Mohamed Hammad

Senior AI/ML Engineer

mohamedhammad1488@gmail.com +971581143623 Dubai, UAE github.com/mo-9

PROFILE

AI/ML Engineer with 5+ years of experience designing, building, and deploying end-to-end machine learning solutions in production environments. Expertise spans machine learning, deep learning, NLP, and computer vision, with a strong emphasis on generative AI, LLM-powered chatbots, Retrieval-Augmented Generation (RAG), and intelligent agent frameworks. Demonstrated success in delivering high-impact models and optimizing real-world applications at scale. Highly proficient in Python, PyTorch, TensorFlow, LangChain, and MLOps, with a proven ability to build scalable, reliable, and production-grade systems that drive measurable business value and innovation.

SKILLS

Programming Languages
- **Python**

Vectordatabase
- **ChromaDB, FAISS, Weaviate, Qdrant, pinecone**

Computer Vision
- **CNN, OpenCV , YOLO, Faster R-CNN , SSD, DETR ,EasyOCR**

MLOps
- **Git, GitHub, GitHubAction, MLflow, DVC, Docker**

Natural Language Processing (NLP)
- **NLTK, SpaCy , Gensim , Hugging Face Transformers , BERT , GPT**

AWS
- **EC2, S3, Serverless compute: AWS Lambda, API Gateway**

Database
- **SQL, MySQL, PostgreSQL, MongoDB**

BackEnd
- **FastAPI,Django,Flask**

Data Processing
- **NumPy ,Pandas, Matplotlib, Power Bi , PySpark, AirFlow**

Machine Learning Frameworks
- **Scikit-Learn, TensorFlow, Keras, PyTorch**

Automation
- **n8n**

Generative AI & LLM Frameworks
- **LangChain, LangGraph, OpenAI chatgpt , Ollama, Qwen3, DeepSeek, CrewAI, Model Context Protocol (MCP),**

Voice Tools and Streaming
- **WebSocket, WebRTC, LiveKit, STT Whisper, TTS kokoro , Elevenlabs**

EDUCATION

Bachelor of Computer Engineering ELshorouk university
09/2018 – 06/2023

Graduated with a strong foundation in Computer Science, Artificial Intelligence, algorithm design, and full-cycle software development, including advanced algorithmic techniques.

LANGUAGES

Arabic
Native

English
Fluent

PROFESSIONAL EXPERIENCE

Senior AI/ML Engineer AT Grand Technology
06/2025 – Present | Cairo, Egypt

- Designed and deployed a **real-time conversational voice agent** enabling low-latency, human-like voice interactions using LiveKit and WebRTC.
- Optimized system performance for real-time streaming, significantly improving user experience and conversational responsiveness.

AI/ML Engineer AT Viganium
01/2025 – 06/2025 | Cairo, Egypt

- Designed and implemented an **AI-powered chatbot** for a car rental platform, enabling automated customer support and reservation handling.
- Integrated internal business data to deliver **personalized, context-aware responses** and real-time booking assistance.
- Built full **Arabic language support**, improving accessibility and customer engagement for regional users.

Ai Engineer AT LamasaTech U.K
03/2024 – 01/2025

Benton, North Tyneside, UK

- **Advanced GenAI Agents:** Developed advanced generative AI agents using **LangChain**, **LangGraph**, and **n8n** to automate workflows and enhance customer interactions, driving efficiency and scalability.
- **AI-powered Marketing:** Led AI-powered marketing initiatives by generating engaging video and audio content, boosting brand reach and impact across digital platforms.

AI/ML Enginner at Aurakore
06/2024 – 12/2024 | Texas U.S

part-Time

- **Advanced AI Agents**: Designed and deployed advanced AI agents leveraging Retrieval-Augmented Generation (RAG) pipelines with LangChain, LangGraph, and OpenAI APIs. Implemented FAISS-based vector similarity search, ensuring scalable, low-latency retrieval across large and complex datasets.

Machine Learning Engineer At Fast Kood Company
10/2021 – 03/2024 | Cairo, Egypt

- **AI/ML Solutions Across Key Sectors**: Led the end-to-end development of advanced AI/ML solutions across industries like healthcare, e-commerce, and speech processing, delivering impactful applications that drive business value.
- **Cross-functional Collaboration**: Partnered with cross-functional teams to deploy machine learning models into production, leveraging Python, PyTorch, TensorFlow, and MLOps best practices to ensure seamless integration and scalability.

PROJECTS

Automated Business Card OCR & Workflow Integration

- Built an **end-to-end automated pipeline** that ingests business cards via WhatsApp, applies OCR to extract structured company and contact data, and stores it centrally.
- Integrated **AI agents with Tavily search** to enrich company information, validate contacts, and improve data accuracy.
- Automated workflows using **n8n**, connecting WhatsApp, Gmail, and Google Sheets, with **auto-generated outreach emails** that significantly reduced manual data entry and follow-up effort.

AI Chatbot for Car Rental Services

Built and deployed an **Arabic (Saudi/Gulf dialect) AI-powered chatbot** for a car rental company, enabling customers to inquire about services, compare vehicle options, and complete **automated reservations** through natural, conversational interactions.

Tools & Technologies: LangChain, FastAPI, Retrieval-Augmented Generation (RAG), OpenAI APIs, Vector Databases (FAISS), Python, NLP, Arabic Language Processing

Gold Price Forecasting Agent

- Designed and deployed an **AI-powered financial forecasting agent** for gold price prediction, leveraging **LangChain and LangGraph** to orchestrate data processing and real-time analytical insights.Integrated **Arima** and **LSTM-based time series models** to forecast future gold price movements using historical market data, enabling data-driven decision support.

AI-Powered Healthcare Assistant

- Built an intelligent healthcare assistant using LLMs and Retrieval-Augmented Generation (RAG) to analyze patient symptoms and medical history, retrieve relevant clinical guidelines, and deliver explainable, non-diagnostic risk insights.
- Implemented vector-based medical knowledge retrieval and deployed the system as a scalable, production-ready FastAPI service.

E-Commerce Conversational Chatbot

- **Product Recommendation:** Designed and implemented an intelligent conversational chatbot that delivers **personalized product recommendations** using **LLMs, RAG**, and customer behavioral data.
- **Selling & Reservation:** Enabled **conversational commerce workflows**, allowing users to browse and compare products, place orders, and complete reservations seamlessly through natural language interactions.
- **Customer Support & Order Tracking:** Integrated automated customer support and **real-time order tracking** by connecting with shipping provider APIs **Aramex** , delivering live shipment status, delivery notifications, and post-purchase assistance.

Smart Retail Analytics System

- Designed and implemented an AI-driven computer vision analytics platform leveraging object detection and multi-object tracking to analyze customer movement and product interactions from real-time video streams.
- Applied behavioral analytics and real-time tracking to measure dwell time, customer flow, and engagement hotspots, enabling actionable, data-driven retail optimization.

Real-Time Voice-Based Conversational Avatar Assistant

- Designed and implemented a real-time, voice-enabled conversational assistant with a visual avatar, enabling natural and low-latency human–AI interactions directly on screen.
- Integrated streaming speech-to-text, LLM-based dialogue orchestration, and real-time text-to-speech, delivering continuous voice-driven conversations with synchronized avatar speech and animations.

Customer Support Agent

- Designed and implemented an AI-powered customer support agent leveraging LangChain, LangGraph, and LangSmith to orchestrate context-aware, multi-step query resolution.
- Integrated the Groq API with Qwen-QWQ-32B to enable low-latency inference and fast, accurate responses, significantly improving customer support efficiency and response quality.

Books Text Summarization

Integrated document summarization pipelines using open-source LLMs to automatically condense long documents as part of enterprise knowledge and customer support systems.

Fine-Tuning LLAMA 3.2 on an Arabic Dataset

Fine-runed a custom LLAMA 3.2 model for Arabic NLP tasks including question answering, summarization, and sentiment analysis. Employed LoRA and QLORA for parameter-efficient training on consumer GPUs using PEFT. Built a memory-optimized pipeline with PyTorch, Hugging Face Transformers, and BitsAndBytes, leveraging mixed-precision (FP16/BF16) for efficient performance.

Restaurant Reservation Voice Agent

- Designed and implemented an AI-powered voice agent for restaurant reservations, enabling customers to book meals, inquire about menus, availability, and special requests through natural, real-time voice conversations.
- Integrated speech-to-text (STT), LLM-based dialogue orchestration, and text-to-speech (TTS) to deliver low-latency, human-like interactions, with automated call handling and confirmations via **Twilio.**