

Muhammad Sahil

✉ muhammadsahil564@gmail.com ☎ 92 308 2575838 in sahil-gul 🌐 portfolio

CAREER SUMMARY

- Experienced and results-driven Machine Learning Engineer with 4+ years of hands-on experience in developing and implementing data-driven, AI-based solutions.
- Proven expertise in data analytics, advanced data-driven methods, and the end-to-end development of machine learning and deep learning pipelines.
- Experienced in AI-driven product development, Data Solutions, and specializing in NLP, multi-agent systems, and deep learning pipelines.
- Expertise in LLMs(Gemma, Llama, Phi, Qwen), Retrieval-Augmented Generation (RAG), and classification models.
- Proven ability to collaborate cross-functionally, lead AI projects using Agile development methodologies, and communicate complex models to non-technical stakeholders.
- Strong background in Production Environment with cloud deployment (AWS, Azure), MLOps, and scalable AI architectures.

SKILLS

Technical Skills: AI/ML, Python, SQL, BASH, NLP, LangChain, LangGraph, Crewai, Multi-Agentic Workflows, PyTorch, TensorFlow, PostgreSQL, Natural Language Processing, Flask, FastAPI, DevOps, MLOps, CI/CD, Serverless APIs, Software Development, Container Apps, Model Inference APIs, MLFlow, AWS, Azure, AWS S3, Docker, Kubernetes, MicroK8s

Soft Skills: Communication Skills, Team Collaboration, Analytical Thinking, Willingness to Learn, Adaptability, Problem Solving, Critical Thinking, Decision-Making, Analytical Skills

PROFESSIONAL EXPERIENCE

Consultant Data Science
Systems LTD

Lahore
Mar 2024 – Present

- Led the design and implementation of an AI-powered medical coding system for a healthcare startup, using LangChain and LangGraph to automate and streamline coding processes.
- Built a multi-agent workflow with open-source LLMs like Gemma 2B/7B, Llama3, and Phi-3 to improve accuracy in medical coding classification tasks.
- Developed a multi-agent workflow for classification tasks using state-of-the-art open-source large language models, including GEMMA2, GEMMA 27B, Llama3, PHI 3, PHI 3.1, ALOE, and Meerkat from Hugging Face.
- Built a knowledge base by collecting data with Selenium and connecting it to vector databases, Neo4j using cypher query language for quick and accurate information retrieval.
- Utilized open-source embedding models to enhance vector search accuracy and scalability for the classification workflow.
- Created two multi-agent systems for accurate SOAP note classification—one specializing in nephrology-specific procedures and the other for Evaluation and Management (EM) coding—leveraging LangGraph for efficient state management.
- Developed two specialized systems for medical coding using LangGraph :
 - **Nephrology coding:** Focused on classifying complex kidney-related procedures.
 - **E&M coding:** Automated billing code assignments to reduce errors and save time.
- Achieved 80%+ accuracy in classifying SOAP notes into procedural and non-procedural categories, ensuring high-quality medical coding outputs.
- Tested and optimized open-source LLMs for on-premises deployment, ensuring full compliance with healthcare data privacy regulations.

- Implemented a seamless end-to-end workflow for CPT code assignment, reducing manual effort and improving coding efficiency.
- Designed a de-identification tool using Gemma/Llama3 to protect patient data and meet regulatory requirements, using NoSQL.
- Added support for multimodal inputs (text, images, PDFs) to streamline medical report uploads and processing.

Machine Learning Engineer
NKU Technologies PVT LTD

Lahore
 Aug 2023 – Mar 2024

- Developed O3Chatbot for real-time historical information retrieval.
- Utilized advanced prompt engineering with a focus on RAG (Retrieval-Augmented Generation).
- Implemented vector databases to enhance contextual understanding.
- Capable of generating charts and tables based on user queries.
- Designed and deployed NLP classification models to improve O3Chatbot responses.
- Created a stateless API endpoint using Azure Functions for scalability.
- Deployed the O3Chatbot application on the Azure Cloud platform for accessibility and reliability.

Machine Learning Engineer
iCustoms

Lahore
 Sep 2022 – Aug 2023

- Developed and deployed infrastructure for predictive models for the classification of HS codes in an AI-based Customs Declaration System (CDS).
- Utilized neural networks, transfer learning, Named Entity Recognition (NER), RNN, LSTM, and fine-tuned BERT models for multi-class classification.
- Designed and deployed NLP models on AWS EC2, enhancing system efficiency and accessibility for predictive modeling.
- Integrated RESTful APIs for seamless communication between cross-functional teams, including Backend and Frontend development.
- Used Azure services for data collection, data scraping, data preprocessing, data preparation, data mining, data analysis, and text processing
- Conducted model training, evaluation, and optimization to improve classification accuracy using Pandas, NumPy, Matplotlib, and Scikit-Learn.
- Implemented testing, experiment tracking, model registry, and monitoring alerting for production models to ensure system reliability and performance.

Associate Data Scientist
Bling.ai

Lahore
 Dec 2021 – Aug 2022

- Worked on data modeling and the development of facial recognition systems in Computer Vision using different ML algorithms.
- Created a web application capable of recognizing multiple faces in images and handling large datasets with 100+ faces using Facenet, and OpenCV.
- Developed the backend of the website using the Flask framework with Postgres SQL, ensuring seamless integration and user experience.

ML Internee
Devicon

Lahore
 Mar 2021 – Nov 2021

- Contributed to an AI-based recommendation system, including data collection, data preparation, feature engineering, clustering, and data modeling.
- Deployed the model using Flask API for efficient access and utilization.

Education

KFUEIT

2017 – 2021

BS Computer Engineering

- **Coursework:** Computer Vision, Machine Learning, Deep Learning, Computer Science, Statistics, Software Engineering, Artificial Intelligence