

AMR MOHAMMED

MACHINE LEARNING ENGINEER • LLM SYSTEMS, RAG & EVALUATION

Cairo, Egypt | +201034184924 | amrm88289@gmail.com
amr-mohammed.com | linkedin.com/in/amr-mohammed01 | github.com/AmrMohamed17

SUMMARY

Machine Learning Engineer building production LLM systems with measurable quality — RAG pipelines served via FastAPI and deployed with Docker, CI/CD, and AWS. Built an evaluation-first RAG platform with a validated golden dataset, a three-metric eval harness, and a CI gate that blocks any change regressing retrieval quality. Experienced with conversational AI agents integrated with SQL data.

TECHNICAL SKILLS

LLM & RAG RAG Pipelines • Vector Search & Embeddings • Cross-Encoder Reranking • Rank Fusion • LLM-as-Judge Evaluation • Hallucination Guards • Prompt Engineering • LangChain
Backend & Data Python • TypeScript • FastAPI • Pydantic • REST APIs • SQL • PostgreSQL/pgvector • MySQL
MLOps & Cloud Docker & Compose • GitHub Actions (CI/CD) • AWS (EC2, RDS) • MLflow • Pytest • Linux • Git
Core ML PyTorch • TensorFlow • Scikit-learn • Pandas • NumPy • Computer Vision • NLP

PROFESSIONAL EXPERIENCE

Research Assistant • Freelance (Remote)

Dec 2025 – Apr 2026

Research project affiliated with King Saud University

- ▶ Conducted systematic literature reviews across scientific databases with documented search strategies and inclusion/exclusion criteria.
- ▶ Synthesized findings into a comparative evidence matrix consolidating results, gaps, and recommendations.

Machine Learning Engineer Intern • Elevvo Pathways

Jul 2025 – Aug 2025

- ▶ Built and tuned 6+ machine learning models (regression, classification, CNNs) using SMOTE and hyperparameter optimization, improving accuracy by 10–25%.
- ▶ Designed reproducible ML pipelines for preprocessing and feature engineering, defining evaluation metrics up front and integrating MLflow for experiment tracking and versioning.
- ▶ Improved model performance by up to 30% using SVD and transfer learning.

PROJECTS

DocuMind — Evaluation-First RAG Platform

Nov 2025 – Present

FastAPI • PostgreSQL/pgvector • Gemini • FlashRank • Docker • AWS

documind.amr-mohammed.com • [GitHub](https://github.com)

- ▶ Built and deployed an evaluation-first RAG platform over technical documentation; improved multi-hop retrieval recall@10 by 58% (0.36 → 0.57) with a two-stage retriever (pgvector dense search → FlashRank cross-encoder → Reciprocal Rank Fusion), holding single-hop recall at 0.97.
- ▶ Defined evaluation criteria before building: a validated 100-question golden dataset and a three-metric harness — recall@k, answer faithfulness (4.82/5 mean via LLM-as-judge), and a hallucination guard (15/15 unanswerable questions correctly refused).
- ▶ Engineered a CI quality gate (GitHub Actions) that fails any pull request whose retrieval recall regresses below a committed baseline — retrieval quality enforced, not merely measured.
- ▶ Measured retrieval latency (p50 809 ms / p95 1441 ms), isolating the embedding API round-trip as the dominant cost; batched query embeddings and pre-computed corpus fixtures — zero corpus re-embedding in CI, deterministic recall run-to-run.
- ▶ Deployed on AWS EC2 via Docker Compose (FastAPI + pgvector + Caddy auto-HTTPS); runs end-to-end from a clean clone with one command.

BeanBuddy — Cloud-Native Conversational AI Agent

Aug 2025 – Oct 2025

Dialogflow ES • FastAPI • MySQL • AWS RDS • EC2

github.com/AmrMohamed17/BeanBuddy

- ▶ Architected a stateful conversational ordering agent that handles multi-turn dialogs through a chat interface, persisting transactions to MySQL on AWS RDS, deployed on AWS EC2 (Linux).

Local Product Finder — Graduation Project (A+)

Jul 2024 – Jun 2025

MobileNet • Sentence-BERT • LightFM • TensorFlow

github.com/AmrMohamed17/Local-Product-Finder

- ▶ Built real-time product recognition with MobileNet (99% accuracy on the project test set) and a hybrid Sentence-BERT + LightFM recommender, integrated into a mobile app with a cross-functional team.

EDUCATION

B.Sc. Computer Science • Ain Shams University

Oct 2021 – Jul 2025

Graduation Project graded A+ (Local Product Finder)

CERTIFICATIONS

DeepLearning.AI & Stanford Online Advanced Learning Algorithms; Supervised Machine Learning (2024)

Imperial College London Mathematics for Machine Learning Specialization (2024)

Harvard University CS50's Introduction to AI with Python (2024); CS50x (2022)