

Saleh Ashref

salehyassien0@gmail.com | +201144840473 | Gharbia, Egypt
in [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | [HuggingFace](#)

SUMMARY

Machine Learning Engineer with hands-on experience building and deploying NLP and computer vision systems. Specialized in Arabic NLP, LLM fine-tuning (LoRA/QLoRA), and production ML APIs using PyTorch, Hugging Face, and FastAPI — with projects deployed live on Hugging Face Spaces and Docker.

PROFESSIONAL EXPERIENCE

SYNC INTERN'S

Machine Learning Intern

Mirzapur, Uttar Pradesh
Dec 2023 – Jan 2024

- Built and deployed 3 end-to-end ML models in 4 weeks: a CNN-based Sign Language Classifier, a real-time Face Mask Detector using OpenCV, and a housing price regression model — covering the full pipeline from data preprocessing to inference.
- Applied transfer learning, data augmentation, and hyperparameter tuning across TensorFlow/Keras projects; evaluated model performance using cross-validation and standard classification metrics.

PROJECTS

Arabic News Agentic RAG | [GitHub](#)

Jul 2026

- Built a LangGraph agentic RAG system with an explicit four-tool state machine (search, summarize, timeline, direct answer) routing Arabic queries over 26,000+ indexed news chunks — chose LangGraph over a ReAct agent specifically for inspectable, visualizable node/edge control flow rather than a black-box loop.
- Implemented hybrid retrieval combining AraBERT v2 dense vectors (768-dim) with BM25 sparse vectors via FastEmbed, fused using Reciprocal Rank Fusion (RRF) in Qdrant — outperforming dense-only retrieval on Arabic queries where keyword matching is critical for morphologically rich text.
- Diagnosed and fixed 6 production bugs through systematic root-cause analysis: Qdrant filter leaking across Prefetch sub-queries; routing prompt/TOOL_MAP spelling mismatch causing silent fallback on every summarize query; AraBERT tokenizer deadlock on Windows resolved by bypassing the sentence-transformers wrapper; and duplicate indexing eliminated with deterministic MD5 hash IDs.
- FastAPI backend with error handling and 30s timeout wrapper; Streamlit frontend with Arabic RTL support; full architecture documented with decision rationale, rejected alternatives, and known limitations.

Syntaxa — Prompt Generation Model (LoRA Fine-Tune) | [HF Model](#)

April 2026

- Fine-tuned Microsoft Phi-3.5-mini-instruct (3.8B) using LoRA (rank 16, alpha 32) with 4-bit quantization on a curated instruction dataset to generate structured system prompts from natural language persona descriptions.
- Trained for 3 epochs with cosine LR scheduling, FP16 precision, and gradient accumulation; merged adapters into the base model and published full weights to Hugging Face.
- Gained 386+ organic downloads within weeks of release with zero promotion — model deployed as the ML backbone of a live prompt engineering platform.

Arabic Sentiment Analyzer | [Live Demo](#)

March 2026

- Built a production FastAPI service for Arabic sentiment analysis using CAMEL-Lab's BERT model with ~87% accuracy and sub-100ms inference latency on live requests.
- Implemented single and batch inference endpoints with confidence scoring, input validation via Pydantic, and containerized the full stack with Docker and Docker Compose.
- Deployed on Hugging Face Spaces with a Streamlit UI; model weights auto-download and cache on cold start, reducing subsequent load time to under 2 seconds.

SKILLS

- Languages & Frameworks:** Python, SQL — PyTorch, TensorFlow/Keras, Hugging Face Transformers, Scikit-learn, FastAPI, Streamlit, NumPy, Pandas, Matplotlib
- ML / NLP:** Supervised & Unsupervised Learning, CNNs, RNNs, Transformers, LoRA/QLoRA Fine-Tuning, Hybrid Retrieval (Dense + Sparse + RRF), RAG Pipelines, Arabic NLP (AraBERT, CAMEL-BERT), Model Evaluation
- Agentic & Orchestration:** LangGraph, LangChain, LangChain-Groq, Groq API (Llama 3.3 70B)
- MLOps & Infra:** Docker, Docker Compose, Qdrant, Git, Hugging Face Spaces, Google Colab, FastEmbed, Postman
- Familiar with:** Vector DB scaling, experiment tracking (MLflow), cross-encoder reranking, NewsAPI integration

EDUCATION

Higher Institute of Management and Information Technology

Kafrelsheikh, EG

Bachelor's in Computer Science

July 2024

- Final year grade: Excellent, Cumulative grade: very good.
- Graduation project grade: Excellent.