

ANDREI CRUCERU

Forward-Deployed AI Engineer | Applied AI Systems Architect | Machine Learning Engineer

London, UK | +447919250103 | andrei.cruceru95@gmail.com | [LinkedIn](#) | [GitHub](#) |

SUMMARY

Forward-deployed AI Engineer and Machine Learning specialist with 2+ years of production experience building enterprise AI, ML, and data products across multiple industries. I work directly with clients to diagnose bottlenecks, design cost-effective solutions, model AI usage and infrastructure costs, and deliver production systems end to end.

Lead developer of the core engine behind DQP, an enterprise data quality platform with 300+ tools for profiling, validation, anomaly detection, workflow automation, AI agents, custom ML, and human-language insight generation. Experienced in turning ambiguous business needs and AI prototypes into scalable, low-latency, cost-efficient systems with measurable operational value.

TECHNICAL SKILLS

- **Programming & Query Languages:** Python, Java, C#, SQL
- **AI Engineering:** Generative AI, AI Agents, RAG, tool-calling workflows, LangChain, LangGraph, MCP, OpenAI API, knowledge retrieval, human-in-the-loop automation, prompt-driven execution flows
- **ML Engineering & NLP:** PyTorch, Transformers, Scikit-Learn, ONNX Runtime, NLP, anomaly detection, custom ML model development, model testing and evaluation
- **Data Engineering & Analytics:** Pandas, NumPy, data profiling, data quality automation, workflow automation, Microsoft Fabric
- **Backend API & Deployment:** FastAPI, REST APIs, Docker, Git, production deployment, API integration
- **Cloud & Infrastructure:** Azure, AWS, GCP, cloud infrastructure design, compute cost estimation, AI usage cost estimation
- **Solution Architecture & Delivery:** client discovery, bottleneck analysis, technical scoping, solution architecture, implementation planning, latency-aware system design, cost-efficient AI system design

WORK EXPERIENCE

PowerServe Global, ML & AI Engineer, London, UK 2024-Current

- Own end-to-end delivery of data science, ML, and AI product workstreams, from client discovery, bottleneck analysis, solution architecture, and cost estimation through to AI workflow development, API integration, deployment, reporting, and client-facing delivery.
- Built AI-enabled workflow systems using Python, LLM orchestration, RAG-style retrieval, structured routing, tool-calling, and human-in-the-loop controls to automate complex business processes while reducing hallucination and execution risk.
- Designed and deployed production ML and AI pipelines across legal, finance, maritime, retail, and hospitality use cases, processing millions of rows on low-cost architectures and writing outputs back into client databases for BI and operational reporting.
- Mentor and manage junior ML engineers by planning workstreams, reviewing technical delivery, supporting architecture decisions, and improving execution quality across client and product projects.
- Produced implementation plans and cost models for AI solutions, estimating development effort, infrastructure requirements, compute usage, LLM/API usage, latency constraints, and deployment trade-offs for client approval.

Profitero, Data Analyst Intern, Maidenhead, UK 2020-2021

- Developed data pipelines for large-scale ecommerce datasets covering millions of products across hundreds of retailers in Europe, the US, and Southeast Asia, supporting analytics for major retailers and global consumer brands.
- Used Python, SQL, Excel, Power BI, and NLP techniques to support product, pricing, availability, and competitive intelligence analysis across large retail datasets.
- Completed the internship between first and second year of university and secured a post-graduation job offer based on performance.

Earlier Career Experience - Project Management

2016-2020/2021-2023

- Managed client-facing projects across pricing, scoping, planning, scheduling, team coordination, execution monitoring, and issue resolution. Produced job estimates, coordinated operatives, managed delivery timelines, handled client communication, and ensured work was completed within operational and budget constraints.
- Built strong foundations in systematic planning, cost estimation, stakeholder management, delivery ownership, and practical problem-solving, now applied to AI, ML, and enterprise software delivery.

PROJECTS & LEADERSHIP EXPERIENCE

DQP, *PowerServe*

2025-2026

- Designed, developed, and deployed the core engine of DQP, an enterprise data quality platform combining data pipelines, custom ML models, profiling systems, validation logic, anomaly detection, AI agents, and 300+ custom data quality tools.
- Built specialised modules for cleaning, standardisation, validation, risk analysis, issue detection, anomaly reporting, and human-language insight generation across high-value and sensitive enterprise data, including addresses, emails, phone numbers, and other structured and semi-structured fields.
- Designed asynchronous processing workflows and scalable data pipelines capable of handling thousands of rows per second, depending on cleaning, profiling, and analysis scope.
- Currently extending the platform with an agentic workflow and MCP integration layer, enabling users to create and execute data quality flows from natural-language prompts while exposing platform capabilities as callable tools for enterprise systems.

Legal Enquiry Automation System, *PowerServe*

2025-2026

- Designing and developed an end-to-end client onboarding and legal enquiry triage system that processes inbound enquiries, filters relevant emails, classifies case type, routes requests, and prepares fee-earner assignment in minutes instead of days.
- Built the system around custom ML inference engines for email filtering, department and case classification, cost and time forecasting, and fee-earner recommendation, reducing reliance on open-ended LLM reasoning and minimising hallucination risk.
- Automated assistant-heavy workflows including enquiry understanding, follow-up question generation, structured case recording, department-head alerts, customer updates, and human-in-the-loop approval for final assignments.
- Integrated retrieval over relevant laws and UK court cases to provide assigned fee earners with contextual case support at the point of assignment.
- Estimated commercial impact included 50% higher onboarding capacity and £300k+ first-year cost reduction through reduced manual triage, faster routing, and improved operational throughput.

Product Classification System, *PowerServe*

2024-2025

- Developed and deployed a transformer-based product classification system for a global maritime client, turning unstructured product descriptions into structured categories across 600+ classes.
- Designed the full ML delivery flow from data gathering, labelling analysis and experimentation through to ONNX Runtime deployment behind a .NET API.
- Supported over 2M predictions per month and achieved 95%+ F1 score on the top 85% of products, improving automated product understanding at enterprise scale.

EDUCATION

- **BSc Data Science (First Class Honours) | Buckinghamshire New University, High Wycombe, UK**