

Professional Summary

Senior AI Engineer / Tech Lead with 5+ years in software engineering & 2+ years delivering production RAG systems on Azure & AWS. Built bilingual (Arabic / English) document intelligence & RAG pipelines using Azure OpenAI, LangChain, FastAPI, Azure AI Search, & vector databases for government & enterprise clients across the GCC. Optimized latency & cost through semantic chunking, prompt/answer-length tuning, & caching; implemented custom safety guardrails for regulated environments. Strong full-stack foundation & client-facing delivery.

Core Competencies

AI & LLM Systems: End-to-end RAG architecture & agentic workflows | LLM integration & evaluation | Prompt & retrieval optimisation | Document intelligence & semantic search | Safety & compliance guardrails

Engineering Stack: Python | LangChain & LangSmith | Vector databases (PostgreSQL + pgvector) | Azure AI Search | Docker & Kubernetes | Cross-cloud deployment (Azure & AWS) | FastAPI | Payload CMS | Frontend frameworks (React/Next.js, Angular)

Domain & Delivery: Technical ownership & architectural decision-making | Multilingual AI systems (Arabic/English) | Client-facing demos, workshops & roadmap discussions | Cross-cultural stakeholder management & team leadership

Verofax — Senior Tech Lead Engineer

Abu Dhabi, UAE (Jun 2023 – Present)

- **Architected & deployed production RAG systems:** Led design & deployment of RAG chatbots using LangChain & Azure OpenAI, processing hundreds of bilingual (English/Arabic) documents.
- **Optimised latency, cost & accuracy:** Improved responsiveness by **~75%**, reducing **end-to-end latency** from **~4–5s to ~1s** (worst-case long responses previously **>10s**) using semantic chunking, retrieval tuning, prompt/answer constraints, & multi-layer caching; reduced token usage & operating cost through answer-length control & prompt compaction.
- **LLMOps & observability:** Instrumented tracing & monitoring (LangSmith + structured logs) to track latency, token usage, retrieval hit rate, & errors; enabled faster debugging & safer prompt/retrieval iterations.
- **Scaled vector DB & search infrastructure:** Built & maintained retrieval infrastructure with PostgreSQL + pgvector & Azure AI Search, deployed via Docker + AKS with production hardening (timeouts, retries, health checks, structured fallbacks)
- **Custom embeddings & on-prem LLMs:** Architected & deployed custom embedding models tailored for Arabic semantic search; deployed & operated open-source local LLMs for on-premises deployments to meet government privacy requirements.
- **Governance & safety:** Implemented client-specific compliance guardrails (content filtering, PII rules, refusal/hand-off behavior, audit logging) for regulated environments.
- **Leadership & client impact:** Led & mentored a team of engineers across RAG architecture, prompt engineering, & production deployment; partnered with C-level stakeholders on AI roadmap, ROI, & delivery plans; delivered Arabic/English demos at **GITEX (2022–2025)**, **Biban 2024**, **KKUH Medical City**, supporting multiple deals.

- **Revamped traceability platform:** Modernised a traceability platform using Angular & Material Design; improved responsiveness & UI/UX, increasing client confidence & helping close multiple deals.
- **Led World Cup “Scan & Win” campaign:** Built the front-end for FIFA World Cup Qatar 2022 “Scan & Win” across 30+ clients, generating thousands of scans & high-profile endorsements.
- **Analytics & full-stack work:** Developed real-time analytics dashboards to monitor demographics & engagement; built features across Angular front-end, Spring Boot back-end & computer-vision models; stepped into mobile development (Flutter/React Native) when needed, & mentored incoming juniors.

Netgreen — Full-Stack Developer *(Concurrent with Verofax)*

(Mar 2022–Jun 2023)

- **End-to-end platform delivery:** Led development of a platform allowing users to fund tree-planting & mangrove projects with digital certificates; built the React front-end from Figma designs & a Node.js/Express back-end.
- **Payments & authentication:** Integrated Stripe payment processing for multiple client/vendor configurations & implemented OAuth-based social logins.
- **Team mentorship:** Mentored a junior developer on back-end architecture & MongoDB, ensuring smooth knowledge transfer.

Education

- **Lebanese International University** — M.Sc. in Computer Science
Specialization: Machine Learning, Information Retrieval, Data Mining
18 Credits Completed
- **Lebanese International University** — B.Sc. in Computer Science
Graduated with a focus on Algorithms, Data Structures, & AI fundamentals
Sept 2017 – Aug 2020
- **University of Cambridge** — GCSEs of High School (A-levels)
Achieved an A-grade in Information & Communications Technology.
Sept 2011 - July 2016

Certifications & Presentations

- **Certifications:** Microsoft Certified: Azure AI Engineer Associate | Azure AI Fundamentals | LangChain for LLM Development (DeepLearning.AI) | Prompt Engineering Specialization | Certified ScrumMaster (CSM) | Supervised Machine Learning: Regression & Classification (Stanford Online)
- **Recognition:** Certificate of Appreciation — Princess Nourah bint Abdulrahman University (PNU) for delivering a **10-hour online course on Neural Networks** (2026)
- **Presentations:** GITEX Global 2025, 2024, 2023, 2022 (Dubai) | Biban 2024 (Saudi Arabia) | KKHU Medical City (KSA) | Multiple student mentorship programs in Saudi Arabia
- **Languages:** English (Fluent), Arabic (Native) – Technical presentations, client negotiations, documentation in both languages