Mutahar Ali

mutahar789.github.io | mutahar.ali@lums.edu.pk | linkedin.com/in/mutahar789
github.com/mutahar789 | +92 331 0495560

Education

Lahore University of Management Sciences — Lahore, Pakistan
BS Computer Science — *Aug. 2019 – June 2023*

- CGPA **3.73/4.00** | Major CGPA **3.84/4.00**
- Graduated with Distinction
- Placed on Dean's Honor List (2019-2022)
- Relevant Coursework: **Systems and Networking:** Distributed Systems, Operating Systems, Network-Centric Computing, Network Security, Blockchain, Topics in Internet Research | **AI:** Deep Learning, Machine Learning, Computer Vision Fundamentals, Data Science | **Theory:** Algorithms, Discrete Mathematics, Theory of Automata | **Architecture:** Computer Organization and Assembly Language, Fundamentals of Computer Systems

Research Experience

Research Associate — Jul. 2023 – Present
Lahore University of Management Sciences

- Working on developing scalable software for LLM-integrated applications.

Undergraduate Research Assistant — Jun. 2022 – Aug. 2022
Lahore University of Management Sciences

- Worked on enhancing robustness in federated learning environments.

Research Intern at CSaLT — Jan. 2022 – Dec. 2022
Lahore University of Management Sciences

- Worked on developing an Image-to-Speech System for Urdu Language.

Undergraduate Research Assistant — Jun. 2021 – Dec. 2021
Lahore University of Management Sciences

- Worked on enabling real-time mobility support for edge applications in 5G.

Research Projects

EdgeCat — Jun. 2021 – Dec. 2023
Advisor: Dr. Zafar Ayyub Qazi (LUMS)

- Explored and evaluated existing techniques for application session migration.
- Implemented reactive app session migration using CRIU, LXC, and Docker, and iterative app session migration with CRIU and LXC, using rsync for proactive app state synchronization with the target edge host.
- Proposed and implemented a method for accurate measurement of application downtime using time logging and NTP-based time synchronization.
- Performed experiments to measure application downtime for CarMap, an open-source edge app, during an app handover when using CRIU and LXC. Results showed that existing techniques for application session migration are slow.
- Identified limitations in existing techniques for application session migration in terms of generality and timeliness.
- Contributed in the selection of real edge applications for evaluation of our design.

Exploring the Impact of System Heterogeneity on Federated Learning — Jun. 2022 – Apr. 2023
Advisors: Dr. Ihsan Ayyub Qazi, Dr. Zafar Ayyub Qazi, Dr. Agha Ali Raza (LUMS)

- Investigated the impact of system heterogeneity on FL by conducting experiments on a small-scale real testbed of heterogeneous mobile devices.
- Implemented CPU profiling by using adb to collect CPU data from /proc/stat at 1 second intervals, and building detailed temporal traces by parsing this data.
- Extended memory profiling support in KotlinSyft, an open-source framework for training and inference of models on Android devices, to collect more detailed data including buffers and cache, swap space, active memory, writeback, mapped memory, shared memory, kernel memory, and allocation availability.
- Collected and analyzed memory and CPU traces of entry-level smartphones during model training and when idle, which provided insights into the optimization of FL processes for heterogeneous devices.
- Conducted testbed experiments to show the benefits of differential model serving in avoiding convergence delays by reducing model training time for slow clients.

GradAssist — Jul. 2023 – Present
Advisors: Dr. Ihsan Ayyub Qazi, Dr. Zafar Ayyub Qazi, Dr. Agha Ali Raza (LUMS)

- Developed an LLM-powered chat bot to provide personalized educational assistance to graduate students.

- Implemented streaming of chatbot responses using Vercel AI SDK and designed the chat API for performance by using Vercel's Edge Runtime so the functions can automatically execute in the region nearest to the user who triggers them.
- Implemented an SMTP server with Google OAuth2 for secure OTP delivery and integrated JWT for authentication.
- Engineered sophisticated prompting techniques to improve instruction following.
- Implemented a storage system using PostgreSQL for persistent storage and a Redis KV cache to improve response times.
- Conducted extensive research and benchmarking on vector databases and set up a scalable API for semantic retrieval using Milvus, an open-source vector database.
- Optimized retrieval to improve RAG performance through multi-vector querying and meta-data filtering.

Image-to-Speech Software for Urdu Language Jan 2022 – Dec 2022

Advisor: Dr. Agha Ali Raza (LUMS)

- Created a dataset of rendered and scanned images to train an OCR model for Urdu.
- Applied a transfer learning approach to train Google's Tesseract Engine for Urdu, achieving a notable reduction in Bag of Character Error Rate (BCER) to 2.94%, despite language resource limitations.
- Designed and implemented a pre-ocr-processing module to handle necessary image pre-processing with a focus on keystone correction, upscaling / super-resolution, and noise removal.
- Designed and implemented a post-ocr-correction module using Google's BERT and a custom masking engine, leveraging Conditional Random Fields (CRFs) for word segmentation followed by lexical lookup.

ConvoLens Sep. 2023 – Oct. 2023

Advisors: Dr. Ihsan Ayyub Qazi, Dr. Zafar Ayyub Qazi, Dr. Agha Ali Raza (LUMS)

- Developed an innovative application for filtering and analyzing call center conversations.
- Created a realistic, synthetic call center conversation dataset using GPT4 and Bark TTS model for demonstrating the app's capabilities.
- Implemented text-based semantic audio search by transcribing audios using OpenAI Whisper, embedding transcriptions with text-embedding-ada-002, and indexing in pgvector vector database, enabling precise query matching based on cosine similarity metric.
- Implemented AI-driven audio tagging using LLMs post-transcription, enabling detailed agent performance evaluation by quantifying interaction quality, customer satisfaction levels, and identifying instances of unprofessional behavior.
- Designed and developed a user-friendly interface for the app using Streamlit.

Teaching Experience

Undergraduate Teaching Assistant Jan. 2023 – May 2023

CS-382: Network-Centric Computing | Instructor: Dr. Zafar Ayyub Qazi *LUMS*

- Assisted in creating, organizing, and grading assignments and quizzes for 110+ students.
- Conducted tutorials and held weekly office hours.
- Utilized Moss for plagiarism detection in code submissions.

Undergraduate Teaching Assistant Jan. 2022 – May 2022

CS-382: Network-Centric Computing | Instructor: Dr. Zafar Ayyub Qazi *LUMS*

- Assisted in creating, organizing, and grading assignments and quizzes for 160+ students.
- Conducted tutorials and held weekly office hours.

Course Projects

Distributed Key-Value Store based on Raft | *Go* Oct. 2022 – Nov. 2022

CS-582: Distributed Systems *LUMS*

- Fully implemented the RAFT distributed consensus algorithm for a key-value storage system.

Peer-To-Peer File Sharing System | *Python* May 2021

CS-382: Network-Centric Computing *LUMS*

- Implemented a fault-tolerant key-value storage system based on Consistent Hashing and Chord's finger tables.

Implementing a Reliable Transport Protocol | *Python* Feb 2021

CS-382: Network-Centric Computing *LUMS*

- Developed a chat application that ensures reliable communication of messages and files by implementing a reliable protocol based on TCP on top of UDP.

Fraudulent Job Prediction | *Python, Scikit-learn, Pandas, Seaborn* Nov. 2022 – Dec. 2022

CS-334: Data Science *LUMS*

- Conducted comprehensive data cleaning and exploratory analysis, identifying key markers to detect fraudulent job offers.
- Trained a classifier model to predict real vs. fake job postings, employing NLP techniques for textual analysis.
- Findings were compiled in a blog post on medium.

Virtual Background | *PyTorch, OpenCV* Feb. 2022 – Mar. 2022
CS-437: Deep Learning *LUMS*

- Implemented background replacement and blur in videos using deep convolutional autoencoders.

Academic Management System | *Node.js, React.js, MongoDB, Git, Postman* Feb. 2022 – Apr. 2022
CS-360: Software Engineering *LUMS*

- Created a platform for schools to manage online education during the pandemic.

FoodSwings | *Node.js, React.js, MySQL, Git* Sept. 2021 – Dec. 2021
CS-340: Databases *LUMS*

- Developed a three-tier food ordering web application.

Technical Skills

Languages:	C, C++, Python, Javascript, Go, Haskell, Solidity, Kotlin, HTML, CSS, SQL
Frameworks/Libraries:	Next.js, Node.js, React.js, FastAPI, PostgreSQL, MongoDB, Milvus, Streamlit, PyGrid, PySyft, KotlinSyft, NumPy, Pandas, Matplotlib, Seaborn, PyTorch, Keras, Tensorflow, Scikit-learn, OpenCV, PyTesseract, PyCrypto
Developer Tools:	Git, Bash, AWS, adb, fastboot, Docker, LXC, CRIU, Jupyter, Wireshark, Postman, Stata