

ARUSHI JAIN

London, United Kingdom | +44 7917304136 | jainarushi651@gmail.com |
<https://www.linkedin.com/in/arushijain1106> | <https://github.com/arushi1106>

ABOUT

AI Engineer and Software Engineer with an MSc in Artificial Intelligence (Merit, QMUL) and hands-on experience building LLM-powered applications, transformer-based NLP pipelines, and cloud-deployed AI services. Currently developing an autonomous multi-agent job search platform using RAG, vector search, and LangGraph. Proficient across the full ML lifecycle — from data normalisation and model fine-tuning to containerised deployment and CI/CD using Python, GPT-4, AWS, and Docker. Available immediately · UK Graduate Visa valid until 2028 · Open to sponsorship.

SKILLS

- **NLP & Transformers:** BERT, CodeBERT, CodeT5, spaCy, NLTK, NER, POS tagging, Hugging Face Transformers
- **LLMs & Fine-tuning:** GPT-4 API, OpenAI API, prompt engineering, text-to-SQL, LangChain, LangGraph, RAG pipelines
- **ML Engineering:** Scikit-learn, TensorFlow, PyTorch, RandomForest, BiLSTM
- **Deployment & MLOps:** Docker, AWS (RDS), Flask, REST APIs, GitHub Actions, Streamlit
- **Data & Databases:** Pandas, NumPy, PostgreSQL, MySQL, SQLite
- **Vector Databases:** Qdrant
- **Languages & Tools:** Python, SQL, Git

EXPERIENCE

Blocvue

London

Lead Design Engineer

July 2026 - Present

- **Sole frontend engineer** on a production compliance platform for UK higher-risk residential buildings, building and shipping the web application to pilot go-live across **23 buildings**.
- Designed and built the platform's **design system from scratch**, including shared components, design tokens and a written design brief adopted as the team's specification of record.
- Built **100+ routes** using **Next.js 15, React and TypeScript**, covering document registers, statutory obligations, risk assessments, asset registers and role-based access.
- Led a **54-item UI audit and remediation programme**, taking each issue from discovery through implementation and verification with code-level evidence.
- Diagnosed and fixed platform-wide defects including 24 broken form navigations, unsaved-work loss, containerised session failures and cache revalidation issues.
- Established a **verification-first development process**, independently validating changes on-screen and cross-checking AI-generated code to catch implementation errors.
- Identified and escalated a **database-layer authorisation gap** during a permissions audit, documenting remediation before pilot data entry.

Sigma.AI DATA Ltd.

London

AI Data Processing Analyst (Contract)

April 2026

- **Enforced** annotation protocols across large-scale speech corpora for ASR and NLP model training, identifying and escalating data quality issues to pipeline leads to maintain downstream model integrity across production datasets.

StatusNeo

India

Software Engineer

Jan 2024 – May 2024

- **Engineered** an end-to-end NL-to-SQL pipeline using GPT-4 and spaCy NER, enabling non-technical business teams to query complex relational databases in plain English, achieving 92% accuracy across finance and healthcare query sets.
- **Presented** technical architecture and model performance findings to non-technical stakeholders, translating NLP pipeline behaviour into business-impact terms that informed product decisions.
- **Optimised** a RandomForestClassifier via GridSearchCV, reducing model error by 15% and processing time by 20%; deployed on AWS via Docker with a GitHub Actions CI/CD pipeline, cutting release time by 50%.

EDUCATION

Queen Mary University of London, London, UK

Sept 2025

Master of Science, Artificial Intelligence - Merit

- **Relevant Modules:** Natural Language Processing, Machine Learning, Deep Learning, Applied Statistics, Neural Networks.
- **Thesis Project:** [AI-Powered Code Assistant](#): Fine-tuned CodeBERT and CodeT5 on CodeXGLUE for C code vulnerability detection, integrating a BiLSTM classifier and GPT-4 explanation layer; evaluated on precision, recall, and F1.

- **Relevant Modules:** Data Structures & Algorithms, Database Management Systems (DBMS), Software Engineering, Web Development, Cloud Computing.
- **Thesis Project:** [Personalised Health Treatment](#): Trained a multi-class symptom classification model on a labelled medical dataset, achieving 92% diagnostic accuracy on a held-out test set; surfaced structured precaution recommendations via a Flask web interface.

PROJECTS

[AI Job Application Copilot](#) | *Python, FastAPI, OpenAI API, LangGraph, Qdrant, PostgreSQL, React, Docker* | *June 2026 - July 2026*

- **Designed** and built an AI-powered job application platform combining LLMs, RAG, vector search, and LangGraph multi-agent orchestration.
- **Implemented** resume parsing, ATS optimisation, job matching, cover letter generation, and live job discovery via external APIs.
- **Built** a RAG pipeline with Sentence Transformers and Qdrant for semantic resume Q&A and contextual retrieval.
- **Delivered** a human-in-the-loop application lifecycle with approval workflows, status tracking, and Playwright-based submission automation.
- **Deployed** a full-stack solution with FastAPI, PostgreSQL, React, and Docker Compose.

[Spotify Reversed Playlist](#) | *OAuth 2.0, Spotify Web API, Flask* | *June 2026*

- **Designed** a backend automation tool handling custom OAuth 2.0 logic, token refresh cycles, and paginated JSON array processing to intercept playlist tracks and update active playback queues in real time.

[AI-Powered Code Assistant](#) | *CodeBERT, CodeT5, BiLSTM, GPT-4, Python* | *Feb 2025 - Sept 2025*

- **Fine-tuned** CodeBERT and CodeT5 on CodeXGLUE for C code vulnerability classification; integrated a BiLSTM sequence classifier and a GPT-4 explanation layer generating developer-readable vulnerability reports alongside predictions.
- **Documented** all experimental parameters and evaluation results (precision, recall, F1) in a reproducible pipeline, enabling full academic replication and clear communication of findings.

[AI-Powered Social Media Automation](#) | *OpenAI GPT, Flask, AWS RDS, Python* | *Jan 2024 - May 2024*

- **Developed** an automated content generation and scheduling system for LinkedIn and Instagram using GPT-4, cutting manual creation time by 50% over a 30-day benchmarked comparison period.

[Automated Text Structuring Pipeline \(NLP & Data Preprocessing\)](#) | *NLP, Spacy, Python, Flask* | *May 2023 - July 2023*

- **Built** a modular NLP pipeline (tokenisation, lemmatisation, NER, relation extraction), converting unstructured documents into structured database records, 35% improvement in downstream data usability vs. manual baseline

CERTIFICATIONS

- Certified in Python, Data Science, and Machine Learning.