



Mohamed Hatem

AI Team Lead · **Globant** · Multi-agent LLM systems, RAG, LLM serving

🌐 mohamedhatem.dev ✉️ hatem@mohamedhatem.dev 📞 +20 114 207 0946 🔗 in/mohamed-hatem-6a5790173

🐙 github.com/mhatem5351 📍 Cairo, Egypt · Remote

AI Team Lead with **7 years in data and AI**, designing, shipping and running production LLM systems: multi-agent orchestration (LangGraph), RAG and hybrid search, self-hosted model serving (vLLM on NVIDIA A100) and LLM evaluation. Built a **14-agent assistant** for ArabsStock and a **6-agent bilingual Arabic/English chatbot** at Wakeb Data. AWS Certified Machine Learning – Specialty (2022).

EXPERIENCE

AI Team Lead · Globant Apr 2026 – Present

Full-time · Remote, from Cairo · Case study ↗

- Leading AI engineering on the **Saudi Tourism Authority (STA) project**: an agentic LLM runtime with retrieval and evaluation for Noura, the visitor assistant, and a staff assistant, in Arabic and English.
- One LangGraph agent behind 96 tools** (27 visible to visitors), with human approval before every action; Noura was accepted in the client's UAT in September 2026.

Senior AI & MLOps Engineer · Wakeb Data Sep 2024 – Apr 2026

Full-time · Giza, Egypt · Case study ↗

Client: **Saudi Ministry of Defense (GADD)**

- Architected and deployed a fully async **6-agent bilingual (Arabic/English) LangGraph chatbot** on Qwen, served with vLLM on NVIDIA A100 GPUs.
- Owned **model serving across multiple production GPU servers**: vLLM configuration, GPU allocation, model versions, health monitoring and automated restarts.
- Built **hybrid RAG** (vector + BM25 with LLM reranking), node-level caching and Brave Search over MCP, shipped as Dockerised FastAPI services behind Nginx.
- Designed **AIP semantic search v2**, built a bilingual enterprise assistant for a Saudi government client with LLM evaluation, and moved key APIs from REST to gRPC.

AI Engineer · ArabsStock Apr 2023 – Dec 2024

Full-time · Remote (Dubai, UAE) · Case study ↗

- Architected and built the **ArabsStock Assistant: 14 LangGraph agents** for semantic search, AI image generation, database operations and content moderation.
- Built semantic search (OpenAI embeddings + Pinecone, automated batch ingestion) over **millions of photos, videos and vector graphics**, plus visual-similarity search.
- Shipped **audio AI** for a beta sound library (generation and search, instrument classification, BPM detection, watermarking) as containerised FastAPI services.
- Set up **agent evaluation** to track reliability, response accuracy and routing in production.

EDUCATION

M.Sc. in Data Science

Cairo University, Faculty of Graduate Studies for Statistical Research

2023 – 2027

(expected)

Diploma in Artificial Intelligence

Information Technology Institute (ITI)

2021 – 2022

B.Eng. in Structural Engineering

Tanta University, Faculty of Engineering

2015 – 2020

SKILLS

LLMS & AGENTS

LangGraph, LangChain, Multi-agent systems, MCP, OpenAI API, Qwen, Prompt engineering

RETRIEVAL

RAG, Hybrid search (vector + BM25), LLM reranking, Pinecone, ChromaDB, Qdrant, pgvector

SERVING & MLOPS

vLLM on NVIDIA A100, Model versioning, Health monitoring, Docker, Kubernetes (EKS), CI/CD, Nginx, Prefect

EVALUATION

LLM and agent evaluation, Observability, Guardrails, Langfuse

BACKEND & DATA

Python, FastAPI, gRPC, PostgreSQL, SQL Server, Cosmos DB, DuckDB, ETL

MACHINE LEARNING

PyTorch, TensorFlow, scikit-learn, NLP (Arabic and English), Computer vision, Spark



Software Engineer · Tabaqat

Part-time · Remote (Saudi Arabia)

Feb 2023 – May 2024

- Designed and managed **all data pipelines in Prefect**, including ETL for GIS data from Saudi government sources.
- Queried Overture Maps with AWS Athena and DuckDB; deployed map serving with MapProxy, MBTileServer, MapLibre and Kepler.gl.

CLOUD

AWS (SageMaker, EC2, S3, Athena, EKS), Azure (Functions, Cosmos DB), Cloudflare Workers

FRONTEND

TypeScript, Astro, Tailwind CSS, WCAG 2.2 AA, Arabic RTL



Data Scientist · Empoweromics

🏆 **Best Engineer of the Year**

Full-time · Cairo, Egypt · Case study ↗

Jan 2022 – Mar 2023

- Unified data from **3,008+ real-estate developers**, each with its own schema, to scale the company's real-time e-map.
- Built lead-lifecycle models (7 stages) for broker prioritisation, and **text + image WhatsApp classification** that routes messages to the right department.
- Built Smart Search (fuzzy lookup over misspelled or partial Arabic/English names) and a PyTorch model inferring gender from Arabic names.
- Designed ETL from Cosmos DB to SQL Server, scheduled with Azure Functions.

CERTIFICATIONS

AWS Certified Machine Learning – Specialty

Amazon Web Services · 2022

Data Scientist with Python

DataCamp · 2021

Data Analyst with SQL Server

DataCamp · 2022

Introduction to Tableau

365 Data Science

AWS Academy Cloud Foundations

AWS Academy

Earlier roles



AI Research Assistant · The British University in Dubai ·

Mar – Nov 2022

Umm Al-Qura University

Research assistant and AI lecturer for master's students in biomedical informatics at BUID; mentored master's research at BUID and Umm Al-Qura University.



AI Instructor · Military Technical College

Aug 2022

Taught AI courses (ML fundamentals, deep learning, deployment) to engineers in the Egyptian Armed Forces.



Freelance Data Scientist · AestheticsPro

Feb – Apr 2022

(US healthcare software)

Scraped, cleaned and loaded healthcare data into SQL Server; built EDA and a stakeholder dashboard. Built ML models predicting patient retention, used to target retention offers.



SQL Developer · Misria Digital Payments (MDP)

Jan 2020 – Jan 2021

Developed and optimised complex SQL stored procedures for production payment systems; query tuning and database administration.

AWARDS

Best Engineer of the Year

Empoweromics

LANGUAGES

Arabic

Native

English

Professional working proficiency

OPEN TO

AI/ML Team Lead

Senior AI / LLM Engineer

MLOps Engineer

Remote

Relocation to the Gulf or Europe

Hybrid in Cairo