Syed Moiz Ali

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

Education

University of Illinois Chicago — Chicago, IL
Ph.D. in Computer Science; Advisor: Prof. Jason Polakis — *Aug. 2025 – Present*

Lahore University of Management Sciences (LUMS) — Lahore, Pakistan
B.S. in Computer Science — *Sep. 2021 – May 2025*

Publications (* equal contribution)

AgentShield: Fortifying Web Applications Against AI Agents and Agentic Browsers *15.8% acceptance rate*
Moiz Ali*, Nafis Karim*, Rownak Islam, Jyotirmay Chauhan, Dhruv Kuchhal, Jason Polakis — **IEEE S&P 2027**

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

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

Experience

Graduate Researcher — Aug. 2025 – Present
University of Illinois Chicago — *Chicago, IL*

- Co-led **AgentShield** (IEEE S&P 2027): tested **14 agents** (e.g., Comet, Atlas, Claude for Chrome) against anti-bot defenses (reCAPTCHA v3, Turnstile, JA4, RFC 9421); local agents evaded network- and environment-level checks.
- Built a **behavioral fingerprinting** system that uses dynamic forms to measure agent reasoning latency, keystroke, and pointer telemetry; a 12-feature classifier flagged every held-out agent, including ones told to mimic humans.
- Discovered **Extinction-induced Policy Drift**, an attack that makes 5 of 6 agents override their own safeguards; released AgentShield components as **npm packages** that sites can add without code changes.

Generative AI Engineer — Jun. 2025 – Aug. 2025
Technology for People Initiative, LUMS — *Lahore, Pakistan*

- Built a **LangChain** multi-agent "AI tax lawyer" that parses tax notices and drafts compliance responses.
- Added complexity-based routing and a **self-correction loop** that cut hallucinations in summaries by **40%**.

Research Assistant — Jan. 2024 – May 2025
Security & Privacy Lab, LUMS (with SRI International, University of Arizona, Google) — *Lahore, Pakistan*

- Built a **RAG-based multi-agent LLM pipeline** that helps software debloaters keep security- and functionality-critical code; hand-built a ground-truth set of 12 coreutils and found **8 new vulnerabilities** in existing debloater output.
- Built **MultitaskBench**, a 2,020-prompt dataset that cut fine-tuning attack success by up to **95%** (COLING 2025).
- Found QA fine-tuning retains injected knowledge **2.8**× better than translation across 5 LLMs (EMNLP Findings 2025).

Projects

Tradesnap.ai | *MongoDB, Express, React, Node.js, Azure, OpenAI API* — Jan. 2024 – May 2024

- Built a chat-based stock trading platform with buy/sell, volatility alerts, and company pages from scraped PSX data.

Social Media Toxicity Classifier | *Llama 2, PEFT, PyTorch* — Jan. 2024 – May 2024

- Fine-tuned Llama 2-7B for 6-class toxicity detection on the Jigsaw dataset (**90%** accuracy, F1 **0.89**).

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

- Scraped consumption data for **3M+** users in Lahore and trained an LSTM to forecast feeder overloading.

Technical Skills

Languages: Python, JavaScript, TypeScript, C, C++, Bash, HTML/CSS, Haskell
ML & LLMs: PyTorch, TensorFlow, scikit-learn, Pandas, OpenCV, LangChain, LlamaIndex, OpenAI API, Ollama
Web & Cloud: Node.js, Express, React, MongoDB, Playwright/Puppeteer, Selenium, Docker, nginx, Azure, Git
Security: reCAPTCHA v3, FingerprintJS, Cloudflare Turnstile & AI Crawl Control, JA4, RFC 9421, Ghidra, GDB, Valgrind, LLVM