USAMA YASEEN

Data Scientist | ML ENGINEER | AI ENGINEER

Johar Town, Lahore| usamayaseen867@gmail.com | linkedin.com/in/usamayaseen867 | 0348-6622291

PROFILE SUMMARY

Experienced Data Scientist specializing in machine learning, deep learning, and large language models (LLMs), with 2 years of hands-on experience. Proficient in Python, NLP, and advanced AI frameworks, with expertise in LLM agents, LangChain, LangGraph, LangFlow, and AI-driven data processing. Skilled in deploying AI solutions on cloud platforms and leveraging big data tools like Hadoop and Apache Spark. Adept at transforming complex insights into actionable strategies and delivering data-driven solutions. Driven by a passion for advancing innovation in AI and data science.

EDUCATION

University of Central Punjab, Lahore

- BS Computer Science

PROFESSIONAL EXPERIENCE

Data Scientist at AugAI, Pakistan   May 2024 - Present

- Served as Team Lead, managing project lifecycles and ensuring the successful delivery of AI solutions.
- Led multiple high-impact AI initiatives, driving measurable improvements in business outcomes.
- Developed and deployed machine learning models, significantly enhancing predictive accuracy.
- Trained and mentored junior data scientists, fostering a culture of learning and collaboration.
- Partnered with diverse teams to efficiently integrate AI into existing products.

Researcher: AI-Powered Research Assistant

- Developed an AI-powered tool for automated research, web scraping, and document generation.
- Handles natural language queries and provides research-based responses.
- Uses BeautifulSoup, Scrapy, and Selenium for web scraping and data retrieval.
- Performs in-depth searches for specialized content.
- Generates downloadable PDF and Word reports using PDFKit and python-docx.
- Integrates GPT-3/4 for summarizing and formatting responses with clarity.
- Built with FastAPI for query handling and data retrieval.
- Efficient research tasks by automating data collection and report generation, enhancing research efficiency.

Data Analyst Agent with Automated Graph Generation

- Developed an agent for automating data analysis, graph generation, and descriptions using **Python** and **FastAPI**.
- Enables CSV file uploads and stores data in **SQLite**.
- Analyzes user queries to suggest relevant graphs using **Matplotlib** and **Pandas**.
- Generates graphs and visualizations with **Matplotlib** and **Pandas**.
- Provides natural language descriptions using **GPT-3/4** via **LangChain**.
- Simplified data analysis through automated visualizations and insights.

Financial Bot for Valuation and Earnings Call Analysis

- Designed a financial bot for real-time valuation and earnings updates.
- Integrated MongoDB for robust database management.
- Utilized embedding models and vector databases for efficient data retrieval.
- Implemented GPT-4 mini for intelligent financial text analysis.
- Developed a secure, high-performance API with FastAPI, incorporating API keys and JWT authentication.
- Enabled agentic architecture for managing multiple financial bots.
- Gained expertise in full-stack development, machine learning, and secure API design.

Financial Assistance AI

- Developed an AI-powered financial forecasting and advisory system.
- Implemented machine learning models to predict personal expenses with high accuracy.
- Created an intelligent chatbot using RAG and fine-tuned LLMs to provide personalized financial advice.
- Integrated historical spending data analysis for context-aware recommendations.
- Resulted in significant improvement in users' financial health over a 6-month period.
- Technologies used: Python, TensorFlow, LangChain, RAG, AWS SageMaker.

Fraud Detection System

- Architected and implemented a comprehensive fraud detection system for banking transactions.
- Consumed advanced anomaly detection algorithms, achieving a high fraud detection rate.
- Designed and deployed the entire system on AWS, leveraging SageMaker, S3, and CI/CD pipelines.
- Implemented an explainable AI bot using Claude model to provide reasoning for fraud flags.
- Considerably reduced false positives compared to the previous system. Technologies used: Python, AWS SageMaker, S3, CloudFormation, Anthropic's Claude API.

PROJECT EXPERIENCE

University Of Central Punjab
August 2023 – May 2024

- Led the development of **ChestXGPT**, transforming chest X-ray analysis with Deep Learning.
- Designed and optimized Deep Learning models using CNNs, RNNs, and LSTMs for medical imaging.
- Built **MediBot**, a custom medical chatbot leveraging LangChain, RAG models, and Llama LLM.

- Integrated advanced Vector Databases for efficient data storage and retrieval in MediBot.
- Deployed AI solutions on ChainLit for secure and decentralized access to medical data.
- Mentored team members, fostering skill development and enhancing team productivity.
- Gained expertise in AI tools like LangChain, RAG, and LLM through independent research.   **ChestXGPT**

- Innovated with Deep Learning to create ChestXGPT, reshaping chest X-ray analysis.
- Engineered efficient data processing techniques, optimizing Deep Learning model performance.
- Gained experience in Deep Learning methodologies like CNNs, RNNs, and LSTMs for medical image analysis.

MediBot - Custom Medical Chatbot

- Developed MediBot, a specialized medical chatbot, integrating LangChain for text embedding and Fassi Vector Database for efficient data storage.
- Implemented the RAG model for context-based information retrieval and Llama LLM for generating responses.
- Deployed MediBot on ChainLit, ensuring secure and decentralized access to medical information.
- Gain proficiency in Chainlit, RAG, LangChain, Vector Database, and LLM through independent research and implementation, contributing to the success of the project.

TECHNICAL SKILLS

- **Machine Learning & Deep Learning**: TensorFlow, PyTorch, Scikit-learn, Keras, XGBoost, LightGBM, CNNs, RNNs, Transformer architectures, Transfer learning techniques, Reinforcement Learning.
- **Natural Language Processing**: Hugging Face Transformers, BERT, T5, RoBERTa, GPT models.
- **Data Processing & Engineering**: Pandas, NumPy, Spark, Airflow, Data Annotation.
- **Model Evaluation & Tuning**: GridSearchCV, RandomizedSearchCV, Optuna; Metrics like Precision, Recall, ROC AUC.
- **Statistical Methods**: Bayesian inference, Hypothesis testing, A/B testing, Statistical modeling.
- **Data Visualization**: Matplotlib, Seaborn, Plotly, Tableau.
- **Vector Databases**: FAISS, Pinecone, Weaviate, Chroma.
- **Cloud Platforms**: AWS (SageMaker, Lambda, S3 , EC2).
- **Big Data Tools**: Hadoop, Apache Kafka.
- **Version Control & Collaboration**: Git, GitHub, GitLab.

CERTIFICATIONS

- ChatGPT Prompt Engineering for Developers
- Building Multimodal Search and RAG
- Intro to Deep Learning
- Computer Vision
- Deep Learning Specialization