

Umair Yousaf

23100053@lums.edu.pk | [Portfolio](#) | [LinkedIn](#) | [GitHub](#) | Lahore

EDUCATION

Lahore University of Management Sciences (LUMS)

BS Computer Science – CGPA 3.6

Class of 2023

Lahore

EXPERIENCE

Reteta, a Visionet Subsidiary (sister company of Systems Limited)

Junior Consultant - AI Team

Nov 2024 – Present

Cranbury, NJ, USA · Remote

- Developed NER-based techniques for anonymizing Personally Identifiable Information (PII) in patient-doctor conversations to ensure compliance with data privacy regulations and ethical standards.
- Designed and implemented methods for generating accurate and structured SOAP notes from encrypted patient-doctor interactions, involving extensive data preprocessing and model training to enhance medical documentation workflows.

Systems Limited

Junior Consultant - AI Department

July 2024 – Present

Lahore · On Site

- Developed a comprehensive **PDF parsing pipeline** capable of handling regular text, headings, subheadings, paragraphs, tables, images, and formulas, utilizing publicly available SOA models.
- Integrated the pipeline with existing **RAG** architectures, resulting in significant enhancements in information retrieval and document querying.
- Developed a clothing recommendation system utilizing **transformer models** and re-ranking techniques.
- Conducted multiple live client demonstrations of our solutions.

Systems Limited

Associate Consultant - AI Department

July 2023 – July 2024

Lahore · On Site

- Created a suite of **Generative AI** solutions for the insurance industry, streamlining operations at every phase of the insurance process. Developed the frontend in **React**, backend in **Flask**, and all **ML & AI pipelines**.
- Used **multimodal LLMs** for **VQA** and image-based information extraction, surpassing in-house OCR methods.
- Developed multiple end-to-end **RAG** pipelines using LangChain and LangGraph.
- Developed a reverse image search module for clothing articles using **Siamese Networks** and contrastive loss.
- Conducted research on SOA methods, hardware/software debugging, and created local LLM inference pipelines.

Teaching Assistant (LUMS)

TA for the course Data Structures

Jan. 2023 – May 2023

Lahore · On Site

- Held office hours, created and checked assignments and quizzes

Aletheia-AI (now BrickAndMortar.AI)

ML and SE Intern

Jun. 2022 – Aug. 2022

Lahore · On Site

- Multi-class vehicle detection, number plate detection, person detection and FR using **YOLOv5**
- Dataset creation, validation, experimentation and model deployment
- Created a dashboard using **React.js** and **Material UI** for real-time monitoring

Wateen Telecom

SE Intern

Jun. 2021 – Jul. 2021

Islamabad · On Site

- Developed a python-based desktop application for calculating cost of deployment of fiber optic cables

PROJECTS

LEGAL-UQA | *Submitted to NAACL 2025 - Under Review*

- Developed the first Urdu legal question-answering dataset with 619 question-answer pairs derived from Pakistan's constitution and corresponding legal article contexts.
- Implemented OCR extraction, manual refinement, and **GPT-4-assisted** translation to create a parallel Urdu-English dataset for domain-specific NLP.
- Evaluated state-of-the-art models, achieving **99.19% human-evaluated accuracy** with Claude-3.5-Sonnet and fine-tuned **mt5-large-UQA-1.0** for adaptation to legal domains.

Early Warning Detection System for Forest Fires | *Final Year Project*

- Employed K-Shot learning on **Siamese Networks** to classify images into 2 categories. Achieved best accuracy of 83% and F1 score of 0.827
- Generated synthetic dataset using video games and **Unreal Engine 4** for use in training other models
- Experimented using a variety of state of the art models such as **Deformable DETR**, **YOLO** etc.
- Submitted a paper to the **ICASSP Conference**

Sentiment Classification using BERT on GoEmotions | *NLP Project*

- Multi-label sentiment classification on GoEmotions dataset using HuggingFace framework

Lane Change Warning System | *Computer Vision/Machine Learning Project*

- Developed a pipeline to read a dashcam video's frames, detect vehicles using **YOLOv7** and lanes using **LaneNet**
- Employed **auto dynamic homography** to generate a top down view and detect any vehicles changing lanes

TECHNICAL SKILLS

Languages: Python, JavaScript, C++, C, Solidity, SQL, Haskell

Python Libraries: LangChain, LangGraph, PyTorch, TF, HuggingFace, Scikit-Learn, OpenCV, Pandas, NumPy

Software Engineering Frameworks: MERN Stack, MySQL, Flask

COURSES TAKEN AT LUMS

Deep Learning | Machine Learning | Computer Vision | Natural Language Processing | Data Mining | Data Science | Statistics and Data Analysis | Applied Probability | Human Computer Interaction

HONORS & AWARDS

Dean's Honor List in Freshman, Sophomore and Junior year on maintaining a CGPA above 3.6
100% Merit scholarship in A Levels (both years)