David A'Hearne

07827 668818 · david.ahearnex@gmail.com · linkedin.com/in/david-ahearne-767614b3 · github.com/DaveAhearne

PROFILE

ML Engineer and engineering leader with 13 years of experience building production systems at scale. I design and ship applied ML services, from NLP scoring pipelines to LLM-integrated APIs, with proper MLOps tooling, not just notebooks. I am currently completing a BSc in Mathematics at the Open University alongside a Cambridge Data Science Career Accelerator, grounding practical ML work in solid theory. I care about reproducible pipelines, pragmatic architecture, and systems that hold up under real-world load.

SKILLS

Languages	Python, Go, C#, SQL
Applied ML	spaCy, HuggingFace Transformers, BERT, sentence embeddings, cosine similarity, OpenAI API, Ollama, RAG pipelines, vector search (sqlite-vec), Stanza NER
Data Science	scikit-learn, XGBoost, Keras, TensorFlow, PyTorch, pandas, numpy, statsmodels, BERTopic
MLOps	MLflow, FastAPI, Docker, AWS ECS, AWS ECR, GitHub Actions CI/CD, pytest
Cloud	AWS (ECS, ECR, S3, IAM), Azure
Architecture	Distributed systems, microservices, CQRS, hexagonal architecture, event-driven design
Leadership	Team building, hiring, mentoring, cross-functional delivery, incident management

EXPERIENCE

Keenu.io — ML Engineer (Placement) *Jan 2026 – Present*

Paid engagement via the Turing Innovation Catalyst to build Keenu.io, an NLP-based IELTS assessment system. Designed and shipped two production systems: Gradia, an automated writing assessment service, and MarkVerify, a scoring platform for inbound test data used as the evaluation set.

Gradia — IELTS Writing Assessment API

- Designed an NLP scoring pipeline using spaCy word vectors and cosine similarity across eight subscores: paragraph cohesion (average pairwise sentence similarity within paragraphs), sentence-to-document centroid similarity (min and max), lexical richness, vocabulary sophistication, spelling error count and category, as well as punctuation placement and type.
- Integrated a HuggingFace BERT-based punctuation restoration model for grammatical analysis, precomputed and cached at startup to keep inference latency low under concurrent load.
- Built a routed LLM client supporting OpenAI (GPT-4o-mini) and Ollama (Qwen 2.5 7B) as interchangeable backends, with structured prompt versioning iterated against a 100-sample student IELTS evaluation dataset. Applied linear regression to calibrate raw LLM scores, achieving 94% of marks within +/- one band and 86% within +/- half a band. Evaluated using MAE and QWK.
- Implemented concurrent subscore computation via ThreadPoolExecutor, allowing all eight NLP subscores to run in parallel per request.
- Set up MLflow for experiment tracking and dataset versioning, and built a full CI/CD pipeline using GitHub Actions: running tests, building Docker images, pushing to AWS ECR, and deploying to AWS ECS. Configured all AWS infrastructure including IAM roles, ECR repositories, ECS task definitions, services, and load balancer health checks from scratch.

MarkVerify — Evaluation Scoring Platform

- Built in Go, MarkVerify is a platform for collecting and scoring inbound IELTS test submissions from real students, used as the evaluation dataset for Gradia. It provided a controlled pipeline for ingesting, storing, and human-scoring test responses to produce the ground-truth labels MLflow experiments ran against.

BrightHR — Delivery Lead *Sept 2023 – Dec 2024*

Owned the BrightSafe product end to end, from ideation and prioritisation through delivery across five countries: Australia, New Zealand, Canada, Ireland, and the UK.

- Hired and led a cross-functional team of 9 to 15 engineers, designers, and QA. Responsible for development plans, 1:1s, mentoring, and performance management.
- Delivered third-party identity server integrations, a rebuilt documents platform, the engineering blog, and back office tooling improvements.

- Acted as the primary interface between engineering, product, and delivery stakeholders, shaping work into deliverable increments while balancing technical constraints and business priorities.
- Produced early-stage prototypes and architecture diagrams using microservices and CQRS patterns. Led live production incident response across all five operating regions.

Schneider Electric — Contract Engineer *May 2023 – Sept 2023*

Contract role on the EcoIQ smart heating platform, a distributed .NET system with teams in France and China.

- Collaborated with lead engineers to modernise shared .NET Standard libraries across teams.
- Worked with embedded systems engineers to design, build, and test integrations with Panasonic heat pumps using .NET 6, applying TDD to ensure reliable communication between backend services and physical hardware.

OpenMoney — Tech Lead *Aug 2020 – May 2023*

Built and led the investments engineering team from the ground up, working closely with product and stakeholders to shape technical direction and identify milestones.

- Hired and technically mentored the team. Architected greenfield solutions, created technical prototypes, and established logging, alerting, and incident post-mortem processes.
- Ran company-wide talks and 1:1 mentoring on TDD, BDD, trunk-based development, feature toggling, safe refactoring, and CI/CD.
- Developed strategies for removing technical debt and ensuring quality delivery through integration, end-to-end, and contract testing.

Raytheon — Senior Software Engineer *Sept 2018 – Aug 2020*

- Worked across multiple teams refactoring monoliths into microservices, applying TDD and industry-standard agile practices. Languages and tools used included C#, Java, JavaScript, Python, Terraform, and AWS.
- Collaborated with product owners to identify thin vertical slices of value, delivering through pairing, mobbing, and independent work.
- Mentored junior and graduate engineers through code reviews, pairing, and leading department Campfire talks.

Zen Internet — Systems Developer *April 2016 – Sept 2018*

- Built and maintained internal and public-facing websites and APIs using C#, MVC, React, Node.js, and .NET Core, supporting order tracking, diagnostics, and network management systems.
- Prototyped AWS Lambda, Azure, and .NET Core to inform technical direction. Refactored legacy code in an agile environment with rapid feedback cycles.

Earlier: Swinton Insurance (Junior C# Developer, Jun 2013 – Apr 2016) Parker Sandfords (Junior Developer, Oct 2012 – Jun 2013)

PROJECTS

Nightshift — Clinical Code RAG Pipeline *Oct 2024 – Present*

Employer project for NICE (National Institute for Health and Care Excellence), built as part of the Cambridge Data Science Career Accelerator. Given a free-text clinical research question, the system parses it into typed clinical entities, retrieves and ranks relevant codes across SNOMED CT, ICD-10, QOF, and NHS reference sets, and returns results with source attribution and confidence scoring.

- Built a two-stage RAG pipeline: an offline ingestion stage encoding clinical code descriptions into biomedical embeddings (SapBERT/BioBERT) stored in sqlite-vec, and an online query stage combining semantic vector search with TF-IDF hybrid retrieval, reranking, and LLM-assisted reasoning.
- Used Stanza NER (bc5cdr and i2b2 biomedical models) to decompose free-text clinical queries into typed entities (conditions, drugs, exclusions), driving structured retrieval across multiple code systems simultaneously.
- Built a FastAPI service with WebSocket streaming of pipeline progress to the frontend, so results appear incrementally as each stage completes rather than after a full round-trip.
- Set up MLflow evaluation tests tracking F1, precision, and recall against gold-standard question-to-code-set pairs, with an acceptance threshold of 0.70 across all three metrics and confidence scoring above 0.65.
- Dockerised with automatic ingestion on first boot and Stanza model caching baked into the image layer to keep startup time predictable across environments.

EDUCATION

BSc (Hons) Mathematics — Open University *Sept 2024 – Present*

Cambridge Data Science Career Accelerator — University of Cambridge, PACE *Oct 2025 – Present*

BSc (Hons) Computing — 1st Class Honours *Sept 2013 – Jun 2014*