Asad Shahid

408-430-2623 | asad.shahid@berkeley.edu | linkedin.com/in/asadshahid04 | asadshahid.netlify.app

Education

University of California, Berkeley — Berkeley, CA

Bachelor of Arts, Statistics & Data Science — *Expected December 2026*

- **Awards:** 2022 CIF Scholar-Athlete, Avi Raina Scholarship, Las Positas Engineering Scholarship
- **Relevant Coursework:** Data Structures & Algorithms, Linear Algebra, Principles & Techniques of Data Science, Probability Theory and Algorithms

Experience

Hewlett Packard Enterprise — May 2025 – Aug. 2025

Software Engineer Intern — *San Jose, CA*

- Cut vulnerability triage time by 90% and reduced false positives by 50% by replacing slow, manual CVE workflows with a custom **Helm chart** deployment of NVIDIA's **Container Security Blueprint** on **Kubernetes**.
- Accelerated vulnerability scans from **hours to minutes**; self-hosted **Llama 3.1 8B NVIDIA Inference Microservice (NIM)** on **L40** and **A100 GPUs** to analyze 300+ Docker images.
- Improved analysis throughput by **3.2x** and cut API redundancy by **60%** by implementing a templated **NGINX caching layer** and orchestrating CVE, SBOM, and repo ingestion into **NVIDIA Morpheus SDK**.

University of California, Davis — Oct. 2024 – April 2025

Undergraduate AI Researcher — *Remote, CA*

- Improved understanding of **Llama 3.2 Vision Models** by successfully loading and executing the **11b model**, enhancing resource management insights.
- Researched quantization techniques, including **LLM.int8()**, **GPTQ**, and **SPQR**, to apply **Post-Training Quantization (PTQ)**, optimizing performance for large LLMs.
- Established benchmark performance metrics through comprehensive testing, collecting data on latency and accuracy to inform model optimization strategies.

Genentech — June 2024 – Aug. 2024

Software Engineer Intern — *South San Francisco, CA*

- Engineered a responsive front-end document revisioning platform that reduced page load time by **40%** through optimized **Vue.js** and **Pinia** state management implementation.
- Accelerated study startup processes by up to **1 week**; developing and integrating key features such as **comments**, **editing history**, and **notifications**, resulting in streamlined protocol authoring and review workflows.
- Improved clinical authors' productivity by **30%** - architected and built a full-stack **Retrieval-Augmented Generation (RAG)** application utilizing **BioBERT**, **MilvusDB**, **Vue.js**, and **FastAPI**, enabling historical protocol data retrieval in under **10 seconds**.

Projects

Postit — *github.com/AsadShahid04/postit*

Built a full-stack Twitter-like platform enabling users to create and oversee posts. Employed **Next.js**, **NextAuth**, **Google Authentication,** and **React** to build front end. Designed **REST APIs** for handling user posts. Leveraged **PostgreSQL** as the backend database using **Railway** and **Prisma (ORM)**. Led continuous integration/deployment using **Git**.

Detecting Financial Fraud Cases — *github.com/AsadShahid04/FinalParallelProcessingWebsite*

Developed and fine-tuned multiple machine learning models, including **Random Forest** and **Support Vector Machine**, using **Scikit-learn** to detect fraudulent transactions in a dataset of over 6 million bank transactions. Presented findings through a **Flask** and **Bootstrap**-based web application, leveraging **Pandas**, **Plotly**, and **Numpy** for data visualization and exploratory data analysis.

Skills & Interests

Skills: Python, C++, JavaScript, Cursor, HTML/CSS, SQL, Kubernetes, Git, Docker, Object-Oriented Programming

Interests: Tennis, Travelling, Aviation, Automotives