Shaoor Munir

530-953-3757 shaoormunir@gmail.com linkedin.com/in/shaoormunir www.shaoormunir.com

EDUCATION

University of California Davis **Sep. 2021 – Dec. 2025 (Tentative)**
Ph.D., Computer Science, CGPA: 3.81/4.0 *Davis, CA*

FAST-NUCES **Aug. 2014 – May 2018**
B.S., Computer Science, CGPA: 3.73/4.0 *Lahore, Pakistan*

RESEARCH EXPERIENCE

University of California, Davis **Sep. 2021 – Present**
Graduate Student Researcher | Online Privacy and Machine Learning

- Developed machine learning pipelines to detect misuse of first-party tracking cookies, link decorations and intrusive JavaScript APIs with state-of-the-art accuracy.
- Designed experiment to detect evasive bots on the web and proposed algorithm to increase detection rates of commercial bot detection services.
- Built and maintained large-scale data collection and analysis pipelines using **OpenWPM**, **Puppeteer**, and **Playwright** to analyze web tracking at scale.

PROFESSIONAL EXPERIENCE

IDPrivacy, Inc. **Mar. 2024 – Present**
AI Research Scientist (Internship) | Privacy and Large Language Models

- Led the design and implementation of a self-serve agentic AI platform built using a custom orchestration layer on **GCP** with end-to-end encryption across microservices, achieving a 73% appointment booking rate and 70% conversion rate using insights across 10,000 real-world customer interactions.
- Engineered and implemented **RAG** (Retrieval Augmented Generation), **TAG** (Table Augmented Generation), and custom **ETL** pipelines using **Terraform** to ingest large scale data and enable deployment of AI agents in secure and private environments.
- Designed an automated wholesale vehicle valuation system, incorporating a pipeline with an **LSTM** in combination with a finetuned self-hosted **LLM** running on **Databricks** using **unsloth**, achieving 98% accuracy across 50,000+ vehicle models for a national wholesale valuation provider.

SELECTED RESEARCH

- **Using ML and Measurements to Protect User Privacy:** *[ACM CCS 2023, ACM CCS 2024, USENIX Security 2024]* Developed machine learning pipelines to crawl websites, model interaction of different elements on a webpage using a Graph representation, and classify tracking behavior using a sophisticated Random Forest Classifier, achieving state-of-the-art accuracy. Leveraged the pipeline to detect misuse of first-party cookies, link decorations, and mixing of JavaScript functions to violate user's privacy.
- **Combining Technical and Legal Analysis to Protect User Privacy** *[Vanderbilt JetLaw Vol. 27, USENIX Security 2026 (under submission)]* Used a combination of technical and legal analysis to propose remedies against interception of communication and collection of extensive user data by both malicious third-party script and first-party browsers.
- **Detecting Evasive Web Bots:** *[IMC 2025]* Developed a semi-automated algorithm to detect inconsistent fingerprints in evasive bots who actively change their fingerprints to evade detection by online bot detection services. Performed a large scale measurement to show that the technique can increase effectiveness of bot detection services by up to 50%.

SKILLS

- **Languages:** Python, SQL, C++, Java
- **Frameworks & Libraries:** PyTorch, Keras, Scikit-Learn, TensorFlow, SciPy, Unsloth, LangChain
- **Cloud & Infra:** Amazon Web Services (Athena, Glue), Google Cloud Platform (Cloud Run, Vertex AI,), DataBricks (DLTs, MLFlow, AgentBricks)
- **Web Scraping:** Selenium, Playwright, JavaScript, Scrapy, Puppeteer

ACHIEVEMENTS

- Research featured in media including Newsweek, The Register and ProMarket
- Appeared on DataSkeptic podcast to discuss future of advertising and privacy
- Gold Medalist (B.S. Computer Science) and Dean's List (Ph.D., UC Davis)