Smit Vasani

Fremont, CA | svasani513@gmail.com | +1 (331) 304-8752 | in/smit-vasani-1980p

Education

University of Illinois Urbana-Champaign, *MS in Computer Science* Sept 2025 – Dec 2026

- Coursework: Numerical Analysis, Distributed Systems, Natural Language Processing

University of Wisconsin-Madison, *BS in Computer Science & Data Science* Jan 2021 – May 2025

- GPA: 3.96/4.0
- Coursework: Machine Learning, Algorithms, Database Systems, Big Data Analytics, Software Engineering

Summary

I'm a Computer Science student with strong skills in data analysis, pipeline development, and database design. I enjoy building systems that support scalable data processing, statistical modeling, and visualization. I have experience with ETL workflows, cloud platforms, and tools like Apache Spark to optimize data reliability and accessibility.

Experience

Research Assistant, *University of Wisconsin-Madison* May 2024 – Present

- Fine-tuned LLMs using SFT and DPO to simulate multi-agent conversations and analyze opinion shifts; evaluated alignment with human responses using semantic similarity, Likert scores, ROUGE, and BLEU.
- Optimized data pipelines and parallelized model inference across multi-GPU systems, reducing inference time by **60%** and enabling evaluation of **2000+ agent conversations** at scale.

Software Engineering Intern, *Hewlett Packard* May 2024 – Aug 2024

- Designed and developed a web plugin for HP printers, enhancing webpage print quality with a three-layered content filtration system, achieving a **75% improvement** in relevant content extraction.
- Collaborated in an agile environment to develop a solution for webpage restructuring, reducing print sheet consumption by **33%** and decreasing callbacks to foundation models, **saving HP $50,000**.

Data Science Intern, *Hewlett Packard* June 2023 – August 2023

- Built predictive maintenance pipelines on Databricks using PySpark and XGBoost to identify early signs of printer hardware failure, and reducing downtime by over **500 hours/month**.
- Deployed ML models to cloud and integrated them into HP's global monitoring system, improving operational efficiency by **80%** and reducing service costs by **$20,000** through automated issue resolution.

Projects

Real-time Weather Data Processing with Kafka

- Built a scalable Python-based Kafka producer-consumer pipeline to ingest and process **20GB+** of weather data daily, enabling **real-time data streaming** and analytics.
- Implemented reliability features such as acknowledgment mechanisms, retries, and checkpoints, minimizing data loss to **<0.1%** across distributed systems.

eBay Data Analysis and Database Design

- Constructed a comprehensive schema for eBay's JSON dataset, enabling the conversion of over **15 million** records into SQLite files with streamlined data integrity and retrieval processes.
- Analyzed eBay's structured data to identify key trends, performing SQL-based aggregations and summaries that provided actionable **insights into user behaviors** and listing dynamics at scale.

Technologies

Languages: Python, C, C++, Java, R, HTML, CSS, JavaScript, SQL, Bash, Assembly

Applications/Tools: Git, Docker, Kafka, Cassandra, Hadoop, Spark, Hive, MongoDB, Elasticsearch, Snowflake