Python scripts, reducing generation time from 3 days to 4 hours.
- Managed end-to-end deployment and maintenance of production models across AWS (S3, Lambda, EC2, RDS), Snowflake, Databricks, and FastAPI.

### AI Engineering Fellow

Remote

DataExpert

Jul. 2025 - Aug. 2025

Completed an intensive AI engineering fellowship focused on architecting production ready AI systems. Specialised in designing advanced retrieval pipelines and autonomous agents to solve complex reasoning tasks.

- **RAG Architecture:** Built advanced retrieval pipelines using GraphRAG to handle complex, multi-hop reasoning over structured data, moving beyond standard vector search limitations.
- **Agentic Workflows:** Engineered agentic workflows using LangGraph and the Model Context Protocol (MCP), creating systems that integrate with external tools to execute asynchronous tasks.
- **Production MLOps:** Implemented LLM Ops best practices, utilizing LangSmith for observability to trace agent reasoning steps, debug latency, and enforce strict safety guardrails against hallucinations.
- **Prompt Engineering Strategy:** Developed systematic testing frameworks for zero-shot and few-shot prompting, optimizing model performance for domain-specific accuracy.

### Data Engineer

Abuja, Nigeria

Nigerian Communications Commission (NCC)

Feb. 2022 - Feb. 2023

- Built data pipelines processing 150-200M daily call records from four major telecom operators for regulatory compliance, maintaining a 50TB+ database covering 18 months of submissions.
- Cut monthly compliance report generation time from 5 days to 6 hours by automating data collection and validation, catching 12-15% of operator submissions with data quality issues each month.
- Optimized PostgreSQL queries by implementing indexing and table partitioning, reducing execution time from 45+ minutes to under 3 minutes for annual compliance reports.
- Created a consumer complaint dashboard used by 15+ regulatory officers to process billing disputes and service quality issues requiring strict 48-hour response times.
- Engineered robust format validators and error handling scripts to prevent pipeline failures caused by unannounced file format changes from operators.
- Recovered 40% of database storage capacity during critical shortages by establishing automated archival and compression for historical data.
- Leveraged Power BI, Python, PostgreSQL, and AWS to deliver solutions within a resource-constrained environment, overcoming legacy system limitations and infrastructure instability.

### AI & Python Tutor

London, UK

FireTech

Sep. 2020 - Sep. 2022

- Mentored 30+ students through Python fundamentals and project-based assignments, achieving a 90% pass rate and consistent 5-star feedback.
- Led interactive sessions on Machine Learning concepts using tools like Google's Teachable Machine, translating complex logic into accessible, hands-on exercises for beginners.

# Projects

---

## Spotify Chatbot

Python, GraphRAG, LangGraph, LLMs, Git

### Built a natural language interface for analyzing personal streaming history using GraphRAG.

- Engineered a Graph RAG Chatbot for interactive Spotify listening history analysis, enabling users to query their personal streaming data through a natural language interface.
- Engineered a chatbot capable of analyzing complex listening history by transforming flat JSON data into a traversable Neo4j graph.
- Deployed LangGraph to manage multi-step reasoning, coordinating agents to retrieve historical data and synthesize answers with high accuracy.

## Transformer-Based Sketch Reconstruction - Thesis Project

Python, PyTorch, Transformers, Seq2Seq, Geometric Deep Learning

### Developed a transformer-based deep learning model to automate the accurate reconstruction of complex 2D concept sketches in 3D for design workflows.

- **Architecture Adaptation:** Specialized the Sketchformer Transformer architecture for generative vector modeling, enabling the high-fidelity reconstruction of concept sketches from the OpenSketch dataset.
- **Geometric Data Pipeline:** Engineered a data ingestion system to process raw 3D wireframes, applying orthographic projection matrices and normalization to generate training-ready 2D vector sequences.
- **Temporal Sequence Modeling:** Implemented the 'stroke-5' vector format to encode pen dynamics (draw, lift, end), successfully preserving the sequential nature of human drawing for the Transformer.
- **Performance Evaluation:** Validated geometric fidelity using Chamfer Distance and Earth Mover's Distance (EMD), achieving a structural reconstruction score of 0.0075.

# Education

---

## University of Surrey

M.Sc. in Artificial Intelligence

Guildford, UK

2023 - 2025

## University of Surrey

B.S. in Computer Science

Guildford, UK

2018 - 2021

# Certifications

---

## AWS Certified Machine Learning Engineer - Associate

Amazon Web Services

2025

## AI Engineering Excellence

DataExpert

2025

# Technical Skills

---

- **Programming Languages:** Python, SQL, Java
- **Frameworks:** Pytorch, Tensorflow, pandas, scikit-learn, LangChain, LangGraph, Neo4j
- **MLOps & DevOps:** CI/CD, Git, Docker, GitHub Actions, MLflow, FastAPI
- **Cloud Platforms:** Amazon Web Services (AWS), Google Cloud Platform (GCP), DataBricks, Snowflake