Laraib Zafar

AI Engineer & Data Scientist

Phone: +92 3258009299 | Email: laraibzafarlaraib@gmail.com | Portfolio: laraibzafar.me
LinkedIn: linkedin.com/in/laraib-zafar-5465a5267

Professional Summary

Final-year Data Science undergraduate with hands-on experience applying machine learning and statistical modelling to real-world problems across computer vision, NLP, and clinical prediction. Skilled in building end-to-end ML pipelines, conducting exploratory data analysis, and deploying data-driven solutions using Python, Scikit-learn, and deep learning frameworks. Experienced with scalable data infrastructure, including PostgreSQL, MongoDB, and Redis. Seeking an entry-level Data Scientist role to extract meaningful insights, build predictive models, and contribute to data-driven decision-making in production environments.

Education

Punjab University College of Information Technology (PUCIT)   **Lahore, Pakistan**

Bachelor of Science in Data Science   *2022 – June 2026 (expected graduation)*

– Relevant Coursework:
Big Data Analytics, Advanced Statistics, Data Warehousing, Machine Learning, Deep Learning, Computer Vision, NLP, Data Structures & Algorithms, Object-Oriented Programming.

Punjab Group of Colleges   **Lahore, Pakistan**

Intermediate in Computer Science (with Physics)   *2020 – 2022*

Distinction: 3rd Position, Lahore board

Technical Skills

ML & AI: Machine Learning, Deep Learning, Computer Vision, NLP, LLMs, Generative AI, Retrieval-Augmented Generation (RAG)

Programming Languages & Libraries: Python, NumPy, Pandas, Scikit-learn

Databases & Storage: PostgreSQL, MongoDB, Redis, Vector Databases (basic)

Data Engineering: ETL Pipelines, Data Wrangling, Exploratory Data Analysis (EDA), Feature Engineering

Backend & System Integration: REST APIs, API Integration, Real-Time Systems, SaaS Architecture

AI Tools & Frameworks: OpenAI API, AssemblyAI, Deepgram, OpenCV, Twilio

DevOps & Deployment: Git, GitHub, Docker, CI/CD Pipelines, System Deployment

Experience

Teacher Assistant – Object Oriented Programming   **Jan 2026 – Present**

Punjab University College of Information Technology | Lahore, Pakistan

– Conducted lab sessions and guided students through OOP concepts including classes, inheritance, polymorphism, and design patterns.
– Evaluated assignments and provided structured feedback to strengthen students' software engineering fundamentals.

Teacher Assistant – Data Structures & Algorithms   **Sep 2025 – Dec 2025**

Punjab University College of Information Technology | Lahore, Pakistan

– Assisted in delivering DSA lab content covering sorting algorithms, trees, graphs, and dynamic programming.
– Supported student assessments and helped build stronger analytical and problem-solving skills among peers.

Projects

ONIX – Organized Notetaker and Insight Extractor | Final Year Project

Tools: AssemblyAI, Playwright, TypeScript, Node.js, React/Next.js, Docker, Google Meet APIs

– **Challenge:** Manual note-taking and action-item extraction from meetings is time-consuming and error-prone at scale.
– **Action:** Developed a full-stack AI meeting assistant that scrapes live meeting captions, generates concise summaries, and extracts action items using AssemblyAI. Built scalable backend services and real-time data pipelines with a React/Next.js interface and Docker-based deployment.
– **Result:** Delivered a production-ready meeting intelligence platform enabling real-time insight extraction for end users.

Polyglot Meeting Data Pipeline (Big Data & Architecture)

Tools: Redis, MongoDB, PostgreSQL, BullMQ, Node.js, TypeScript, Prisma, Mongoose

Challenge: Meeting captions arrive as a continuous high-volume stream across concurrent sessions, requiring low-latency ingestion, durable transcript storage, and asynchronous post-processing without blocking the capture path.

Action: Designed a three-database architecture — Redis for real-time caption buffering with distributed locking and atomic flushes, MongoDB for chunked batch transcript persistence, and PostgreSQL for relational metadata and analytics. Implemented a BullMQ job queue with deduplication and a Dead Letter Queue for failed writes.

Result: Delivered a fully staged, fault-tolerant pipeline capable of handling high-volume concurrent meetings, with atomic processing guards preventing duplicate jobs and a processing worker computing speaker stats, summaries, and action items post-meeting.

AI Voice Agent – Pharmacy Assistant

Tools: Twilio, Deepgram, OpenAI GPT, Python

- **Action:** Developed an AI-powered voice agent for a pharmacy that handles inbound calls, provides real-time information on medicines and pricing, and enables phone-based order placement.
- **Result:** Delivered a fully functional conversational AI system integrating real-time speech recognition with LLM reasoning.

Palm Vein Authentication System

Tools: Python, Deep Learning, OpenCV, Scikit-learn, Siamese Neural Networks, Gabor Filters, LBP

Action: Built a contactless biometric authentication system using Near-Infrared (NIR) imaging, implementing and comparing two approaches — classical methods (Gabor filters and Local Binary Pattern feature extraction) and a Siamese Neural Network for deep learning-based biometric matching. Evaluated both pipelines on the VERA-Palmvein dataset.

Result: Delivered an end-to-end authentication pipeline demonstrating strong understanding of both traditional computer vision fundamentals and modern deep learning techniques for secure, contactless identity verification.

Facial Emotion Detection

Tools: Python, OpenCV, Scikit-learn, NumPy, Pandas

- **Action:** Built an image-based facial emotion recognition system using computer vision and machine learning to classify emotions from facial expressions in real time.
- **Result:** Achieved reliable multi-class emotion classification across core facial expression categories.

Predictive Analysis of Thalassemia

Tools: Python, Scikit-learn, Statistical Modelling

- **Action:** Conducted research on predicting Thalassemia Major using multiple machine learning and statistical models, analysing patient data to improve early diagnosis accuracy.
- **Result:** Produced a comparative analysis of model performance to identify the most effective clinical prediction approach.

Space Missions Analyzer

Tools: Python, Streamlit, Pandas, Matplotlib

- **Action:** Developed an interactive Streamlit web app showcasing an Exploratory Data Analysis (EDA) of a comprehensive dataset covering all historical space missions.
- **Result:** Delivered a filterable dashboard enabling users to explore mission timelines, success rates, and agency-level trends.