

KSHITIJ TAPALE

AI Engineer

📍 Oxford, UK ✉️ kshitijtapale.uk@gmail.com [🌐 LinkedIn](#) [🐙 GitHub](#)

[🌐 Portfolio: https://www.kshitijtapale.com/](https://www.kshitijtapale.com/)

Professional Summary

AI Engineer with 7+ years of software engineering experience, including 1.5+ years in production AI. Specialised in LLM-based systems, RAG pipelines, vector databases, and multi-agent architectures. Strong in Python, Azure/AWS, scalable APIs, and data pipelines. Proven impact in low-latency AI deployment, system optimisation, and reliable AI outputs through validation and fallback mechanisms. Developed data ingestion, embedding, and context retrieval systems that enhance the accuracy and trustworthiness of AI outputs.

Technical Skills

Programming:	Python (FastAPI, Django, Flask), JavaScript (React, Node), HTML, CSS
AI/ML:	LLMs, RAG Systems, NLP, LangChain, Hugging Face, Prompt Engineering, Confidence Scoring
Cloud & Azure:	Azure (AI Services, AI Search, OpenAI Service, Prompt Flow, AKS), AWS, Cloud Security Basics
DevOps & CI/CD:	Docker, Kubernetes, CI/CD Pipelines, Git (GitHub/GitLab), Terraform, Dynatrace
Data & ETL:	Pandas, Numpy, PySpark, Talend, ETL Pipelines, Large Dataset
Databases:	Vector Databases (Pinecone, ChromaDB), SQL, NoSQL, SQLAlchemy ORM

Professional Experience

Premedgo | AI Engineer

Oxford, UK | Nov 2025 – Present

- Design and operate an adaptive AI system using multi-agent LLM orchestration with multi-layer validation, handling 900+ categorised NHS records with 3-level classification logic.
- Architect hybrid RAG systems integrating SQL databases with vector search (Pinecone), delivering sub-500ms query response times across large datasets.
- Build model governance layers with confidence scoring and fallback logic for LLM outputs, ensuring data accuracy and predictable system behaviour.

Independent Consultant | AI Engineer

Oxford, UK | Jan 2025 – Sept 2025

- Built production RAG system using Claude API, ChromaDB, and multi-model embeddings, achieving 93%+ accuracy and reducing manual data interpretation time by 42% (10+ hours/week saved).
- Engineered advanced search and ranking pipelines combining multiple matching algorithms, improving data retrieval relevance across diverse query types.
- Designed fault-tolerant document ingestion pipelines with multi-stage parsing and OCR fallbacks, achieving 99%+ extraction success for complex documents.

Oxford Brookes University | Research Software Engineer

Oxford, UK | May 2024 – Nov 2024

- Developed ML-based data classification solution achieving 97.8% accuracy for URL threat detection across large-scale datasets.
- Optimised data ingestion and preprocessing pipelines with dynamic encoding detection and feature extraction, reducing processing time by 35%.
- Implemented adaptive learning mechanisms for evolving data patterns, improving system resilience by 23%.

Tata Consultancy Services | Systems Engineer

Pune, India | Mar 2022 – Aug 2023

- Reduced API latency by 32% through python upgrade, database schema optimisation and server tuning, improving response times across core services.
- Delivered 16 new features and resolved critical issues in database queries and data integrity logic, improving system reliability.
- Collaborated with ETL and BI teams to automate data workflows using Talend and Python, reducing manual effort by 80% (120 work-hours saved per quarter).
- Implemented performance monitoring solutions using Dynatrace, identifying system bottlenecks and reducing downtime.

Acloudzyn Technologies | Software Engineer

Pune, India | Feb 2020 – Mar 2022

- Built responsive UI/UX using React and Figma, improving customer engagement and cross-device compatibility.
- Developed secure OAuth2 authentication and 24 RESTful APIs using Django REST and SQLAlchemy ORM.

eClerx | Analyst

Pune, India | Mar 2019 – Apr 2021

- Designed reporting automation pipelines using Excel VBA and SQL, enhancing global ISDA compliance workflows saving 20+ minutes per task.
- Maintained compliance and precision in global derivative trades (IRD, CDS, FX, Equity) by working closely with international teams to increase operational efficiency.
- Played a significant role in resolving product discrepancies and financial breaks, ensuring transaction accuracy. Generated and meticulously validated diverse daily reports, upholding data integrity and reliability.

Achievements

- **UK Government Turing Scholarship** – Awarded for academic excellence in Advanced Computer Science.
- **TCS On the Spot Award** – Exceptional Project Delivery & Technical Innovation
- **TCS Xcelerate Warrior Award** – Outstanding Leadership in Agile delivery teams

Education

M.Sc. Advanced Computer Science

Oxford Brookes University, UK

Sept 2023 – Nov 2024

B.Tech. Automobile Engineering

Kolhapur University, India

Jun 2013 – Jun 2018