

# Umar Abdul Rehman

Islamabad, Pakistan | [umarabdulrehman126@gmail.com](mailto:umarabdulrehman126@gmail.com) | +923105203485 | [LinkedIn](#) | [GitHub](#)

## Experience

---

**AI/ML Engineer**, Blue Surge pvt ltd – Islamabad

July 2024 – Present

- Trained computer vision models such as YOLOv9, Faster R-CNN, and Detectron2 for small object detection tasks in UAV and surveillance datasets.
- Fine-tuned YOLOv8 models for pedestrian detection and UAV tracking, optimized for real-time inference on Jetson Nano, Orin NX, and AGX Xavier devices.
- Developed a gait analysis system by integrating object tracking outputs with pose estimation frameworks to extract spatio-temporal motion features.
- Designed and deployed scalable microservices on Google App Engine using Docker containers, enabling seamless communication and autoscaling for AI inference.
- Built a Retrieval-Augmented Generation (RAG) based chatbot using Django backend, integrating FAISS vector search, Hugging Face Transformers, and custom document chunking pipelines.
- Implemented Dockerized microservices for the RAG system, separating document preprocessing, embedding generation, vector storage, and LLM querying as independent services.
- Orchestrated the RAG system components with Docker Compose, enabling reproducible development, versioned deployments, and modular backend updates.
- Leveraged CI/CD pipelines to deploy chatbot and vision services on cloud infrastructure, ensuring consistent model updates and fault-tolerant scaling.

**Summer Intern**, NRTC - Islamabad

Aug 2023 - Sep 2023

Engineered a vehicle number plate recognition system with improved detection accuracy and applied machine learning to optimize greenhouse resource forecasting during my summer internship at NRTC.

## Education

---

**Institute of Space Technology Islamabad**, BS in Computer Science

July 2020 - July 2024

### Co-Curricular/Lead Activities

- Served as President of ACM-IST Chapter Islamabad, Pakistan (Mar 2023 - Jun 2024). Founded the IST Chapter and organized 10+ technical events,
- Worked as Lead Event Manager at Google Developer Student Club - IST, Islamabad, Pakistan. Coordinated 8+ tech-related events, enhancing student engagement.

## Projects

---

### – Multi-Agent LLM Platform for Real-Time Enterprise Support

Built a scalable multi-agent system using LangChain, FastAPI, and Redis Pub/Sub for real-time support ticket handling. Integrated OpenAI GPT-4, Google Gemini, and HuggingFace models for hybrid task routing. Enabled RAG (Retrieval-Augmented Generation) with FAISS vector store and PDF/CSV ingestion. Added voice query handling using Whisper, and image input with CLIP. Deployed on Kubernetes with CI/CD via GitHub Actions and Prometheus monitoring.

### – Sugar Care | Hybrid Deep Learning for Early Diabetes Detection

Built hybrid deep learning models (CNN-BiLSTM, LSTM-FNN) using PyTorch with 88.04% accuracy on BRFS data.

Integrated SHAP for explainability and deployed predictions via a Flutter and MERN stack app. Included a neural network-based diet recommender trained on 500K+ recipes, a Streamlit-based chatbot using Google GenAI, and insulin reminders.

- **Customer Complaint Analysis & Resolution Prediction**

Designed an ETL pipeline and SQL-driven complaint categorization system. Built Power BI and Tableau dashboards for telecom support teams. Integrated predictive analytics to forecast escalations, proactively resolving 40% of cases.

- **Fashion Footprint | Outfit Recommendation App**

Led development of an outfit classifier using VGG16/InceptionV3 (95% accuracy), YOLOv8 for body/skin tone analysis, and CNNs on Fashion MNIST. Deployed using AWS, Docker, and Jenkins with A/B tested recommendation logic.

- **LLM-Based Context-Aware Q&A System with LangChain & FastAPI**

Developed a context-aware document Q&A system using LangChain, OpenAI embeddings, and vector stores. Served via FastAPI with role-based access and JWT auth. Integrated frontend using React and tested deployment on Railway and Azure App Service.

- **CI/CD Pipeline for AI Microservices using GitHub Actions & Docker**

Automated testing and deployment of multiple AI microservices including FastAPI apps and LangChain agents. Configured GitHub Actions for linting, testing, and Docker image builds; deployed to GCP Cloud Run with secret

## Skillset

---

**AI & Machine Learning:** Python, AI, Machine Learning, Deep Learning, Predictive Modeling, TensorFlow, Keras

**Technologies & Frameworks:** Microservices, FastAPI, Docker Compose, Kubernetes, CI/CD, GitHub Actions, Django, Jenkins

**Transformers & RAG:** Hugging Face Transformers, Lang Chain, FAISS, Document Chunking

**Computer Vision:** YOLO, Faster R-CNN, Detectron2, UAV Tracking, Object Tracking, Pose Estimation, Gait Analysis

**Machine Learning Models:** OpenAI GPT-4, Whisper, CLIP, VGG16, InceptionV3

**Data Engineering:** ETL, Data Preprocessing, SQL/NoSQL Databases, Redis

**Business Intelligence:** Power BI, Tableau

**Cloud Deployment:** GCP, AWS

## Research Paper

---

Moeza Noor, Hamza Bhatti, **Umar Abdul Rehman** et al, Hybrid Deep Learning for Early Diabetes Prediction using Explainable AI, Natural Computing and Applications Springer, July 2024, Under Consideration

## Courses and Certifications

---

- Oracle Cloud Infrastructure 2024 AI Foundations Associate (1Z0-1122-24)
- Oracle Cloud Infrastructure 2024 Data Foundations Associate (1Z0-1195-24)
- Oracle Cloud Infrastructure 2024 Foundations Associate (1Z0-1085-24)
- Coursera Deep Learning Specialization
- Coursera BI Foundations with SQL, ETL, and Data Warehousing Specialization