

Alex Lewis

Senior AI Engineer & Architect | PhD Astrophysics | LLM Systems · High-Performance Backend

ajrlewis.com | hello@ajrlewis.com | +44 7535 791970 | linkedin.com/in/ajrlewis | github.com/ajrlewis

PROFESSIONAL EXPERIENCE

Senior Software Engineer (Contract)

Oct 2025 – Mar 2026

Digital Speed

London, UK

Drove the development of high-impact core libraries and AI-integrated pipelines to accelerate agency delivery.

- Developed core Java libraries within an existing CI/CD pipeline to support multi-language SDK generation; integrated ISO 20022 (PAIN.001) validation and secure XML preprocessing with a focus on cross-language logic reuse.
- Built Speedy, an internal Slack-based agent in TypeScript/Next.js using the Vercel AI SDK, handling team operational queries through a multi-step tool-calling loop over handbook data (pgvector) and live integrations (GitHub via MCP, Productive.io); added structured logging and evals to iterate on response quality.
- Engineered an 'AI Skills' framework enabling Claude and Cursor agents to ingest Markdown-based procedural knowledge, standardizing design and development workflows across the agency.
- Delivered rapid client-facing prototypes using React/Vite, transitioning validated concepts into production-ready Next.js applications with SSR and SEO considerations.

AI Solutions Architect (Contract)

Aug 2025 – Nov 2025

Downing LLP

London, UK

Architected and delivered a Python-based, Azure-native, model-agnostic AI orchestration platform (LangGraph) to automate internal support triage.

- Engineered a manifest-driven multi-agent framework enabling dynamic registration of department-specific bots, establishing the architecture for enterprise scaling.
- Delivered a production-ready Phase 1 pilot with multi-provider switching (Anthropic, Google, OpenAI) and HITL validation via Microsoft Teams, supporting internal digital team workflows through automated query classification.
- Implemented secure, thread-aware state management using Azure PostgreSQL and Managed Identity, ensuring reliable conversation persistence and traceability.
- Centralized prompt management and evaluation using LangSmith, improving debugging visibility and iteration speed.
- Established a gated, trunk-based CI/CD pipeline in Azure DevOps, enabling progression from R&D to UAT readiness.

Technical Founder

Feb 2025 – Jul 2025

W4lkies Ltd

London, UK

Built open-source W4lkies (w4lkies.com), a custom operations platform for a small pet-care business, streamlining bookings, customer/dog profiles, invoicing, and expenses.

- Designed and deployed a full-stack application (Next.js, FastAPI, PostgreSQL, Docker), replacing manual/admin-heavy workflows with a unified system.
- Streamlined day-to-day operations through booking management, history tracking, and automated invoice generation, reducing time spent on admin tasks.
- Implemented authentication and role-based access to support multiple users and safeguard customer data.
- Developed lightweight operational tooling (seed data, backups) to improve reliability and maintainability of the system.

Lead AI Architect (Contract)

Mar 2024 – Jan 2025

Hyperhumans OÜ

Remote

Architected the Neuralead lead-intelligence API, delivering a scalable platform for automated company discovery and multi-source LLM enrichment.

- Engineered complex multi-stage enrichment pipelines integrating Bing Search, OpenAI, and Hunter.io to automate high-volume contact data extraction and lead scoring.
- Led the architectural migration from a Flask prototype to a high-performance FastAPI backend, implementing robust Pydantic schema validation to ensure 99.9% API contract reliability.
- Designed and implemented a commercial-grade API monetization engine, featuring credit-metering and Role-Based Access Control (RBAC) to support tiered subscription models.

- Optimized large-scale data access patterns and pagination using SQLAlchemy, significantly reducing response latency for downstream consumers of high-density datasets.
- Contributed to production deployment workflows through pull requests and main-branch integration, working within CI/CD pipelines supporting deployment to Google Cloud Run.

Head of Optimization (Lead Backend/Data Engineer)

Mar 2019 – Feb 2024
London, UK

Itim Ltd

Architected and delivered high-throughput forecasting systems powering supply chain decisions for tier-1 retailers, specializing in performance optimization and backend architecture under constrained infrastructure.

- Engineered high-performance FastAPI services delivering 1M+ daily SKU/store forecasts, selectively applied to high-signal inventory to optimize compute efficiency at scale.
- Developed a SQL-based data triage layer using CTEs and stored procedures to classify inventory by signal strength, routing low-signal items to lightweight in-database forecasting and reserving ML capacity for high-impact SKUs.
- Improved forecast stability in highly variable retail environments by implementing rolling-window retraining with Holt-Winters models, continuously updating parameters to reflect recent demand patterns and mitigate model drift.
- Designed and deployed end-to-end pipelines spanning data ingestion, triage, model execution, and persistence of 98-day forecasts in SQL Server, with aggregated views for downstream consumption.
- Architected advanced ML prototypes for dynamic pricing, markdown optimization, and recommendation engines to support data-driven retail strategies.
- Led technical delivery across a distributed team (London/Pakistan), establishing engineering standards, hiring and mentoring a junior engineer who successfully assumed ownership of the system after my departure.

TECHNICAL RESEARCH & OPEN SOURCE

AI Contract Risk Review

A distributed legal-tech pipeline for automated clause extraction and risk detection. Built with LangGraph and Celery for async multi-stage processing, utilizing Docling for OCR on complex PDFs and MinIO/S3 for persistent document storage and retrieval.

GitHub: github.com/ajrlewis/employment-checker

Tools: [LangGraph](#) · [FastAPI](#) · [Celery](#) · [Redis](#) · [MinIO](#) · [Next.js](#)

Competitive Musical Strategy

A low-latency, turn-based music recognition system built with Next.js. Implemented secure server-side audio proxies and complex game-state logic to prevent client-side data leakage during competitive play.

GitHub: github.com/ajrlewis/passthetrack | Website: passthetrack.com

Tools: [Next.js](#) · [TypeScript](#) · [Vercel](#) · [Serverless](#) · [Tailwind](#)

EDUCATION

The University of Edinburgh

Aug 2014 – Nov 2018

PhD in Astrophysics

- First-author publication in The Astrophysical Journal (2018) on early-universe galaxy clusters; PhD fully funded by the European Research Council (ERC) and supervised by the Head of Science at ESO.
- Engineered custom analysis workflows for large-scale sub-millimetre survey data (Herschel, JCMT, APEX), implementing Monte Carlo simulations and empirical noise estimation to isolate faint signals in crowded fields.
- Developed SED modelling and clustering pipelines to infer galaxy properties and detect statistically significant structures in sparse multi-band datasets.

The University of Edinburgh

Sep 2008 – Aug 2014

BSc in Astrophysics

- Completed core studies in physics, cosmology, statistical data analysis, and scientific computing.

TECHNICAL SKILLS

Languages & Core: Python (FastAPI, SciPy, Pydantic), TypeScript / JavaScript, Java

AI & Orchestration: LangGraph & LLM Orchestration, RAG Systems (pgvector), Model Context Protocol (MCP), LangChain / LangSmith

Backend & Data: PostgreSQL & SQLAlchemy, API Design (FastAPI / REST)

DevOps & Cloud: CI/CD Workflows (GitHub Actions, Azure DevOps), Cloud Platforms (Azure, GCP Cloud Run)

Frontend: Next.js (App Router), React / Tailwind CSS