Advait Yadav

602-565-4271 | advaity2@illinois.edu | linkedin.com/in/advaityadav | github.com/advtydv

Education

University of Illinois Urbana-Champaign **Expected May 2027**
Bachelor of Science in Mathematics and Computer Science GPA: 3.93/4.00
Coursework: Data Structures, Algorithms, Low-level Programming, Operating Systems, Computer Architecture
Leadership: CS124 Honors Project Manager, Machine Learning Club

Skills

Languages: Python, C, C++, HTML, CSS, JavaScript, Go, SQL
Libraries and Frameworks: PyTorch, TensorFlow, Scikit-learn, React, WebGL, Three.js, Django, Flask
Developer Tools: AWS, Azure, Google Cloud, Git, Unix/Linux, Docker, Kubernetes, Terraform, CI/CD

Experience

Research Fellow, ML Alignment & Theory Scholars **June 2025 – Present**
- Mentored by Oliver Sourbut and Sid Black from UK AISI; Built and ran information-asymmetric simulations to discover cascading failures in multi-agent systems.
- Analyzed individual agent's Theory-of-Mind capabilities to detect and mitigate collusion, deceptive strategies, and coordination breakdowns.
- Designed early-warning metrics, a taxonomy of failure modes, and a pre-deployment evaluation protocol.

Researcher, DREAM group **August 2024 – Present**
- *InfoFlood*: demonstrated an information-overload jailbreak that exposes a core safety weakness in frontier black-box LLMs; arXiv preprint. Featured in 404 Media, POLITICO, IT Brew.
- Engineered automated red-teaming framework that generates and iteratively refines attack prompts, achieving up to 3× higher attack success rates vs. prior methods.
- Researched latent-space guardrails to prevent malicious editing in multi-modal models.

Software Developer, SHaDE Lab **November 2023 – August 2024**
- Led development of MaRTiny, a low-cost atmospheric and visual detection robot across 20+ deployment sites.
- Optimized YOLOv8 algorithm for low-quality images on edge devices, improving processing speed and detection accuracy on video segmentation and activity patterns.
- Enabled real-time data collection, delivering insights to the Government of Phoenix Urban Planning team.

Undergraduate Researcher, ASU Biodesign Institute **September 2023 – July 2024**
- Developed molecular diffusion model that improved the accuracy of protein trajectory predictions, enhancing the reliability of lab-grown molecule analysis.
- Optimized GPU simulations and parallelized cloud infrastructure by rewriting TensorFlow code in JAX, cutting compute time by 30% and throughput time by 25%.

Data Science Intern, KiwiTech **June 2022 – August 2022**
- Built machine learning models for MNC clients, improving forecast accuracy for revenue reports by 20%.
- Optimized data pipelines, improving system uptime from 95% to 99%+.

Projects

PrepU: Interview Analysis – Winner at SunHacks (Amazon)
- Led development of interview analysis system using AWS and React, processing 3,000+ mock interviews.
- Engineered scalable microservices architecture in Docker and Kubernetes to handle 5,000+ user requests utilizing CI/CD integration, delivering detailed candidate evaluations to employers.

Tax Return Search Engine
- Steered team to develop a search engine with sub-second response time across 1M+ records, extracting tax form data from 100,000+ IRS 990 filings, cutting search time by 70%.
- Tested by affiliated non-profits, enabling faster financial planning, and enhancing operational efficiency.