AVIDAN (AVI) SHAH

65 Dorison Drive, Short Hills NJ 07078 | amcshah@berkeley.edu | 973-902-3032

EDUCATION

University of California, Berkeley | May 2025 **Cumulative GPA:** 3.8

B.A. Computer Science & B.A. Applied Mathematics (Data Science Concentration)

Awards & Honors: UCB Dean's List, National Merit Finalist, National AP Scholar, Nationally Certified EMT (2019)

PROFESSIONAL / RESEARCH EXPERIENCE

Millennium Management June – August 2023, June – August 2024

Quantitative Research Intern *New York, NY*

- Implemented a generative adversarial network for unsupervised anomaly detection on market data
- Built an automated classifier to detect emails that would result in disruption of data delivery or PM trading
- Used natural language processing techniques on textual market data to generate and evaluate trading signals
- Performed data ingestion and assisted the maintenance of systematic data pipelines used daily by over 300 different investment teams

Berkeley Artificial Intelligence Research (REDS Group) September 2022 – Present

Undergraduate Researcher *Berkeley, CA*

- Conducting research as part of Professor David Wagner's group in both ML for security and security for ML, working with Julien Piet and Chawin Sitawarin
- Designed, built, and tested a model using transformer architecture for unsupervised and semisupervised anomaly detection on keystroke data over SSH connections
- Developing a framework to assessing efficacy of transfer attacks on black-box LLMs

MIT Data to AI Laboratory May – August 2022

Undergraduate Researcher *Cambridge, MA*

- Researched and developed ML pipelines for more accurate unsupervised anomaly detection in time series data using the lab's open-source Python libraries
- Built a fully automated, end-to-end workflow for continuously updating public data acquisition, model driven anomaly detection, and visualization via GitHub pages
- Joined the development team of the Signal Intelligence (Sintel) project's Orion anomaly detection library

PROJECTS / PUBLICATIONS

Deep Learning for SSH Traffic Anomaly Detection September 2022 – Present

Undergraduate Researcher *Berkeley, CA*

- Developed three different unsupervised learning models for time series data to detect potential network intruders using inter-keystroke timings in SSH
- Currently building a new model for a semi-supervised learning context combining keystroke data and server information in order to defend against timing attacks

Improving Universal and Transferable Jailbreaks: Targeting Safety Features Directly January 2023 – Present

Undergraduate Researcher *Berkeley, CA*

- Currently analyzing factors that correlate to transferability and universality of adversarial suffixes trained on white-box LLMs, as well as their applicablity to black-box models, with the goal of improving robustness
- Using sparse autoencoders to evaluate the potential for residual activations and safety feature vectors to be leveraged in an adversarial attack on a language model

Efficient Bus Bunching Mitigation through Adversarial Curriculum Learning December 2023

CS285 (Deep Reinforcement Learning) Final Project *Berkeley, CA*

- Developed a novel approach to curriculum learning utilizing adversarial model to increase bus system efficiency
- Adversarial curriculum setter model performs well without requiring extensive domain knowledge or training

SKILLS, PERSONAL INTERESTS

Skills: Python, Java, SQL, PyTorch, Pandas, Spanish (Limited)

Interests: Strategy Games, Piano, Swimming, Writing Flash Fiction, Emergency Medicine, Sigma Chi Fraternity