Muhammad Umer Khan AI Engineer | Machine Learning Specialist

+92 3432187868 | muhammadumerk546@gmail.com | Portfolio | Github

Summary

AI Engineer experienced in designing, fine-tuning, and deploying scalable ML and LLM-powered systems. Proficient in building robust AI pipelines, integrating MLOps workflows, and leveraging tools like LangChain, Hugging Face, and Streamlit for NLP, generative, and predictive applications. Proven ability to deliver intelligent solutions that automate processes, extract insights, and solve real-world problems.

Skills

Programming & Tools: Python, SQL, Git, VS Code, JupyterLab
Machine Learning & AI: Supervised & Unsupervised Learning, Generative/Agentic AI, LLM Fine-Tuning, Prompt Engineering
Frameworks & Libraries: Scikit-Learn, TensorFlow, PyTorch, Hugging Face, LangChain / LangGraph
MLOps & Deployment: MLflow, FastAPI, Streamlit, Docker, CI/CD
Cloud & Infra: Microsoft Azure, Amazon Web Service (AWS)
NLP & LLMs: OpenAI API, Groq API, LangGraph Agents, LangSmith
Databases & Vector Search: PostgreSQL, MySQL, MongoDB, FAISS, ChromaDB
Visualization: Power BI, Matplotlib, Seaborn, Plotly
Languages: English (Professional), Urdu (Native)

Experience

AI Developer
Bright Solutions | August 2025 - Present | Hybrid
- Delivering end-to-end client AI products with a focus on automation, Generative AI, and agentic AI solutions.

AI Intern
SaylaniTech Limited | July 2025 - Present | Onsite
- Contributing to LLMs, NLP, and AI product development with modular pipelines, deployment, and API integration.

AI Intern
Visdalytics | May 2025 | Remote (1-month)
- Built AI chatbots, TTS tools, Roman-to-Urdu converter, and CV/ML pipelines (YOLOv8, Transformers, TensorFlow).

Projects

• DineMate: Agentic AI for Automated Food Ordering
Personal Project | 2025
- **Challenge:** Replaced **manual** food ordering with a **real-time**, multi-role AI system featuring voice input and analytics.
- **Action:**
  - Engineered **LangGraph-based multi-agent** backend to **orchestrate** orders, kitchen, and admin tasks.
  - Developed a **voice-enabled** food ordering **chatbot** using Qwen, Whisper ASR, Torch TTS, LangChain, and SQLite.
  - Integrated **SQLite** to simplify **local deployment** and **implemented secure**, **role-based authentication**.
  - Designed **analytics dashboards** to monitor **peak hours**, **sales trends**, and **customer spend patterns**.
- **Result:**
  - Boosted **ordering efficiency** by **30%**, reducing **manual operations** across all user roles.
  - Enhanced **decision-making** via **dashboards**, improving **business visibility** by **25%**. Live Demo.

• SupportGenie: Dual Fine-Tuned LLM Customer Support Chatbot
Personal Project | 2025
- **Challenge:** Upgraded a generic **RAG** bot into a **domain-specialized assistant** using **fine-tuned LLMs** on banking FAQs.
- **Action:**
  - Fine-tuned **Mistral-7B** and **LLaMA-3 (8B)** using **QLoRA** on a custom Hugging Face-hosted dataset.
  - Built a Streamlit app with **MongoDB**, **FAISS** retrieval, **dual-LLM inference**, and **sentiment analytics.**
  - Benchmarked model outputs, selecting **Mistral** for its **domain alignment** and concise **factual responses**.
  - Released a Colab **notebook** and integrated real-time usage **dashboards** with sentiment tracking.
- **Result:**
  - Improved answer relevance by **20%** versus **base models** through targeted fine-tuning and **validation.**
  - Deployed a **public chatbot** with **dual inference**, **feedback analytics**, and FAQ explainability. Live demo.

- **Diagnosify: LLM-Powered Medical Report Insights** *Hackathon Project | 2025*
  - **Challenge:** Engineered a reliable **AI system** to **interpret** and **explain** unstructured medical reports using **LLMs**.
  - **Action:**
    - Combined **OCR**, **prompt-tuned** LLMs, and **PDF parsing** to extract structured clinical data.
    - Built a **RAG-enabled chatbot** for real-time, document-grounded medical **question answering**.
    - Integrated **RAGAS** to **evaluate answer faithfulness** and ensure response accuracy.
    - Logged all **chatbot interactions** and **metrics** in **MongoDB** for auditability and **analysis.**
  - **Result**:
    - Delivered a **production-ready health AI tool** within 12 hours during live hackathon.
    - Demonstrated trustworthy, domain-specific LLM performance with built-in evaluation logging. Live demo.
- **ChurnSage: MLOps-Driven Customer Retention System** *Personal Project | 2025*
  - **Challenge:** Predicted **customer churn** from **imbalanced** telecom data with scalable **ML infrastructure**.
  - **Action**:
    - Built a full **ML pipeline** with SMOTEENN, encoding, scaling, and **hyperparameter tuning** using **GridSearchCV.**
    - Trained and tracked Logistic, Random Forest, and SVC models using **MLflow** for **reproducibility.**
    - Leveraged **Azure Cloud** Notebooks for scalable training and managed experiment logging.
    - Deployed the best model with **FastAPI** (REST API + SQLite) and integrated **LLMs** to explain user churn from reviews.
  - **Result**:
    - Achieved **~94% test accuracy** with robust handling of class imbalance and **optimized modeling**.
    - Delivered a **CI-ready**, **batch-enabled** ML pipeline with LLM explainability. Live demo.

Certifications

- Deep Learning Specialization - **DeepLearning.AI** (Coursera), 2025
- Natural Language Processing Specialization - **DeepLearning.AI** (Coursera), 2025
- Machine Learning Professional Certificate - **IBM** (Coursera), 2024
- Artificial Intelligence and Data Science - **S.M.I.T.**, 2025

Education

Sindh Madressatul Islam University *Karachi, Pakistan*

B.S. in Computer Science

- **Relevant Coursework:** Machine Learning, Deep Learning, NLP, Data Structure & Algorithms, Web Technologies.

Achievements

- **Top 10 Finalist** – SMIT AI Hackathon (2025) for Diagnosify: Medical Insight System.