Syed Moiz Ali

✉ 25100149@lums.edu.pk   in moizali01   in moizali01

EDUCATION

Lahore University of Management Sciences – B.S Computer Science *2021-2025*

- **Relevant coursework:** Topics in Computer and Network Security, Deep Learning, Machine Learning, Network Security, Topics in Large Language Models, Computer Vision, Probability, Algorithms.

RESEARCH EXPERIENCE

LLM-Integrated Source Code Debloating *Mar 2024 - Present*

Security & Privacy Lab, LUMS *Lahore, Pakistan*

Collaborators: Dr. Fareed Zaffar (LUMS), Dr. Ashish Gehani (SRI International), Dr. Sazzadur Rahaman (University of Arizona), Dr. Fahad Shaon (Google)

- Developed a novel **RAG** based multiagent LLM pipeline to assist traditional **debloaters** by retaining code critical for functionality and generality in debloated software.
- Designed specialized LLM prompts informed by manual analysis to address limitations in current debloaters, enhancing relevance of retained code.
- Leveraged **LLVM coverage** and **code semantics** to introduce stability-focused heuristics, reducing critical functionality loss while maintaining program security.
- Achieved improved generality and stability across benchmarks when integrated with three existing debloaters, with minimal size impact.

LLM Safety Alignment *Feb 2024 – Sep 2024*

Security & Privacy Lab, LUMS *Lahore, Pakistan*

Collaborators: Dr. Fareed Zaffar (LUMS), Dr. Yasir Zaki (NYU Abu Dhabi)

- Investigated task-specific safety degradation in fine-tuned LLMs, identifying vulnerabilities related to downstream task.
- Comprehensively analysed the shortcomings of existing safety solutions, including fine-tuning datasets, and external guard and moderation models.
- Created and curated a **multitask safety dataset** that enhances task-specific safety guardrails in fine-tuned models, ensuring comprehensive alignment across tasks.
- Fine-tuned models using the safety dataset, showing significant improvements in security, with findings under review at COLING 2025.

LLM Hallucination Benchmarking *Oct 2024 – Present*

LUMS *Lahore, Pakistan*

- Designing a probabilistic framework to detect hallucinations in LLM outputs, analyzing log probabilities and top-k token distributions.
- Conducting initial benchmarks on datasets such as MMLU, MedMCQA, and ScienceMCQ to identify patterns distinguishing confident and hallucinated responses.
- Exploring threshold-based metrics to improve error identification, with preliminary findings guiding further refinement of the methodology.

Applied ML Intern *Jun – Aug 2023*

Center for Urban Informatics, Technology and Policy, LUMS *Lahore, Pakistan*

- Developed a high-performance web scraper using selenium and multithreading, extracting **electricity consumption** data of over **3 million** users across Lahore.
- Engineered an LSTM-based time series **forecasting** model to predict **feeder overloading**, aiming to improve grid management.
- Analyzed **seasonal consumption** patterns to identify **poverty hotspots**.

Syed Moiz Ali

✉25100149@lums.edu.pk in moizali01 in moizali01

Publications

Preprints

Multitask Mayhem: Unveiling and Mitigating Safety Gaps in LLMs Fine-tuning ***2024***
Essa Jan, Nouar AlDahoul, **Moiz Ali**, Faizan Ahmad, Fareed Zaffar, Yasir Zaki
ArXiV:2409.15361

Experience

Teaching Assistant *Sep 2023 - Present*
Lahore University of Management Sciences *Lahore, Pakistan*

- Computer Vision Fundamentals (Fall 2024, Dr. Murtaza Taj): Designed and evaluated programming assignments and quizzes in a graduate-level course.
- CS100: Computational Problem Solving (Spring 2023, Dr. Fareed Zaffar): Conducted tutorials, graded assignments for 90 students, and offered individual support.
- CS200: Introduction to Programming (Fall 2023, Dr. Shafay Shamail): Led tutorials, graded labs and quizzes for 80 students, and provided individual assistance.

PROJECTS

Tradesnap.ai | *MERN, Selenium, Azure Cloud, OpenAI* *Jan 2024 – May 2024*

- Developed a conversational stock trading platform using OpenAI's Assistant to enable multilingual stock trading via chat interface
- Integrated features like buying/selling stocks, educational content, and personalized volatility alerts
- Scraped data from PSX for platform backend and built detailed company pages with advanced React charts
- Implemented automated testing for the application using Selenium to ensure platform reliability

Nighttime Wildlife Monitoring | *CycleGAN, Image Processing, OpenAI CLIP* *Jan 2024 – May 2024*

- Developed a hierarchical model leveraging CycleGANs to enhance nighttime camera trap images for snow leopard detection
- Used OpenAI's CLIP for image classification and fine-tuned it for challenging nighttime conditions
- Collected and curated training data from the Snapshot Serengeti Database, achieving 0.95 accuracy and 0.89 F1-score

Social Media Toxicity Classifier | *Llama2, PEFT, Jigsaw Dataset* *Jan 2024 – May 2024*

- Developed a model to detect and flag harmful social media content, fine-tuning Llama2-7B (PEFT) for toxicity classification
- Achieved 90% accuracy and an F1-score of 0.89 across 6 toxic classes using the Jigsaw Toxic Comment Classification Dataset
- Reached a ROC of 0.85, ensuring effective detection of harmful content

Technical Skills

- **Language:** Python, JavaScript, C++, Haskell, HTML, CSS, Bash Scripting
- **Technologies/Frameworks:** PyTorch, TensorFlow, OpenCV, MERN, TypeScript, LLVM, LangChain, Pandas, Scikit-learn, LlamaIndex, OpenAI Platform, Google AI Studio, Selenium