ANSHUMAN SHARMA

Sunnyvale, California | (949) 279-7105 | anshums3@uci.edu | Linkedin | Portfolio

EDUCATION

University of California Irvine | Master of Science in Data Science   Sep 2022- Dec 2023
Honor: Program Ambassador MDS

Amity University | B.Tech Computer Science and Engineering   Aug 2016- Jul 2020
Honor: Recipient of Dr. Ashok K Chauhan Merit Scholarship

SKILLS

Backend:	Python, Flask, FastAPI.
Frontend:	ReactJs, Javascript, HTML, CSS.
Data Pipelines:	Apache Airflow, Spark, Kafka, Data Warehousing (Redshift, Snowflake, SAP BW), Cassandra, DBT, Hadoop Grafana, Prometheus, ELK Stack (Elastic Seach, Logstash, Kibana), SQL.
Miscellaneous:	Docker, Kubernetes, Amazon AWS, Power BI, Looker, Git.

EXPERIENCE

Visceral d/b/a Natural Language Labs | Software Engineer | New York, NY   Jun 2024 - Present

- Implemented an LLM auto-parser using **Hermes Theta2** (LLaMA 3) on Modal, reducing survey building time by 95%.
- Developed an application for report generation across the platform, utilizing **FastAPI** to interface seamlessly with Hermes.
- Engineered automated plot generation to visualize underlying data for analysis and presentation.
- Designed and deployed a Retrieval-Augmented Generation (RAG) system using **LLaMA 3.1**, containerizing it with **Docker** for efficient deployment on client systems.
- Created a **Chroma DB** database to support the RAG system, incorporating annotations to improve response accuracy and context.
- Designed and developed consumer data access, deletion, and opt-out mechanisms, ensuring adherence to **CCPA** and **GDPR** requirements.
- Established a **PostgreSQL** database to initiate data collection for the Reinforcement Learning with Human Feedback (RLHF) pipeline.
- Set up evaluation metrics for Hermes Theta2, monitoring improvements in our RLHF-enhanced version of the model.
- Implemented and deployed an auto-classifier API on **AWS Lambda**, enabling scalable and efficient classification.

City of Irvine | Machine Learning Engineer for Capstone Project | Irvine, CA   Sep 2023 - Dec 2023

- Developed an LLM-powered chatbot using **FastAPI, React,** and **Llama2**, achieving a 97% reduction in time to find contract details.
- Optimized a 1M+ document dataset with the **Llama Index** and fine-tuned the **Llama2** model via prompt engineering.
- **Dockerized** the application for consistency and portability, and orchestrated its deployment using **Kubernetes**.

Dell Technologies | Data Engineer Intern | Round Rock, TX   Jun 2023 - Aug 2023

- Uncovered manual reporting inefficiencies via 50+ stakeholder interviews. Derived actionable KPIs for performance improvement.
- Streamlined reporting for Marketing team (FMMs), saving over 5,000 hours annually, and an additional 10,000 hours in the PAN ISG initiative.
- Spearheaded the integration of **Apache Airflow**, enabling automated and scalable data orchestration for the data platform.

Accenture | Data Engineer | Bangalore, KA   Oct 2020 - Jul 2022
Electrolux E-commerce Search Optimization

- Developed a hybrid recommendation system using collaborative filtering and content-based filtering**,** leading to a 12% conversion lift.
- Implemented **Airflow** logging, achieving 20% faster troubleshooting and error identification for data pipelines.

Personalized Next-Best-Offer Recommendations

- Designed dimensional data models in **SAP BW** on **HANA**, integrating customer & sales data for efficient analysis.
- Optimized data pipelines, achieving a 20% reduction in North American data transfer delays.
- Led a team of 5 to achieve a 98% on-time delivery rate for critical BI reports, exceeding SLA benchmarks.

Samsung Research Institute (SRI-D) | Software Engineer Intern | Noida, UP   Feb 2020 - Apr 2020

- Automated 90% of manual TV testing workload with **Robot Framework**, saving an estimated 200 man-hours per week.
- Built a real-time data ingestion system using **Apache Kafka** to capture and process test logs, enabling the generation of reports with **Grafana**.

PROJECTS

Chicago Transit Authority (CTA) | Real-Time Train Monitoring

- Designed and implemented a real-time train monitoring system using **Apache Kafka** (Kafka Connect & REST Proxy) for data ingestion and processing, including caching static data (routes, stations) with **Redis**.

Match Fit | AI-powered talent acquisition system

- Built a talent acquisition system using RAG with **Chroma DB** and **Llama 2**, allowing efficient search of resumes based on skills.