Abdul Wajid
Karachi, Sindh, Pakistan.
+92-311-0351773 engineer.wajid.ds79@gmail.com

**EDUCATION** LinkedIn GitHub Kaggle Hugging Face learn.microsoft

**Mehran University of Engineering Technology, Jamshoro, Pakistan** **2020 - 2024**
*B.E., Computer Systems Engineering*

**PROFESSIONAL EXPERIENCE**

**AI Engineer**
X Flow Software Technology LLC (Bluescarf Artificial Intelligence)
Full Time | May 2025 – Present

- Fine-tuned LLMs (GPT, Claude, Llama, Mistral) using LoRA, QLoRA, PEFT, increasing model efficiency by 28%.
- Created data pipelines using Apache Spark and Databricks, reducing ETL processing time by 35%.
- Built RAG systems with Pinecone and Chroma, improving query response accuracy by 40%.
- Deployed ML pipelines with MLflow integrated into CI/CD workflows, achieving 99.5% system uptime.

**Abdul Majid Bhurgri Institute of Language Engineering (AMBILE), Hyderabad**
Under the administrative control of the Culture, Tourism, Antiquities & Archives Department, Government of Sindh

- **Key Projects:**

  - **Sindhi Bhittai ChatBot** – RAG-based Sindhi chatbot using LangChain on Bhittai's works.
  - **Sindhi Law Assistant (AI Wakeel)** – Legal Q&A chatbot in Sindhi with custom NLP dataset.
  - **Sindhi Speech-to-Text Model** – Fine-tuned Transformer ASR for Sindhi transcription.
  - **Sindhi OCR** – Trained Tesseract 5 on 10K pairs; achieved 45% WER, 19% CER.
  - **Data Cleaning Pipeline** – Regex + Spark/Pandas workflows; boosted data cleaning by 75%.
  - **Sindhi LLM Data Pipeline** – Preprocessing/tokenization for 118M-token Sindhi corpus.
  - **Sindhi dataset**: Ambile dataset made and managed by me

**PROJECTS**

*Final Year Project (FYP): Automated Chest X-ray Report Generation*
- Developed an AI-powered system to classify chest X-ray images into four categories: Normal, Pneumonia, Tuberculosis, and COVID-19, achieving a test accuracy of 92.58% using VGG16.
- Integrated OpenAI's GPT-4 API to generate professional, contextually relevant radiology reports.
- Built a user-friendly interface with Radio, enabling seamless interaction for clinicians.
- Conducted model training and evaluation on a combined dataset of 13,500+ X-ray images from Kaggle.
- Applied advanced deep learning techniques and frameworks, including TensorFlow and PyTorch, for robust model implementation.

*Prompt Optimization Engine*
- Applied prompt engineering, Langchain, and GPT-4 to optimize AI responses.

*AI E-commerce Chatbot*
- Developed using Langchain and GPT-4 for enhancing online shopping experiences. *RAG Based on AI Support Chatbot*
- Integrated Pinecone vector database with GPT-4 for dynamic query responses. *AI Personal Document Chat Model*
- Utilized Langchain and Chroma vector database, Gemini - Google's large language model. *Text to Image Generation Web App*
- Created using Dall-E 3, Flask, showcasing generative AI capabilities.

## TECHNICAL SKILLS

- **Programming Languages**: Python
- **Machine Learning**: Supervised and Unsupervised Learning, Classification, Regression, Clustering, Anomaly Detection
- **Deep Learning**: CNNs, RNNs, LSTMs, BERT, GPT-3.5/4, Gemini
- **Natural Language Processing (NLP)**: Word2Vec, BERT, GPT-NER, RAG, LangChain
- **Vector Databases**: Pinecone, Chroma, FAISS, Weaviate, Milvus
- **Cloud & Deployment**: AWS, Azure, Docker, CI/CD
- **Frameworks & Tools**: PyTorch, TensorFlow, LangChain, Flask, Streamlit, Gradio
- **Big Data Technologies**: Spark, PySpark
- **Database Management**: SQL, MySQL, PostgreSQL
- **Others**: Git/GitHub, VS Code, Jupyter Notebooks, Google Colab

## CERTIFICATIONS

•**Introduction to Python**: DataCamp & hackerrank
•**Foundations of Data Science**: Google, Coursera
•**AI Agents Fundamentals AI Agents Fundamentals** Hugging Face