

Omar Baig

AI / Machine Learning Engineer

Lahore, Pakistan | omarbaig8070@gmail.com | +92 334 0101065
omarbaig.tech | github.com/OmarBaig05 | LinkedIn

Summary

AI/ML engineer with 2+ years of experience across AI/ML projects, freelance work, internships, and research. Builds LLM applications, document-retrieval pipelines, computer vision models, and recommendation systems. Delivered 10+ client AI/ML projects, with hands-on work across Python APIs, databases, authentication, subscriptions, and deployment. MITACS Globalink Research Intern and first author of a paper accepted for oral presentation at IEEE FLLM2026.

Experience

Freelance AI & Machine Learning Engineer

Oct 2025 – Present

Fiverr (Level 1 Seller)

Remote

- Delivered 10+ client AI/ML projects with a 5-star rating, including a deployed chatbot with document retrieval, FastAPI, React/TypeScript, MongoDB, Stripe billing, Clerk authentication, and an admin dashboard.
- Built Groq-backed LLM services with vector search, sentiment analysis, rate limiting, and subscription-based query limits; deployed services to a production VPS.
- Trained and optimized models on client datasets and integrated model outputs into applications and automated decision-support workflows.

MITACS Globalink Research Intern

Jul 2026 – Present

University of Winnipeg

Winnipeg, Canada

- Built a full-stack research application integrating LLMs with health data for a smartwatch health platform.
- Owned application architecture and worked with a faculty supervisor and undergraduate researchers to translate research requirements into software components.
- Developed TypeScript frontend and backend services, REST API integrations, database-backed workflows, and reusable LLM and retrieval components.

Generative AI Intern

Jun 2024 – Aug 2024

Xavor Corporation

Lahore, Pakistan

- Built a mobility-assessment chatbot with a developer-clinician team, using OpenCV and MediaPipe to automate Functional Reach Tests.
- Developed a real-time pose-estimation and geometric measurement pipeline; project ranked first among 10+ internship teams.

Projects

Tallby — AI Mathematics Tutoring Platform | tallby.no

TypeScript, React, Hono, tRPC, MongoDB, OpenAI, Whisper, Paddle

- Built a full-stack AI tutoring application supporting guided explanations, image-based questions, voice input, and generated practice problems and study materials.
- Implemented token-based usage budgets and model-cost accounting across subscription tiers, with separate transcription usage tracking.
- Integrated JWT authentication with refresh-token rotation, Paddle subscriptions, English/Norwegian support, and a 24-panel admin analytics dashboard.

RAG-Powered Medical Report Interpreter

Python, FastAPI, Pinecone, MongoDB, Next.js

- Built an AI application combining OCR, document retrieval, report interpretation, comparative analysis, and visual summaries for clinician and patient use.
- Developed APIs with JWT authentication, asynchronous database access, vector search, and rate limiting; deployed the frontend on Vercel and backend on Hugging Face Spaces.

Model Development

AiroDx — Multi-View Chest X-ray Classification

PyTorch, timm, DINO-ViT, PadChest | Oct 2025 – May 2026

- Developed and evaluated a dual-backbone vision transformer pipeline combining frontal and lateral X-ray features through cross-view attention across 14 respiratory pathologies.
- Built data preparation workflows for PadChest, pairing views by StudyID and addressing class imbalance with weighted binary cross-entropy and OneCycleLR.
- Benchmarked against a reimplemented REMEDIS ResNet-50 baseline and added a consensus view of model agreement and disagreement.
- Ranked first among approximately 55 final-year projects at UET Lahore.

Hybrid Movie Recommendation System

PyTorch, Sentence-BERT, Neural Collaborative Filtering | Nov 2024 – Jan 2025

- Combined metadata embeddings with neural collaborative filtering on 2.5M+ interactions, adding a content-based fallback for cold-start recommendations.
- Evaluated recommendation quality using RMSE, precision, and mean average precision; achieved RMSE approximately 1.08, precision 0.80, and MAP 0.61.

Research

FedGraph-LoRA: Hierarchical Preference Graphs and Federated LoRA Adaptation for Personalized LLM Agents

First author | Accepted for oral presentation at IEEE FLLM2026

- Developed hierarchical preference retrieval, local/global LoRA adapters, federated aggregation, and query routing for personalized text-based LLM agents.
- Evaluated retrieval and response alignment against vector retrieval and federated baselines using a synthetic five-user dataset.

HAVIT: Attentive Probing of a Frozen Chest X-ray Foundation Model

Submitted to FIT 2026 | 2026

- Developed a parameter-efficient attentive-probing framework with a reported 6.24-point mean-AUC improvement over a matched frozen-feature baseline.
- Used statistical comparisons and ablation experiments to evaluate model quality; found that explicit cross-view fusion did not improve performance in the frozen, low-data setting.

Technical Skills

Languages: Python, TypeScript, JavaScript, SQL; additional experience with C# and C++.

Model development: PyTorch, Scikit-learn, TensorFlow, Hugging Face Transformers, timm, OpenCV; model training, evaluation, fine-tuning, LoRA/PEFT.

LLM applications: OpenAI, Groq, Whisper, RAG, LangChain, LangGraph, Pinecone, LLM agents.

Backend and data: FastAPI, Flask, Hono, tRPC, MongoDB, REST APIs, JWT authentication.

Frontend and delivery: React, Next.js, Tailwind CSS, Docker, VPS deployment, Vercel, Hugging Face Spaces; Stripe and Paddle billing.

Education

B.S. Computer Science

University of Engineering and Technology (UET), Lahore

Dec 2022 – Jun 2026

CGPA: 3.56/4.0

Thesis: AiroDx — Context-Aware AI for Multimodal Differential Diagnosis of Respiratory Diseases.

Awards and Certifications

- MITACS Globalink Research Internship Award, 2026.
- First Place, Final Year Project Exhibition, UET Lahore; First Place, Generative AI Internship Project, Xavor Corporation.
- Machine Learning Specialization, DeepLearning.AI / Stanford; Oracle OCI 2025 Data Science Professional; Oracle OCI 2024 Generative AI Professional.