Muhammad Umair Farooq

AI Engineer

◇ +923084624629 ◇ farooq.intellecta@gmail.com ◇ LinkedIn ◇GitHub

SUMMARY

Machine Learning Engineer specializing in Generative AI, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and agentic AI systems. Experienced in designing and deploying production-ready AI applications including real-time voice agents, healthcare automation workflows, legal reasoning assistants, and domain-specific AI copilots. Skilled in Python, PyTorch, LangChain, FastAPI, and vector databases (FAISS, Pinecone), with hands-on experience integrating LLMs into scalable backend systems. Passionate about building reliable, explainable, and end-to-end AI solutions that solve real business problems through modern AI orchestration and deployment practices.

EXPERIENCE

AI Engineer — *ArcSirius, Islamabad*   *Nov 2025- Present*

Engineered an AI voice agent for car dealerships enabling automated appointment booking and customer handling through real-time speech recognition and LLM-driven conversations using LiveKit infrastructure.

Machine Learning Intern — *ZAA Soft, Islamabad*   *June 2024 – Aug 2024*

Built LLM-powered applications using OpenAI and Hugging Face models. Implemented FAISS and Pinecone databases for efficient retrieval and search. Contributed to end-to-end deployment of internal AI prototypes.

Machine Learning Intern — *Air University, Islamabad*   *Jul 2023 – Sept 2023*

Completed an enriching internship in the Department of Creative Technologies at Air University, where I honed my skills of Machine Learning.

Backend Developer — *128 Technologies, Islamabad*   *Jul 2023 – Aug 2023*

Created a management system for hospitals to keep track of their doctors' schedules and to provide appointments to patients relative to their emergencies. And gained hands-on experience in coding with python, problem-solving, and project management.

SKILLS

AI & Generative AI Technologies: Large Language Models, Retrieval-Augmented Generation, LangChain, OpenAI API, Transformers, Agentic Workflows, Vector Search
Machine Learning & Deep Learning: PyTorch, TensorFlow, Keras, Scikit-learn, Deep Learning, Model Training & Evaluation, Feature Engineering

AI Deployment & Backend Engineering: FastAPI, REST APIs, Docker, AI System Integration, Backend Development
Vector Databases & Data Management: FAISS, Pinecone, ChromaDB, MySQL, SQLite, Embedding Storage & Retrieval

Programming: Python, C++, C

PROJECTS

Invinci Voice Agent- A real-time conversational voice agent integrating STT, LLM dialogue management, and TTS to handle end-to-end appointment booking.*[Demo-Voice Agent](#)*

Agentic Healthcare Workflow Automation System – Built and deployed an agentic AI system that processes unstructured healthcare data (text, PDFs, scanned documents) using LLMs and OCR, validates information completeness, and executes decision-based workflows with full audit logging. *[Healthcare Agentic System](#)*

Mednix – AI-Powered Drugs Information Assistant – Built a FastAPI-based RAG system integrating OpenAI and medical databases to deliver accurate, context-aware drug information. *[Mednix-Click Me](#)*

Legal AI Assistant – Law Retrieval & Analysis System— developed a hybrid BM25 + FAISS retrieval engine with LLM summarization for intelligent legal document search and reasoning. *[Legal Agent-Click Me](#)*

FYP – Neuro-Flex – EMG-Based Hand Movement Classification— Designed a deep learning pipeline using LSTM models to classify EMG signals and control a 3D prosthetic hand in real-time

Credit Enrich System – Creditworthiness Automation: — Automated financial document parsing and data enrichment using OCR, regex, and AI-driven validation pipelines.

EDUCATION

BS Artificial Intelligence — *Air University, Islamabad* **—** CGPA: **2.71**   **Sep 2021 - July 2025**