Mohammed Abdul Omer

Computer Science Engineering (AI and ML)

+91 96521 59548 | mohammedabdulomer99@gmail.com | Hyderabad, India

linkedin.com/in/mohammad-abdul-omer | github.com/MOHD-OMER | mohdomer.vercel.app | huggingface.co/mohdomer

PROFESSIONAL SUMMARY

Final-year Computer Science Engineering student (AI and ML), graduating May 2026. I build AI systems end to end — from fine-tuning large language models and designing RAG pipelines to shipping FastAPI backends and monitoring models in production. My work spans NLP, computer vision, multi-agent AI, and MLOps, with 5 live applications deployed on Hugging Face Spaces and Render. I am looking for a full-time role as an AI Engineer or ML Engineer where I can keep building things that work in the real world.

TECHNICAL SKILLS

Programming Languages: Python, SQL

AI and Machine Learning: supervised learning, unsupervised learning, model training, model evaluation, hyperparameter tuning, feature engineering, data preprocessing, Scikit-learn, PyTorch, TensorFlow

Deep Learning and Computer Vision: convolutional neural networks (CNN), recurrent neural networks (RNN), transfer learning, object detection, image classification, gradient class activation mapping, OpenCV, Hugging Face Transformers

Natural Language Processing (NLP): text classification, information retrieval, large language models (LLM), LangChain, LangGraph, CrewAI, retrieval-augmented generation (RAG), vector databases, prompt engineering, Mistral 7B, LLaMA 3, Groq, Ollama

LLM Fine-Tuning: quantized low-rank adaptation (QLoRA), parameter-efficient fine-tuning (PEFT), supervised fine-tuning, Weights and Biases

Backend and APIs: FastAPI, Django, REST API design, asynchronous programming, Streamlit, Gradio

MLOps and DevOps: MLflow, Weights and Biases, Evidently AI, Docker, model deployment, data drift monitoring

Tools and Platforms: Git, GitHub, Hugging Face Hub, Kaggle, Vercel, Jupyter Notebook, VS Code

Soft Skills: problem solving, analytical thinking, cross-functional collaboration, technical communication, project ownership, adaptability

PROFESSIONAL EXPERIENCE

AI Engineer Intern *March 2025 to June 2025*

TechZone Software Academy — Hyderabad, India

- Architected and delivered 4 production-grade AI and ML systems for a live educational platform, streamlining academic workflows and reducing manual overhead across 3 departments
- Designed a retrieval-augmented generation (RAG) tutoring assistant using LangChain, a vector database, and LLaMA 3 via Groq API, applying NLP and information retrieval techniques to significantly cut educator query response time
- Engineered an automated language proficiency test engine with LLM-based scoring across reading, writing, and listening modules, improving student test-readiness through personalised AI feedback
- Built an automated MCQ generation and attendance tracking system that eliminated manual administrative workflows; orchestrated all 4 systems as production-ready FastAPI REST services with continuous reliability

AI and ML Intern *November 2024*

RAM Innovative Infotech — Hyderabad, India

- Developed and evaluated a supervised machine learning disease-prediction classifier using Scikit-learn on a structured medical dataset, applying data preprocessing and feature engineering to optimise model accuracy
- Integrated the trained model into a live Django web application, gaining end-to-end experience with REST-based AI model deployment serving real users

PROJECTS

Building Safety Smoke Detection | *CNN, object detection, transfer learning, machine learning, Django* *B.E. Major Project*

- Engineered a dual-module fire and smoke detection system on an IoT sensor dataset of 62,630 readings across 13 sensor channels; trained 7 ML classifiers (Random Forest, SVM, Gradient Boosting) achieving AUC-ROC above 0.999
- Applied transfer learning to deploy a MobileNet CNN reaching 96.98% validation accuracy on image classification; integrated YOLO object detection for real-time bounding box prediction on fire and smoke regions

- Developed and delivered the complete system as a role-based Django web application with Railway deployment

Multi-Document RAG Chatbot | *NLP, LangChain, LangGraph, vector database, FastAPI, Streamlit*

- Implemented a hybrid retrieval agent combining dense semantic embeddings with sparse keyword search and a cross-encoder reranker to improve retrieval precision across multi-document corpora
- Built an asynchronous FastAPI backend with server-sent event streaming for real-time responses; integrated a LangGraph ReAct agent loop with persistent memory for multi-turn conversation; published on Hugging Face Spaces

LLM Fine-Tuning: Mistral 7B Medical QA | *QLoRA, PEFT, supervised fine-tuning, Weights and Biases huggingface.co/mohdomer/mistral-7b-medical-qa-qlora*

- Fine-tuned a 7-billion-parameter large language model using 4-bit quantized low-rank adaptation on a Kaggle T4 GPU, cutting GPU memory requirements by over 75% versus full fine-tuning
- Logged all training runs in Weights and Biases; achieved consistent ROUGE-1, ROUGE-2, and BLEU score gains over the untuned baseline on medical question answering benchmarks; published the merged model to Hugging Face Hub

Multi-Agent Research System | *multi-agent AI, CrewAI, Tavily search, LLaMA 3, Python huggingface.co/spaces/mohdomer/nexu research*

- Orchestrated a CrewAI multi-agent pipeline with live Tavily web search, coordinating specialised agents for research planning, information retrieval, and structured synthesis; automated the full research workflow from query to formatted report with citations

AI Orchestrator: Multi-Provider LLM Router | *FastAPI, Groq, Gemini, OpenRouter, Ollama ai-orchestrator-nu86.onrender.com*

- Designed a unified LLM routing system supporting 17 or more models across 4 providers with ML-based automatic task classification and a benchmark mode for side-by-side latency and output quality comparison

MLOps Pipeline | *MLflow, Weights and Biases, Evidently AI, Docker*

- Constructed a modular, reproducible ML pipeline with MLflow experiment tracking, a model registry for version control, and Evidently AI automated data drift monitoring; containerised the entire workflow with Docker for consistent deployment across environments

PulmoScan AI: Tuberculosis Detection | *CNN, transfer learning, gradient visualization, FastAPI pulmoscanai-app.hf.space*

- Trained a tuberculosis detection CNN with transfer learning and gradient-based saliency maps for clinical explainability; automated PDF diagnostic report generation per scan; published as a live application on Hugging Face Spaces

TruthLens: Fake News Detector | *NLP, text classification, CNN-RNN, Keras Mini Project — truthlens-uqo4.onrender.com*

- Built a misinformation classifier using a combined CNN and RNN architecture trained on over 40,000 labelled news articles, achieving 94.2% classification accuracy; released as a live inference web application on Render

EDUCATION

Bachelor of Engineering, Computer Science Engineering (AI and ML) *2022 – 2026*
Lords Institute of Engineering and Technology, Hyderabad, India

Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, Data Structures and Algorithms, Database Management Systems, Artificial Intelligence, Software Engineering

ACHIEVEMENTS

- Published 5+ deployed AI applications on Hugging Face Spaces and Render with functional inference pipelines accessible to real users
- Fine-tuned and released a Mistral 7B Medical QA model using QLoRA and PEFT with documented ROUGE and BLEU improvements over the untuned baseline
- Built AI systems spanning four domains: Retrieval-Augmented Generation, Computer Vision, Multi-Agent AI, and MLOps end-to-end pipelines
- Delivered production AI solutions supporting academic operations across multiple departments of a live educational platform during internship