

Rahul Rathod

AI ENGINEER

Leeds, UK | raaulc@outlook.com | raaulc.com | [GitHub](#) | [LinkedIn](#)

SUMMARY

AI Engineer building and deploying production LLM systems at Flutter UK & Ireland, one of the world's largest online betting and gaming operators. Design, ship, and operate AI and automation solutions that serve millions of users and transform how trading and risk teams work day to day. Founded and lead an internal AI Accelerator that embeds AI into department workflows and upskills traders, analysts, and risk specialists. Combine applied GenAI depth (LLM integration, RAG, agentic workflows, evaluation, MLOps) with 16+ years of software engineering rigour delivering reliable systems at scale in fast-paced, regulated environments.

CORE SKILLS

Languages TypeScript (Node.js), Python, SQL

Agents & LLMs Agent harnesses, tool use and multi-agent orchestration, Model Context Protocol (MCP) servers and gateways, Claude Agent SDK, OpenAI Agents SDK, LangGraph, PydanticAI; context engineering, prompt caching, structured output; Anthropic Claude, OpenAI GPT, Google Gemini, open-weight models (Llama, Qwen, DeepSeek); fine-tuning (LoRA)

Recommendation, Search & Ranking Real-time personalisation, two-stage retrieval and learned ranking (LightGBM LambdaRank), hybrid search (BM25 + vector), embeddings, RAG, vector databases (pgvector, OpenSearch, FAISS)

Evaluation & Experimentation Offline metrics (NDCG, recall, coverage), A/B and champion/challenger testing, versioned eval sets with regression gates, LLM-as-judge, groundedness checks, guardrails, tracing (Langfuse, LangSmith)

Cloud & Platform AWS (Lambda, ECS/Fargate, API Gateway, DynamoDB, S3, Bedrock, SageMaker, CloudWatch, X-Ray), infrastructure as code (AWS CDK, Terraform), Docker, Kafka, Redis, FastAPI, PostgreSQL, CI/CD (GitHub Actions, Jenkins), OpenTelemetry, Grafana; latency budgets and load testing

Leadership AI transformation and enablement, technical leadership and mentoring, cross-functional delivery, stakeholder management

EXPERIENCE

AI Engineer – Automation, Transformation and AI

Jun 2026 – Present

Software Engineer

Jan 2023 – Jun 2026

Flutter UK & Ireland (UKI Trading)

Leeds, UK

- Design and ship AI and LLM-powered solutions that make traders measurably more efficient, automating manual, repetitive workflows across trading and risk operations.
- Deploy AI systems into production serving millions of users, owning the full lifecycle from problem discovery and prototyping through evaluation, release, monitoring, and iteration.
- Founded and lead the department's AI Accelerator, a programme that embeds AI into day-to-day work for all traders and risk analysts and drives measurable, department-wide AI adoption.
- Build agentic and retrieval-augmented workflows that connect LLMs to internal trading data and services, with evaluation and guardrails suited to a high-stakes, regulated domain.
- Lead AI transformation and enablement across the department: running upskilling sessions, defining safe-usage patterns, and turning high-value use cases into reusable, production-grade tooling.
- Partner with product, engineering, risk, and compliance to prioritise AI initiatives and ship solutions that are reliable, observable, and cost-aware rather than one-off proofs of concept.

Assistant Consultant (via TCS)*Marks & Spencer***Jan 2020 – Jan 2023***London, UK*

- Led delivery for Click & Collect and loyalty features across UK & ROI; integrated account, payments, and promotion APIs.
- Diagnosed and resolved a critical login regression impacting thousands of daily users; restored access and reduced support tickets.
- Worked closely with backend teams to consume Kafka-driven personalization feeds; partnered with analytics to enable data-informed product decisions.

Assistant Consultant (via TCS)*Best Buy***May 2017 – Dec 2019***Chennai, India*

- Delivered modules for product search, recommendations, and dynamic pricing; integrated with Java microservices and caching layers.
- Enhanced responsiveness by optimizing API usage and background tasks; improved stability via monitoring and targeted fixes.

Technical Lead*Capgemini***May 2012 – Jul 2016***Mumbai, India*

- Delivered enterprise-grade mobile apps for clients including 3M, ING, and Dako, achieving 4.5+ app store ratings on average.
- Managed distributed teams and QA/release pipelines, reducing post-release issues by 30% through performance and stability work.

Software Engineer*Persource***Apr 2011 – Jun 2012***Ahmedabad, India*

- Built 15+ apps for global clients including Yahoo and Sundance Film Festival; owned full-cycle delivery from design to release.

Software Engineer*BYPT***Sep 2010 – Apr 2011***Ahmedabad, India*

- Developed live sports scoring apps with real-time updates and performant UI, recognized for usability and stability.

SELECTED AI PROJECTS

Independent, shipped projects — full write-ups at raaulc.com/projects.

Real-time Game Recommender — Two-stage recommender (item-kNN and SVD retrieval, LightGBM LambdaRank ranker) taken from a Python research prototype to a TypeScript service. Ranks 23% better (NDCG@10) than the best single model on 3.15M plays from 53k players, at 3.87 ms p99. Updates from play events in real time, applies cooldown, exclusion and diversity guardrails, explains every pick, and is tested for parity with the prototype. Live demo at raaulc.com/projects/realtime-recs. *TypeScript, Node, Python, LightGBM, SciPy*

AI Paper Radar — Daily ranking pipeline over new arXiv AI papers: scores every paper for relevance and novelty with a typed-decision model, ranks them, and publishes a browsable archive. Runs unattended on a scheduled GitHub Actions workflow with retries and per-category fault tolerance. *Python, Jev, GitHub Actions, arXiv API*

Self-RAG Pipeline — Stateful LangGraph workflow where the LLM grades its own retrieved documents, checks for hallucinations, and validates its answer before responding. *LangGraph, LangChain, Chroma, Groq, HuggingFace*

EDUCATION

Bachelor of Engineering, Electronics & Communication*Dharmsinh Desai University — Gujarat, India***May 2009**