

Raian Khan

London, UK | +44 7904 210204 | raiank2021@outlook.com | [linkedin.com/in/raian-khan](https://www.linkedin.com/in/raian-khan) | github.com/raian2021 | raian.uk

PROFESSIONAL SUMMARY

Junior AI Engineer with a strong infrastructure background, building production AI systems including fine-tuned LLMs, RAG pipelines, full-stack AI applications, and cloud-native ML inference platforms, alongside hands-on enterprise experience in AWS, Entra ID, and M365. Brings infrastructure depth that most AI candidates lack, with real deployment experience across global environments. Currently pursuing an MSc in Artificial Intelligence at the University of Liverpool (part-time, commenced 2026, Expected 2028) and AWS Machine Learning Associate certification.

PROFESSIONAL EXPERIENCE

IT Engineer Odgers Berndtson | London, UK

Jun 2025 - Present

- Engineered and maintained a hybrid AWS/Azure/M365 environment for 800+ users across EMEA, APAC, and AMER, responsible for identity, endpoint management, cloud infrastructure, and operational continuity.
- Identified a recurring operational bottleneck in ticket handling, scoped a solution end-to-end, and led delivery of an LLM-based triage system driving ~35% efficiency improvement.
- Built identity and endpoint automation across Entra ID, Intune, and M365 using Python and PowerShell, reducing manual workload by ~35% and improving provisioning reliability.
- Designed AWS EC2/IAM lab environments for pre-production testing; hardened access governance via Entra ID access reviews and updated CI/CD security baselines.

IT Support Engineer Integrating Solutions Limited | London, UK

2024 - 2025

- Automated provisioning workflows across Azure, AWS, Google Workspace, and M365 for 85+ MSP clients, cutting setup time by ~40% with ~98% success rate.
- Standardised documentation and runbooks across a multi-tenant environment, reducing recurring ticket noise by ~20%; collaborated with senior engineers to scope and deliver infrastructure improvements.
- Operated across Windows, macOS, Intune, VMware, and core networking in a multi-tenant cloud environment, translating client pain points into technical requirements.

IT Support Apprentice Integrating Solutions Limited | London, UK

2023 - 2024

- Completed Level 3 ICT Apprenticeship with Distinction. Managed 300+ onboarding/offboarding workflows and standardised asset lifecycle processes across multiple client organisations.

SELECTED PROJECTS

FinLit: AI-Powered Financial Literacy Dashboard | Python, FastAPI, Next.js, Claude API, PostgreSQL, Docker

- Built full-stack personal finance app ingesting Monzo/Starling bank data via CSV with auto-categorisation across 20+ spending categories.
- Designed privacy-first AI architecture: raw transactions processed locally, only aggregated summaries sent to Claude Haiku for cost-efficient insight generation (~0.11/user/month). Containerised with Docker Compose.
- Implemented compound growth simulator modelling opportunity cost of spending patterns over 1 to 30 year horizons.

Ticket Triage LLM: LoRA Fine-Tuning | Qwen2.5-0.5B-Instruct, LoRA/PEFT, PyTorch, FastAPI

- Fine-tuned Qwen2.5-0.5B-Instruct using LoRA adapters on ~800 synthetic samples for structured ticket classification. Deployed as a FastAPI microservice with ~95% schema-valid JSON output compliance.

RAG API Platform | FastAPI, PostgreSQL, pgvector, SentenceTransformers

- Built a RAG backend using 384-dim embeddings, IVFFlat-indexed pgvector store, and cosine similarity top-k retrieval at sub-100ms latency. Full API specs with request/response contracts.

Cloud-Native ML Inference Platform | AWS ECS Fargate, ECR, ALB, Terraform, GitHub Actions

- Building a production ML inference API on ECS Fargate behind ALB with full IaC via Terraform and CI/CD via GitHub Actions with AWS OIDC keyless auth.

JobOS: AI Job Application Tracker | Next.js, Supabase, Vercel, Claude API

- Built a full-stack job application tracker with AI-powered CV tailoring, cover letter generation, and application status management. Deployed on Vercel with a Supabase backend, reducing manual job search overhead significantly.

TECHNICAL SKILLS

AI / ML: LLM fine-tuning (LoRA/PEFT), RAG pipelines, PyTorch, HuggingFace Transformers, SentenceTransformers, scikit-learn, pgvector

Backend & APIs: Python, FastAPI, Next.js, REST API design, PostgreSQL, Docker, Terraform, GitHub Actions, PowerShell, CI/CD, Git

Cloud & Identity: AWS (ECS Fargate, ECR, ALB, IAM, EC2, CloudWatch), Azure AD / Entra ID, M365, Intune, Okta

Process & Delivery: Requirements definition, API contract design, agile delivery, stakeholder communication, technical documentation

EDUCATION & CERTIFICATIONS

MSc Artificial Intelligence University of Liverpool (Part-time, commenced 2026)

Expected 2028

AWS Machine Learning Associate In Progress

Expected 2026

Level 3 ICT Apprenticeship Distinction

2024