

Ryan Perkins

r.perkins925@gmail.com | github.com/codingperks

AI engineer building production LLM agents; consulting background with client-facing technical delivery across enterprise and regulated industries.

Skills

AI Engineering (LangGraph, MCP, RAG, LLM-as-judge evals), **Engineering** (Python, FastAPI, PostgreSQL, Docker, Kubernetes), **Cloud** (AWS, Azure OpenAI), **Observability** (LangFuse, Grafana)

Experience

AI Engineer, Ten Lifestyle Group – London, UK Feb 2025 – Present

- Built Talia, a LangGraph multi-agent WhatsApp concierge deployed to enterprise and private member cohorts; 130+ production conversations, rolling out to 3 regions and 3 enterprise clients.
- Designed Talia's agent orchestration, planner workflows, tool execution, and subgraph architecture, cutting typical response latency from 52s to ~35s via prompt caching and tighter context management.
- Created the company's first agent evaluation framework, using 50+ labelled customer conversations, golden datasets, execution traces, and prompt versioning to regression-test agent behaviour.
- Developed live LLM-as-judge metrics for user frustration and task quality, surfacing production health through Grafana dashboards and Slack alerting when behaviour drops below thresholds.
- Built and deployed a shared MCP tool server with FastMCP, standardising tool interfaces, error contracts, and server architecture across 5 product tools and 25 DevOps/Kubernetes tools.
- Rolled out LangFuse observability and onboarding playbooks across four AI products and nine engineers, enabling trace-level debugging, message replay, and incident investigation.
- Evaluated and extended an internal browser-use PoC for restaurant booking automation, assessing reliability on dynamic UIs and authoring a production-readiness recommendation.

Senior AI Engineer, Capita – London, UK Sep 2023 – Feb 2025

- Delivered AI engagements as a technical consultant for enterprise clients across utilities, broadcasting, and the public sector; ran scoping and live requirements gathering with client stakeholders and their end-user delivery teams, product demos, and implementation support across the engagement lifecycle.
- Ran live product demos of AgentSuite, an LLM-powered contact centre platform, with utility client stakeholders during paid pilot, addressing technical questions and contributing to conversion into full deployment across two national utility providers.
- Implemented AgentSuite's backend and prompt engineering for live-call RAG retrieval and QA automation, cutting call handling times by 20%; worked directly with client operational SMEs to shape agent workflows.
- Scoped, designed, and delivered Capita's first RAG pipeline over 50,000+ internal documents for Capita's legal business unit; led discovery sessions with stakeholders, secured internal R&D funding, and cut clerical due-diligence effort by 90%.
- Proposed the accepted architecture for an AI complaint automation system for a major broadcaster, presenting trade-offs and implementation plans to the operational delivery team and Capita leadership.

Senior Economic Consultant, London Economics – London, UK Oct 2018 – Jun 2022

- Led £500k+ data and technology consulting engagements end-to-end: client scoping, stakeholder workshops, executive presentations, and delivery planning for senior decision makers across government and industry.

Projects

GeoHarness: Agent Architecture Experiment (github.com/codingperks/geoharness) – Pure-Python ReAct agent over geospatial data sources (NASA POWER, OpenTopography), built to test whether primitive-tool architectures can match graph-based agents on complex tasks. Hand-labelled eval set of 11 locations with documented scoring; Langfuse tracing across the act/observe/reflect loop.

Education

Imperial College London – MSc Computing, Distinction Sep 2023

University College London – MSc Economics, Merit Sep 2018

Royal Holloway, University of London – BSc Economics, First Class Jul 2017