

Mohammad Shaharyar Ahsan

Lahore, Pakistan ★ +92 322 4018296 ★ shaharyarahsan1@gmail.com ★ [Linkedin](#) ★ [Github](#)

I am pursuing a Computer Science degree at LUMS, and am passionate about data science, machine learning and NLP, particularly in the fields of computational hermeneutics and linguistic analysis for social impact.

Education

Lahore University of Management Sciences (LUMS) – B.S. Computer Science Sept 2021 - June 2025

- ★ GPA: 3.80/4.0 , Major GPA: 3.89/4.0
- ★ **Relevant Coursework:** Machine Learning, Intro to Artificial Intelligence, Principles and Techniques of Data Science, Probability, Topics in Large Language Models, Internet of Things, Tech and Text: Digital Approaches to the Humanities, Databases, Internet Governance and Technology Policy

Awards & Honors

- ★ **Dean's Honor List**, Academic Year 2023-2024
- ★ **Dean's Honor List**, Academic Year 2022-2023
- ★ **Merit Scholarship** - O and A Levels, Academic Year 2021-2022 (LUMS)
- ★ **Outstanding Cambridge Learner Award** in O Level English Literature, 2019

Research Experience

Research Assistant – Embedded Systems & Networks Labs (SysNet/EmNets), LUMS June 2024 - May 2025

Advisors: Dr Naveed Anwar Bhatti (Assistant Professor), Dr Hamad Alizai (Associate Professor)

Worked on the Guardian Angel project, which involved using IoT data and generative AI to develop a smartwatch which can help democratize healthcare and wellness beyond economic and linguistic obstacles.

- ★ Implemented a small-scale prototype using an ESP-32 and off the shelf sensors for a course project. Used the Telegram Web API and Gemini's free plan to build the chatbot, which earned me the highest percentage in the course
- ★ Rebuilt the project in the EmNets lab, designing the tech stack with the research goals in mind of creating an affordable and intuitive solution
- ★ Tracked the project budget and led a team of three, coordinating with the research supervisors to keep them abreast of our progress
- ★ Helped write C code and determine the sensors to be used in the final PCB version of the watch. Learned about the GATT Bluetooth protocols and used them to transmit data using a BLE module to save energy
- ★ Sourced light libraries for use on a space-constrained MCU, and most notably fixing a critical issue with the oximeter values not being read correctly by replacing a library
- ★ Sourced a BLE library on GitHub and wrote an Android app in Java for data collection, intermediate processing (structuring into JSON etc.) and sending to the inference server
- ★ Wrote and deployed an inference server in Node.js, using OpenAI and Langchain JS packages, using in-context learning mechanisms like RAG and chain of thought prompting to improve results. Built function-calling capabilities into the server by adding tools and cron-jobs and iterative engineered prompts to obtain the best possible results.
- ★ Used the Meta Developer platform and Meta API to integrate the server with WhatsApp
- ★ Created a vector-store in Pinecone, sourced documents and set up a Python notebook for vectorizing and storing them using OpenAI's embedding model
- ★ Conducted extensive testing of the system to make improvements. Sourced datasets with the same sensor values and wrote code which used them to simulate real-time traces which were used to test the system for accuracy. Wrote python code to create visualizations and calculate measures like the mean absolute percentage error.

Research Intern – Center for Speech & Language Technologies, LUMS Jan 2024 - May 2024

Advisors: Dr Agha Ali Raza (Assistant Professor)

Worked on the Khpal Tabib project, backed by funding from the Gates Foundation. The project involved creating an AI chatbot for Pashto speakers to allow them to have accessible healthcare where the penetration of clinics and healthcare professionals is low in their native language

- ★ Conducted a thorough literature review on biomedical NLP research, identifying articles which aligned with the research goals of the project and outlining how they did so. Focused on biomedical prompt engineering and alignment.
- ★ Wrote a python script which allowed me to use a Reddit API wrapper to scrape the moderated AskDocs subreddit for responses by medical professionals and create a dataset with the alpaca prompt format, using regex methods for cleaning and preprocessing the text.

- ★ Trained a Gemma model using a pooled dataset with Unsloth, saving the LoRa adapters and model weights for reproducibility.
- ★ Suggested possible evaluation metrics sourced from state-of-the-art industry methods and the latest research

Teaching Experience

Teaching Assistant – Syed Babar Ali School of Science and Engineering, LUMS

Sept 2023 - Dec 2023

Course: CS 100 – Computational Problem Solving

- ★ Assisted in creating weekly lab tasks which tested concepts taught in the lectures for that week, as well as the marking schemes for that lab.
- ★ Invigilated labs to provide feedback to struggling students, as well as checked the lab tasks either in-lab with feedback or after the fact. Held contestations to allow students to discuss their issues and understand mistakes.
- ★ Invigilated graded components such as the midterm, final and quizzes held throughout the semester.
- ★ Was the lead for a subset of students for their final project. Helped them finalize their ideas, provided guidance for design and implementation and graded their projects.
- ★ Held office hours weekly to provide struggling students with feedback.

Teaching Assistant – Mushtaq Ahmad Gurmani School of Humanities and Social Sciences, LUMS

Sept 2023 - Dec 2023

Course: SS 100 – Writing and Communication

- ★ Marked attendance and class participation in weekly sessions (2x lectures a week)
- ★ Assisted in grading components of the course, such as presentations, quizzes and writing assignments
- ★ Held office hours weekly to provide struggling students with feedback and allow contestations.
- ★ Taught three sessions on conducting academic research and formatting citations, as well as separate seminars for individualized feedback to help students narrow their chosen research question for the final essay

Teaching Assistant – Suleman Dawood School of Business, LUMS

Sept 2024 - Dec 2024

Course: MGMT 348 – Internet Governance and Technology Policy

- ★ Marked qualitative CP in weekly sessions (2x weekly), as well as offered feedback on how to improve for struggling students.
- ★ Helped create, invigilate and check quizzes, exams and in-class writing assignments over the course of the semester, offering students constructive criticism in contestations. This included limiting the use of AI by negatively marking and offering feedback.
- ★ Helped manage the logistics of presentations, such as scheduling conflicts and auditorium bookings for block classes.
- ★ Held office hours regularly to offer struggling students feedback and help them improve.
- ★ Assisted the professor by using a database join system to speed up data entry on the LUMS portal for the final grading.

Academic Projects

Hook, Time and Singer - An Analysis of Songwriting as Cultural Memory

Sept 2024 - Dec 2024

Course: Tech and Text: Digital Approaches to the Humanities

- ★ Used data science and machine learning methods to conduct a sentiment and linguistic analysis of song lyrics from the past 7 decades (1960s-2020s)
- ★ Built a web-scraper using Python, specifically the BeautifulSoup library, to first create a dataset of GRAMMY-nominated songs from the past 7 decades. This dataset, along with the LyricsGenius library and the Spotify Web API, was used to generate the corpus for this research, including the lyrics and selected track features provided by Spotify.
- ★ Used the NLTK, WordCloud and HuggingFace transformers libraries to conduct sentiment and thematic analysis of the songs in quantitative fashion
- ★ Made arguments to support my thesis regarding the sentiment of songs and linguistic structure using existing literature and a qualitative close reading of selected lyrics, submitting a 20-page final paper
- ★ Currently working on preparing this project for publication in a journal of digital humanities/cultural analytics

Portal-LLM: Character-based Vicarious Learning through Chatbots

Sept 2024 - Dec 2024

Course: Topics in LLMs: Systems, Applications & Impacts

- ★ Identified a lack of interest in traditional reading practices in the education vertical and used existing literature to identify why students are uninterested in long-form narratives
- ★ Designed and built a project which aimed to transfer the useful aspects of vicarious learning through the experiences of characters in narrative form to text-messaging using LLMs to appeal to younger audiences using the Telegram API
- ★ Built a RAG pipeline which automatically vectorized documents sent in the chat, using Pinecone as the vector-store, as well as a web retriever which added web results
- ★ Built a LLM QA pipeline which classified prompts into 'code-switching' to allow changing characters in the chat and 'question-answering', which drew from the vector store and used an engineered prompt to answer as those characters
- ★ Tested the performance of multiple models from the Google family, including Gemini 1.5 Flash and 2.0 Experimental, as well as LearnLM 1.5, a fine-tuned model for educational tasks.
- ★ Tested the performance of two embedding styles offered by Google to see which would perform best for our use-case.
- ★ Tested the accuracy of the results using a Harry Potter trivia dataset and manually annotating results, achieving an accuracy of 83% in the best case

Tabular Regression on the Blue Book for Bulldozers dataset

Sept 2023 - Dec 2023

Course: Machine Learning

- ★ Used the Blue Book for Bulldozers dataset on Kaggle to perform 5 experiments using various ML models
- ★ Extensively pre-processed the data, exploring mechanisms for filling in null data, standardizing representations, using encoding mechanisms to represent categorical data etc.
- ★ Compared multiple methods for feature extraction, such as Mutual Information Gain and correlation matrices
- ★ Trained and compared multiple models, including Multiple Regressions, K-Neighbors regressor, Neural Network, SVM and Random Forest. Used standard metrics for error such as R^2 and RMSE.

Heart Disease Classification

Jan 2023 - May 2023

Course: Principles and Techniques of Data Science

- ★ Used a heart disease dataset on Kaggle to conduct a data science project related to EDA and model training
- ★ Cleaned the dataset by replacing or dropping missing data and generated encodings for categorical labels
- ★ Used box-and-whisker plots and histograms to analyze the data before and after cleaning to iteratively judge the spread of data and the effectiveness of the cleaning process
- ★ Plotted correlations to identify useful features using density plots, verifying our choice of features by using bootstrap sampling
- ★ Trained multiple models (SVM, Logistic Regression, KNN, Decision Tree, Gradient Boosting and CatBoost) and plotted the confusion matrix for each to see which would perform best, achieving 84%+ accuracy for all, with the best one being 90%.
- ★ Wrote and published a blog post detailing the process on [Medium](#).

On Geoblocking

Sept 2023 - Dec 2023

Course: Internet Governance and Technology Policy

- ★ Acting as a fictional NGO, our group created hypothetical policy guidelines to prevent geo-blocking as a mechanism for power retention for our final presentation.
- ★ Conducted extensive research on how geo-blocking manifests in third-world countries, particularly in terms of healthcare and content creation, and related existing policy measures.
- ★ Suggested policy measures which can be used to allow greater distribution of useful features globally, advocating for a more open and inclusive internet by citing how walled apps for diabetes monitoring, for instance, can be harmful
- ★ Included a budgeting plan for implementing suggesting policies
- ★ Conducted and presented a short case-study on Spotify, exploring how geo-blocked content and features can have implications for the market

Zaamin: A Secure and Compliant Data Portal

Jan 2024 - May 2024

Course: Software Engineering

- ★ Designed and hosted an HR portal with a strong emphasis on data security and compliance, externally evaluated by a software house (Devsinc).
- ★ Maintained extensive project documentation, including an SRS, SDS, and a GitHub project board

- ★ Designed the application on Figma, implemented it using React JS, HTML and CSS and deployed both the front-end and back-end on DigitalOcean and Vercel.
- ★ Implemented optimizations such as robust encryption methods (AES-256 and AES-CBC) for data protection.
- ★ Implemented audit logging and role-based access control (RBAC), as well as two-factor authentication using external APIs like the Hunter API, Visual Crossing and EmailJS.
- ★ Followed the AGILE methodology throughout the project, alternating roles as SCRUM master, frontend/backend developer, and designer.

Miscellaneous Projects

- ★ **Learning Management System:** Designed and implemented a learning management system using MongoDB with a ReactJS interface
- ★ **Auction App:** Created a live auction web app using the MERN Stack, using sockets for communication
- ★ **UNIX Shell:** Wrote a custom implementation of a UNIX Shell with a command line interface in C
- ★ **File System Implementation:** Wrote a custom implementation of a file system in C
- ★ **Compliment Bot:** Created a telegram bot which uses Gemini to respond to compliments on uploaded images
- ★ **Nonsense Outros Generator:** Compiled a dataset of Sabrina Carpenter's nonsense outros and fine-tuned a Gemma model to generate them based on an input city
- ★ **Open Source Bug Fix:** Closed an issue on GitHub where Nearby Share stopped working after the Quick Share merger on LineageOS by finding a workaround

Skills & Toolkit

- ★ **Research Methods:** Qualitative and Quantitative Research, Literature Reviews, Dataset Generation, System Design
- ★ **Data Analysis Tools:** SQL, MongoDB, Pandas, Matplotlib, NLTK, Sci-Kit Learn, HuggingFace, MATLAB
- ★ **Design and Prototyping:** Figma, LucidChart
- ★ **Proficient Languages:** Flutter, MERN Stack, Python, R, Java, JavaScript and Typescript, C, C++, HTML, CSS, LaTeX
- ★ **Project Management:** GitHub, Git (version control), Agile methodologies
- ★ **Soft Skills:** Presentations, Leadership, Teaching, Team management

Leadership & Extracurriculars

Member - Feminism Society (FemSoc) at LUMS

Sept 2021 - May 2022

- ★ Engaged in social media marketing and physical campaigns to
- ★ Advertised, helped organize and attended panel talks on pertinent issues
- ★ Actively participated in reading circles to learn and promote feminist ideas and literature.

Member - Drama Society (Dramaline) at LUMS

Sept 2021 - May 2022

- ★ Collaborated with the team to design and caption social media posts to attract a larger audience
- ★ Helped organize the 'Skit Tamasha' event by calling and networking with potential sponsors, as well as promoting the event on social media
- ★ Helped organize stalls to promote the annual play and film promotional material at them.

Volunteer - Lahore Literary Festival 2020

March 2020

- ★ Assisted in coordinating sessions and facilitating communication between speakers and attendees.
- ★ Managed on-ground logistics to ensure a smooth event experience for participants.
- ★ Was invited by the management to audit sessions on journalism, publication, arts and culture in Pakistan

Magazine Editor - Editorial Society at LGS JT (A Levels)

Jan 2020 - May 2020

- ★ Edited and curated articles, poetry, and artwork for the school's annual magazine.
- ★ Used tools like Canva, Photoshop and Adobe InDesign to design the magazine layout
- ★ Led a team of writers to develop and source content, maintaining high editorial standards.