

Karim Arafa

Dubai, UAE · arafakarim05@gmail.com · linkedin.com/in/karim-arafa05 · invenro.co · tarif.ae

SUMMARY

AI & Computer Science graduate, University of Birmingham Dubai — **First Class Honours, 4.0/4.0 GPA** — with production experience shipping a multi-tenant AI SaaS to a paying UAE retail client. Built end-to-end systems spanning embedding pipelines, RAG agents, async insight engines, and transactional data infrastructure. Seeking graduate AI/ML engineering roles.

EDUCATION

University of Birmingham — B.Sc. Artificial Intelligence & Computer Science Dubai · July 2026
First Class Honours (4.0/4.0 GPA). Final-year dissertation: Invenro.
Relevant modules: Neural Computation, Natural Language Processing, Human-Computer Interaction.

PROFESSIONAL EXPERIENCE

DataFlow — Software engineering intern Sept 2026 – present
Dubai, UAE · Primary Source Verification (PSV) provider

- **Arabic AI agents** (Claude on AWS Bedrock) orchestrated by AWS Step Functions to automate verification workflows.
- **Full-stack development** of the core verification platform in Next.js and Node.js on AWS.
- **Playwright-based browser automation** for the agent pipeline.

Invenro — Founder & full-stack engineer Nov 2025 – present
Dubai, UAE · invenro.co

- **Multi-tenant inventory SaaS shipped to a paying UAE retail client** managing 600+ active SKUs across 5 sales channels (Amazon, Noon, Noon Minutes, Sharaf DG, Carrefour), unifying disparate channel data into a single "golden record" with full tenant isolation and role-based access control.
- **4-step product identification waterfall** (internal SKU → platform listing memory → barcode → lazy create) paired with a parallel pgvector cosine-similarity merge engine that flags cross-channel duplicates above an 85% threshold for human review, eliminating manual SKU reconciliation.
- **Sentinel insight engine**: two-tier async system on Upstash QStash webhooks — Tier 1 runs SQL-based anomaly detection (dead stock, restock-adjusted velocity spikes/drops), Tier 2 enriches findings via LLM — with composite-key deduplication and debounced scheduling for idempotency.
- **Streaming conversational AI agent** (AI SDK + OpenRouter) with 7 tool-calling functions over inventory, insights, trends, and product details, plus a RAG knowledge base (PDF, DOCX, TXT, MD) with semantic chunking and pgvector retrieval.
- **Transactional upload-undo system** capturing pre-upload inventory snapshots as JSONB for single-transaction reversal of CSV/XLSX imports; FIFO batch costing and restock-adjusted velocity tracking with historical backfill.
- **Data-correctness campaign** — 36 pull requests and ~95 permanent regression tests until dashboards, ledgers, and the AI agent agreed on every number.

Vieu — ML & software engineering intern June – Sept 2025
Seattle, WA (remote)

- **~50% API cost reduction**: developed a hybrid GPT-4.1 + rule-based degree-classification pipeline that cut token usage by 51%.
- **Student capability engine**: vector embeddings and k-means clustering to map semantic similarities between degree programs, improving student–alumni matching accuracy.
- Delivered production-ready Python and documentation into the core product.

PROJECTS & COMPETITIONS

Tarif — Solo build; the UAE driving-cost app 2026 · v1 complete, unreleased

- Prices any UAE route to the dirham before driving it: deterministic Salik/Darb toll engine (geometric gate detection), monthly official fuel pricing, and budget forecasting; 1,000+ automated tests. Marketing site live at tarif.ae (Lighthouse 100 in every category), waitlist open.

AI Voice Financial Advisor — Alibaba Cloud Hackathon, 2nd place Sept 2025

- 2nd place in Dubai, building an AI voice financial advisor in 24 hours exclusively on the Alibaba Cloud AI stack (Qwen LLMs, speech/embedding APIs); owned onboarding and the document-ingestion pipeline enabling the agent to reason over user-uploaded financial statements live in conversation.

TECHNICAL SKILLS

AI / ML	OpenAI API, vector embeddings, pgvector, RAG pipelines, LLM tool-calling, k-means clustering, AI SDK, OpenRouter, Alibaba Cloud AI stack (Qwen)
Programming	Python, TypeScript, JavaScript, SQL, C++, Java
Backend & DB	Node.js, PostgreSQL (Neon), Drizzle ORM, Better Auth, Upstash QStash
Frontend	Next.js (App Router), React 19, Expo / React Native, Tailwind CSS, Astro
Tools	Git, Docker, Vercel, GitHub Actions, Playwright
Languages	Arabic (native), English (fluent)