

# Sheikh Hamza Elahi Sodana

520-437-3298 | [sodana@arizona.edu](mailto:sodana@arizona.edu) | Tucson, Arizona 85719 | [GitHub](#) | [LinkedIn](#)

## EDUCATION

---

### University of Arizona

*PhD in Computer Science (Incoming)*

Tucson, AZ, USA

*Starting Aug. 2026*

### Lahore University of Management Sciences (LUMS)

*Bachelor of Science in Computer Science; CGPA: 3.80*

Lahore, Pakistan

*Aug. 2022 – May 2025*

## AWARDS

---

- Placed on Dean's Honor List for 2021-2022, 2022-2023, 2023-2024, 2024-2025
- Graduated with High Distinction

## RESEARCH EXPERIENCE

---

### Research Assistant - Lahore University of Management Sciences

June 2025 – August 2025

*Advisor: Dr. Zafar Ayyub Qazi*

- Performed literature reviews, data preprocessing, and experimental result analysis

### BGP Data Analysis and Internet Shutdowns

May 2024

*Advisors: Dr. Zafar Ayyub Qazi, Dr. Zartash Afzal Uzmi, Dr. Qasim Lone, Dr. James Cowie*

- Investigated BGP level route changes in countries to understand the impact of political events on country-wide network connectivity
- Analyzed BGP announcements, withdrawals, path inflation, routing anomalies and AS Graphs crafted from the BGP State of Autonomous Systems
- Explored the role of physical and logical centralization as well as state-owned ASes on the level of disconnection in the country
- Observed different network infrastructures and how intra-network communication retained in spite of expected country-wide shutdown
- Performed a comparative analysis across countries with varying centralization and hegemony
- I plan to submit out findings to IMC '26

### Efficacy of Large Language Models on PDF Data Extraction

October 2023

*Advisors: Dr. Zafar Ayyub Qazi, Dr. Ihsan Ayyub Qazi*

- Evaluated GPT-4 and other LLMs for extracting structured data from Pakistan Telecommunication Authority's (PTA) QoS unstructured datasets
- Designed and implemented a JSON-based data extraction pipeline, enhancing accuracy to 96.8%
- Developed automation tools using Selenium for PDF processing and QGIS for geo-referencing telecom data
- Conducted ablation studies on prompting strategies, achieving significant improvements with few-shot and chain-of-thought approaches
- Developing an application by integrating scraped broadband data, cell tower locations and PTA data to provide comprehensive telecom insights

### Mutation Localization in Proteins Using Learned Representations

August 2025

*Advisor: Dr. Claire McWhite*

- Designed a machine learning framework to localize unseen point mutations in protein sequences under realistic experimental constraints
- Utilized transformer-based Protein Language Models (PLMs) to extract learned sequence representations for mutation localization
- Adopted a two-stage training strategy where a model trained with access to both wild-type and mutated sequences guides a second model trained only on mutated sequences
- Evaluated localization accuracy using distance-based metrics derived from known protein structures

### Investigating Donation Scams on Twitter

December 2024

*Advisors: Dr. Fareed Zaffar, Dr. Afsah Anwar, Dr. Bhupendra Acharya*

- Analyzing cases where scammers impersonate authentic Twitter users posting about charitable causes to divert donations to fraudulent PayPal addresses

- Scraped archive.ph for snapshots of reported scam profiles, collecting data on impersonation tactics and scam patterns
- Planning to report findings to Twitter, highlighting trends in fraudulent donation schemes and their impact on users
- Developing a honeypot experiment to understand scammer behavior, including how they identify and target posts for impersonation

## TEACHING EXPERIENCE

---

**Teaching Assistant (CS-200: Introduction to Advanced Programming)** Jan. 2024 – Dec. 2024

*Lahore University of Management Sciences*

- Created and graded quizzes and assignments for over 250 students
- Invigilated exams, organized labs, held tutorials, and office hours

**FutureTech Program ([website](#))** Aug. 2024

*Lahore University of Management Sciences*

- Guided high school students through interactive modules on Internet fundamentals, focusing on routing, data reliability, congestion control, and Internet security

**Teaching Internship** July 2019 – Aug. 2019

*Lahore Literati Montessori and High School*

- Assisted in teaching various subjects to elementary students

**Robowars - PSIFI Event ([website](#))** Sept. 2022 – Oct. 2022

*Lahore University of Management Sciences*

- Guided students in building RC cars from scratch
- Provided parts and demonstrations

## PROJECTS

---

**Distributed Key-Value Store (RAFT) | *Go, Raft Consensus, gRPC* | [GitHub](#)**

- Implemented a distributed fault-tolerant key-value store using the Raft consensus algorithm for leader election and log replication.
- Designed a persistent WAL (write-ahead log) and snapshotting mechanism enabling crash recovery.

**Chat-Server-Summarizer | *Python, FastAPI, SQLAlchemy, PostgreSQL, OpenAI API* | [GitHub](#)**

- Developed a backend service that ingests messages in real time from Discord, Slack, and Telegram, storing them in a PostgreSQL database for unified archiving and analysis.
- Implemented on-demand summarization via a Discord slash command using GPT-4, generating per-platform, per-channel summaries through FastAPI and asynchronous pipelines.

**ScholarRAG | *Python, LangChain, FAISS, OpenAI API, Wikipedia* | [GitHub](#)**

- Built a hybrid Retrieval-Augmented Generation (RAG) system that answers research queries using a FAISS-indexed vector store of academic papers with optional live Wikipedia retrieval.
- Implemented document ingestion, embedding, and context-assembly pipeline via LangChain with citation tracking and a REST API for research Q&A, achieving 99% retrieval accuracy on benchmark queries.

## TECHNICAL SKILLS

---

**Programming Languages:** Python, JavaScript, TypeScript, C, C++, Go, Haskell

**Machine Learning & Data Analysis:** Scikit-Learn, Hugging Face, LangChain, NumPy, Matplotlib, Pandas, MySQL, MATLAB

**Network Tools:** RIPE Atlas, RIPE NCC, BGPlay, Wireshark, BGPstream

**Tools:** MS Office Suite, Git, LaTeX, Selenium, Qiskit

## RELEVANT COURSEWORK

---

**Undergraduate-Level Courses:** Algorithms, Data Structures, Databases, Software Engineering, Operating Systems, Network-Centric Computing, Network Security, Introduction to Quantum Computing

**Graduate-Level Courses:** Distributed Systems, Topics in Internet Research, Topics in Large Language Models, Topics in Computers and Network Security