Mohamed Abdalkader

AI Engineer

(+20) 102 327 7913 | Mohameed.Abdalkadeer@gmail.com | Cairo, Egypt
LinkedIn | GitHub | Portfolio

Professional Summary

AI Engineer focused on LLMs, RAG systems, and applied AI solutions. Built and deployed 8 end-to-end ML projects spanning medical imaging and NLP. Designed RAG pipelines indexing 20+ medical textbooks with sub-second retrieval. Reduced model memory footprint from 17GB to 7GB through 4-bit quantization. Secured $3,450 in competitive research grants from ITAC and ASRT.

Experience

Freelance AI Engineer **May 2025 – Feb 2026**
Deep Learning, Computer Vision, NLP, Medical AI, MLOps *Remote*

- Fine-tuned Qwen 2.5 7B VLM with LoRA rank-32 on 35,000 retinal images for diabetic retinopathy diagnosis.
- Improved retinopathy detection accuracy from 20% to 76% and edema from 72% to 94%, reducing diagnosis time by 75%.
- Built ensemble pipeline (2 CNNs + RBF-SVM) improving ROUGE score from 0.17 to 0.44.
- Deployed RAG system indexing 20+ medical textbooks via FAISS (0.3s retrieval latency) with Flask API in Docker.
- Reduced inference memory from 17GB to 7GB via 4-bit quantization with no measurable accuracy loss.

AI Research Intern **Aug 2024 – Sep 2024**
Neuronetix

- Built XGBoost + Random Forest ensemble on 5,000+ patient records; achieved AUC 0.92 and F1 0.89.
- Ranked Top 3 of 150 in customer churn challenge with 93.5% accuracy using gradient-boosted trees.
- Authored model cards and deployment documentation for production handoff.

Machine Learning Intern **Mar 2024 – Jun 2024**
ShAI

- Developed music genre classifier (MFCC + XGBoost) achieving 95% accuracy across 8 classes.
- Built price prediction models with MAE < $50 via feature engineering and Optuna hyperparameter tuning.

Research & Competitive Funding

ITAC Research Grant | *70,000 EGP ($2,300 USD)* **2023**

- Built cloud-monitored solar generation system with LSTM forecasting; achieved 92% accuracy with a 48-hour predictive window.

ASRT Research Grant | *35,000 EGP ($1,150 USD)* **2023**

- Integrated IoT sensors with Transformer-based forecasting for energy consumption estimation; reduced prediction error by 23%.

IEEE Graduation Project Award | *National Recognition* **2023**

- Recognized nationally for a renewable energy forecasting framework spanning 21 deep learning architectures.

Projects

Medical VLM with Ensemble Architecture | *Qwen 2.5, PyTorch, LoRA* **NDA**

- Fine-tuned 7B VLM with LoRA rank-32 on 35,000 retinal scans for automated ophthalmic diagnosis.
- Improved retinopathy detection from 20% to 76% and edema classification from 72% to 94%.
- Designed ensemble pipeline (2 CNNs + RBF-SVM) for decision fusion, improving report generation quality.
- Applied 4-bit quantization, reducing VRAM from 17GB to 7GB with negligible accuracy drop (<0.5%).

NeuralChat | *FastAPI, LangChain, Cohere, OpenAI, Groq, Railway* **GitHub | Demo**

- Multi-provider LLM chatbot supporting 10+ models (Cohere, OpenAI, Groq) switchable from the UI.
- Supports 5 prompting strategies: Zero-Shot, Few-Shot, Chain-of-Thought, Memory Chain, and Structured JSON.
- Streams responses via SSE with per-session memory, 6 AI personas, and real-time cost and latency display.
- Deployed on Railway as a unified service; FastAPI serves both the REST API and frontend.

Mo-Bot — Personal AI Telegram Bot | *Cloudflare Workers, Groq, LLaMA 3.3, JavaScript* **Live Bot**

- Serverless AI assistant deployed on Cloudflare Workers, representing Mohamed professionally 24/7 at zero infrastructure cost.
- Implements a three-model fallback chain (LLaMA 3.3 70B → 3.1 70B → 3.1 8B) across independent Groq token pools.
- Uses keyword-based CV extraction to send only relevant sections per query, minimizing token consumption.
- Supports bilingual Arabic/English input with Arabizi detection and pronoun resolution via conversation history.

Face Recognition & Similarity System | *PyTorch, Streamlit, SQL* **GitHub** | **Demo**

- Built facial recognition system with desktop GUI and SQL database managing 1,000+ identity profiles.
- Deployed Streamlit app with cosine similarity matching; achieved 0.92 accuracy at 15 FPS on CPU.

Technical Skills

LLMs & NLP: Qwen 2.5, LLaMA, BERT, GPT, LoRA/QLoRA, RAG (LangChain, LangGraph, FAISS, Chroma), Few-Shot & Chain-of-Thought Prompting, BLEU/ROUGE/BERTScore, Hallucination Detection

Computer Vision: Vision-Language Models, CNNs (DenseNet, EfficientNet, ResNet), YOLO, OpenCV, MediaPipe, 4-bit/8-bit Quantization, Knowledge Distillation

MLOps & Deployment: Docker, MLflow, FastAPI, Flask, Azure, AWS, CI/CD Pipelines, Railway, Cloudflare Workers

Tools & Data: PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, Pinecone, Streamlit, SQL

Programming: Python (Expert), JavaScript, Java, SQL, Bash

Education

B.Sc. Computer Science **2019 – 2023**
Zagazig University *Cairo, Egypt*

- Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, Algorithms, Databases.
- **Graduation Project:** Python framework with 21 DL models (LSTM, GRU, Transformers) for renewable energy forecasting; achieved <15% MAPE on solar/wind prediction.

Digital Egypt Pioneers Initiative (DEPI) **Apr 2024 – Oct 2024**
Machine Learning Engineer Scholarship — Microsoft Track

- 6-month program (300+ hrs) covering production ML, MLOps, and Azure deployment.
- **Capstone:** Breast cancer classifier at 97% accuracy on 15,000+ images using DenseNet transfer learning.

Professional Certifications:

- Deep Learning & Machine Learning Specializations — DeepLearning.AI / Stanford University, Andrew Ng (2022–2024)

Soft Skills

ML System Architecture · Model Debugging & Performance Optimization · Prompt Engineering · Technical Documentation · Cross-functional Collaboration · Algorithm Problem-Solving (100+ LeetCode) · Fast Technology Adoption