Syed Moiz Ali

+1 872 (265) 8628 | moizsyedali.ma@gmail.com | linkedin.com/in/moizali01 | github.com/moizali01

Education

Ph.D. in Computer Science, University of Illinois at Chicago — Chicago, IL
Ph.D. in Computer Science — *Sep. 2025 – Present*

Lahore University of Management Sciences — Lahore, Pakistan
Bachelor of Science in Computer Science — *Sep. 2021 – May 2025*

Relevant Coursework: Topics in Computer and Network Security, Deep Learning, Machine Learning, Network Security

Research Experience

Graduate Researcher — Sep 2025 – Present
University of Illinois at Chicago — *Chicago, IL*

- Researching security and privacy vulnerabilities in LLM-based browsing agents, focusing on fingerprinting and privacy risks. Developing auditing frameworks to identify attack vectors, enabling developers to build more secure systems.

Research Assistant — Jan 2024 – May 2025
Security & Privacy Lab, LUMS — *Lahore, Pakistan*

- Collaborated with **SRI International**, **University of Arizona**, and **Google** to develop a novel RAG-based multiagent LLM pipeline that assists traditional debloaters by retaining code critical for security and functionality. Achieved improved generality and stability across benchmarks when integrated with existing debloaters.
- Manually created a ground truth dataset of **12 coreutil programs**, moving away from traditional test case-based evaluation. Conducted manual code review and security auditing of existing debloater outputs against the ground truth dataset, discovering **8 unique vulnerabilities** in the debloated code that had not been reported earlier.
- Evaluated knowledge internalization across **5 LLM architectures** and **4 model sizes** (1.5B-72B parameters), analyzing how scale and task complexity affect generalization. Found that comprehension tasks (QA) retain knowledge **2.8× more effectively** than mapping tasks (Translation), with scaling improving retention up to **72%** (**EMNLP Findings**).
- Investigated task-specific safety degradation and exploit vectors in finetuning of LLMs, uncovering security vulnerabilities in tasks such as code generation, translation, and classification. Developed **MultitaskBench**, a cross-task safety alignment dataset of **2,020** prompts, reducing attack success rates by up to **95%** (**COLING 2025**).

Publications

MultitaskBench: Unveiling and Mitigating Safety Gaps in LLMs Fine-tuning | *COLING* — 2025
Essa Jan, Nouar AlDahoul, **Moiz Ali**, Faizan Ahmad, Fareed Zaffar, Yasir Zaki.

Data Doping or True Intelligence? Evaluating the Transferability of Injected Knowledge in LLMs | *EMNLP Findings* — 2025
Essa Jan, **Moiz Ali**, Saram Hassan, Fareed Zaffar, Yasir Zaki.

Industry Experience

Generative AI Engineer — June 2025 – Aug 2025
Technology for People Initiative (TPI), LUMS — *Lahore, Pakistan*

- Developed a LangChain-based multiagent AI tax lawyer with automated document parsing and response generation, reducing manual processing overhead and improving compliance accuracy across regulatory workflows.
- Designed scalable refine chain pipelines with complexity-based routing, enabling differentiated processing paths for consumer vs. enterprise tax scenarios and significantly reducing manual review requirements.

Projects

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

- Developed a conversational stock trading platform using OpenAI's Assistant to enable stock trading via chat.
- 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.

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

- Developed a hierarchical model using **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.

Urban Electricity Analytics | *Selenium, LSTM, Python, Pandas* Jun 2023 – Aug 2023

- Developed a **Selenium** based web scraper to extract electricity consumption data for over **3 million** users across Lahore.
- Engineered an **LSTM**-based time series forecasting model to predict feeder overloading, improving grid management strategies. Conducted analysis of seasonal consumption patterns to identify **poverty hotspots**.

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 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

Languages: Python, JavaScript, C, C++, Haskell, HTML, CSS, Bash
Technologies/Frameworks: PyTorch, TensorFlow, OpenCV, MERN, TypeScript, LLVM, LangChain, Pandas, Scikit-learn, LlamaIndex, OpenAI Platform, Google AI Studio, Selenium, Azure Cloud, Ghidra, GDB, Valgrind