




# Mohsin Shah

(Data Science, Machine Learning, NLP, GenAI)

 engineer.mohsinshah@gmail.com

 +923139373802

 Islamabad, Pakistan

 linkedin.com/in/engineer-mohsin-shah

 github.com/Engineer-Mohsin-Shah

## About me

---

I am a dedicated Computer Software Engineering graduate with a strong foundation in Data Science, Machine Learning, and NLP, complemented by practical experience in developing innovative AI solutions. My background includes hands-on work with IBM courses and projects that have strengthened my skills in supervised and unsupervised learning, deep learning models, and large language models (LLMs). I have experience in fine-tuning LLMs, building Retrieval-Augmented Generation (RAG) systems, and creating chatbots using LangChain, combining these tools to deliver advanced, interactive applications. Additionally, I am well-versed in NLP technologies like TTS, STT, and Hugging Face. I am seeking a Junior Data Scientist, ML Engineer, or AI Engineer role where I can leverage my technical expertise and passion for problem-solving to contribute to forward-thinking projects in a collaborative, AI-driven environment.

## Experience

---

### AI/ML Engineer (Merik Solution)

07/10/2024-Present

- working on a Medical Billing and Coding project focused on automating billing processes and mapping medical codes for improved accuracy and compliance.
- Preparing detailed project documentation, including workflows, implementation steps, and guidelines to streamline project understanding and integration.

### NLP Internship (Itsolera PVT LTD)

07/2024 - 10/2024

- During my internship, I gained hands-on experience in Natural Language Processing (NLP) using various tools and libraries, including NLTK, Spacy, Gensim, and Hugging Face.
- I worked on key projects that involved text summarization, text-to-speech (TTS) generation, Speech to Text (STT) generation, Topic modeling and Chatbot Development.
- One of my significant accomplishments was successfully completing a project focused on text-to-speech generation, where I implemented and fine-tuned models to convert text into natural-sounding speech.

### Machine Learning Internship (Ezitech Institute)

04/2024 - 08/2024

- Conducted Exploratory Data Analysis (EDA) and utilized Scikit-learn for various algorithms.
- Used TensorFlow and Keras to build and train ANN, CNNs, and pre-trained models.
- **Worked on projects:** Cotton Plant Disease Prediction, Brain Tumor Detection, Dog vs Cat classification, and Dubai properties dataset ETL.

## Technical Skills

---

- **Data Science:** Pandas, NumPy, Matplotlib, Seaborn
- **Machine Learning:** Scikit-Learn, TensorFlow, Keras, Pytorch, CNN, ANN, RNN, LSTM, GRU
- **GenAI & NLP:** NLTK, TF-IDF, TTS, Transformers, Huggingface, SpaCy, STT, LangChain, RAG, Fine-tuning, Chatbots development, CrewAI, Groq, Llama, Gemini, BERT
- **Model Deployment:** FastAPI, Streamlit, Gradio
- **Data Base:** MYSQL, Postgres, Vector Database (FIASS, Chroma etc.)

- **Language:** C++, Python, SQL

## Projects

---

### Smart Doorbell (FYP) (IOT)

- **Description:** Developed a smart doorbell that captures visitor images and sends them to the user via email. The user can view the image and unlock the gate through a mobile app.
- **Technologies:** Raspberry Pi 4, Camera, Gmail API, Python, OpenCV

### Heart Attack Prediction (ML)

- **Description:** Analyzed a heart disease dataset, visualized data insights, and built a machine learning model to predict the likelihood of heart disease.
- **Technologies:** Python, Scikit-Learn, Panda, Numpy, TF-IDF, Naive Bayes classifier

### Cotton Plant Disease Prediction (DL)

- **Description:** Developed a model to identify diseases in cotton plants by analyzing images with Convolutional Neural Networks (CNNs).
- **Technologies:** Python, Tensorflow, Keras, CNN (Resnet50, VGG, etc.)

### Text-To-Speech Generation and Text-To-Speech (NLP)

- **Description:** TTS and STT generation, where I implemented and fine-tuned models to convert text into natural-sounding speech and speech to Text. Text-to-speech generation of 17 different languages, and Speech-to-text generation of 100 different languages
- **Technology:** Python, TTS model of hugging face, XTTS-v2, whisper, Gradio etc

### Chatbot development (NLP, Gen AI, BERT, LLAMA)

- **Description:** Multi Document RAG Chatbot, QA Chatbot, Finetuned Distilled BERT for Sentiment Analysis, Text Summarization etc
- **Technologies:** Langchain, FAISS, Chroma, Gradio, FastAPI, Streamlit, LORA, RAG, Transformers, LLAMA, BERT, Gemini

## Education

---

### BSc Computer Software Engineering

University of Engineering and Technology Mardan

**2018 - 2022**

### FSC Pre-Engineering

Government Degree College Mingora

**2015 –2017**

## Certificates

---

- **IBM Data Science Professional Certificate:** [Link](#)
- **Machine Learning Specialization by Stanford:** [Link](#)

## Languages

---

English, Urdu, Pashtu