

Mostafa Ali

AI Engineer | NLP · Generative AI

+201276913999 • mstfyshrqawy520@gmail.com • mstfyshrqawy520-alt.github.io/my-Portfolio/

Linkedin: linkedin.com/in/mostafa-el-shargawi-66b3a7352 • GitHub: github.com/mstfyshrqawy520-alt • Cairo, Egypt

PROFESSIONAL SUMMARY

Results-driven Artificial Intelligence Engineer with over 2 years of hands-on experience architecting and deploying production AI systems across NLP, Generative AI, Agentic Systems, and Computer Vision. Launched Hire-Mind — a production RAG platform achieving 35% API cost reduction and 92% document extraction accuracy, awarded an A+ grade and selected as a university showcase project. Proficient in the full AI lifecycle: data engineering, LLM orchestration, fine-tuning, agentic design, and scalable MLOps deployment on AWS and GCP.

TECHNICAL SKILLS

Languages: Python (NumPy, Pandas, Scikit-learn), SQL, Bash.

ML / DL Frameworks: PyTorch, TensorFlow, Keras, Hugging Face Transformers, Scikit-learn.

LLM APIs & Frameworks: OpenAI API, Anthropic Claude, Google Gemini, LangChain, CrewAI, Hugging Face Hub.

RAG & Retrieval: Hybrid Search, Cross-Encoder Re-ranking, ChromaDB, Qdrant, Sentence Transformers.

Prompt Engineering: Chain-of-Thought, Structured Output, Function Calling, Few-shot, Prompt Chaining, Pydantic.

Computer Vision: YOLOv8, EfficientNet-B4, OpenCV, PyQt5, Object Detection, Image Classification, Real-time Inference.

Cloud & MLOps: AWS (EC2, S3, Lambda), GCP (Vertex AI), Azure (AZ-900), Docker Compose, MLflow, CI/CD, Linux.

Databases: PostgreSQL, MongoDB, Redis, ChromaDB, Qdrant.

PROFESSIONAL EXPERIENCE

Valeo Egypt

Smart Village, Giza, Egypt (On-site)

AI Intern

Jun 2026 – Present

- Selected for Valeo Egypt's highly competitive AI Summer Internship Program, collaborating with senior engineers on production-grade AI systems.
- Built scalable applications across 3 AI domains — LLM, Retrieval-Augmented Generation (RAG), and Computer Vision — for real-world automotive AI use cases.
- Developed and deployed AI backend services across 4 core technologies — Python, FastAPI, Docker, and cloud infrastructure — with a focus on scalability and maintainability.
- Delivered end-to-end AI engineering workflows spanning 5 lifecycle stages — data preprocessing, model integration, evaluation, optimization, and API deployment.
- Collaborated in Agile cross-functional teams, contributing across 3 core areas — software architecture, code reviews, and engineering best practices — to deliver production-ready AI solutions.

FlyRank AI

Remote, USA

General AI Fluency Intern, 1 Month

June 2026

- Executed a 4-week remote AI Fluency program focused on prompt engineering, critically evaluating over 50 AI outputs, and responsibly using Generative AI tools.
- Presented and shipped 1 comprehensive capstone project leveraging 3 distinct AI tools, earning a verifiable credential and a recommendation letter upon graduation.

IEEE Beni Suef Student Branch

Hybrid

AI Committee Member

Feb 2026 – Present

- Spearheaded 10+ AI-focused technical initiatives, workshops, and knowledge-sharing sessions covering machine learning, deep learning, and real-world AI applications.
- Partnered cross-functionally to organize 5 major events discussing emerging AI trends, strengthening community impact for over 500 active members.

Confidential Financial Services Client

Remote

Freelance AI Engineer, 5 Months

Aug 2025 – Dec 2025

- Architected a secure, enterprise-grade Generative AI platform spanning 3 integrated modules — document ingestion, RAG-based compliance search, and an agentic query assistant — to automate financial report analysis across multiple report types.

- Designed a specialized Retrieval-Augmented Generation (RAG) pipeline for compliance documents, achieving high retrieval accuracy and reducing manual review time by 40%.
- Deployed an autonomous Agentic AI assistant using LangChain, delivering 2 core capabilities — complex financial query resolution and structured risk-assessment summarization.
- Containerized the NLP microservices across all 3 platform modules using Docker, deploying on secure cloud infrastructure that strictly adheres to financial data privacy and low-latency standards.

Digital Egypt Pioneers Initiative (DEPI / MCIT)

Hybrid

Generative AI Engineer, 8 Months

Nov 2025 – June 2026

- Graduated from an 8-month government-sponsored Generative AI track, covering Large Language Models, RAG architectures, PyTorch, and production deployment across 200+ training hours.
- Delivered 5+ hands-on enterprise projects in LLM fine-tuning and agentic system design, achieving a 95% evaluation score and a nationally recognized MCIT certification.

KEY PROJECTS

Hire-Mind — AI Recruitment & Document Intelligence Platform, [GitHub](#)

Grade: A+ (University Showcase)

Graduation Project

2024–2025

- Architected a multi-provider LLM orchestration layer with intelligent prompt routing, reducing API operational costs by 35%.
- Created an end-to-end RAG pipeline using Hybrid Search and Cross-Encoder Re-ranking, achieving an 88% relevance score and sub-200 ms API latency.
- Engineered a Document Intelligence pipeline achieving 92% extraction accuracy for Arabic and English documents using NLP, semantic parsing, and Named Entity Recognition.
- Integrated an Agentic AI Interview Coach with role-specific question generation and real-time structured candidate feedback.
- Containerized all microservices using Docker Compose, enabling independent horizontal scaling per component across the modular architecture.

Stack: Python, LangChain, OpenAI API, FastAPI, MongoDB, ChromaDB, Docker, Sentence Transformers.

FruitGuard — Real-Time Industrial Computer Vision System, [GitHub](#)

120 hrs

NTI Scholarship

2024

- Built a multi-stage CV pipeline utilizing YOLOv8 detection and EfficientNet-B4 classification, achieving sub-100 ms end-to-end inference.
- Reduced false-positive rate by 30% via multi-class confidence thresholding with fallback logic.
- Implemented full MLOps setup including MLflow versioning, GPU inference servers, and custom PyTorch data augmentation strategies.

Stack: PyTorch, YOLOv8, EfficientNet, MLflow, PyQt5, FastAPI, Docker, OpenCV.

EDUCATION

Egyptian e-Learning University (EELU)

Cairo, Egypt

Bachelor of Science in Computer Science — Artificial Intelligence

June 2026

Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, MLOps, Data Structures, Algorithms.

CERTIFICATIONS

- Microsoft Certified: Azure Fundamentals (AZ-900).
- NVIDIA Deep Learning Institute: Deep Learning (90 hrs) and Building LLM Applications.
- Digital Egypt Pioneers Initiative (DEPI / MCIT): Generative AI Track — 8 Months.
- NTI / ITIDA: Machine Learning Specialization — 90 hrs.
- ITI: Artificial Intelligence & Machine Learning — 72 hrs.
- DataCamp: Machine Learning Scientist, Deep Learning, NLP, Computer Vision.
- Sprints & Microsoft: AI & ML Summer Camp — Credential: SPR-TN945C.
- Google Cloud Skills Boost: Generative AI Fundamentals.
- FlyRank AI: General AI Fluency — Verifiable Credential.
- Prodigy InfoTech: Generative AI Engineer Training Certificate.

LANGUAGES

Languages: Arabic — Fluent | English — Professional Working Proficiency.