Hansal Shah

GitHub LinkedIn Portfolio Google Scholar hms262@cornell.edu 607-262-1984

EDUCATION

Cornell University Ithaca, New York
Masters of Engineering in Computer Science: GPA 4.02/4.3 *August 2023 - May 2024*

Relevant Courses: Large Scale Machine Learning | NLP | Computer Vision | Parallel Computing | Robot Learning

Pandit Deendayal Energy University Gujarat, India
Bachelors in Computer Science & Engineering: CGPA - 9.99/10 - **Gold Medalist** *June 2019 - May 2023*

RELEVANT EXPERIENCES

Reunion | *Data Science Intern* December 2022 - June 2023
- Managed data processing enhancements for real estate legal sales, handling complex JSONs and legal PDFs.
- Collaborated with the team to create an **ELT Data Pipeline** on **Databricks** using **Python, DBT, Spark-SQL, and Elastic Search.**
- Implemented **Natural Language to SQL Query conversion using GPT-3.5-turbo and NER tagging.**
- Lowered the query execution time by a **factor of 1.8** using data pre-fetching, caching, and parallel job processing.

DEED Mobility Solutions | *SDE Intern* May 2022 - August 2022
- Led the development of a full-stack web application using **ReactJS**, **NodeJS**, **ExpressJS**, and **MongoDB**.
- Implemented an IoT-based feature in the application for **real-time EV tracking on Google Maps**, improving customer experience and vehicle management with a real-time **delay of less than 1 millisecond**.

OPEN SOURCE CONTRIBUTIONS

dbt-labs/dbt-core | *Contributor* March 2023
- Resolved Issue #7138 in dbt snapshots, **enabling tracking and management of non-chronological data insertions.** by optimizing 'default__snapshot_merge_sql' and 'default__snapshot_staging_table' macros

PROJECTS

Cloneadvisor - Chatbot | Vertical Search Engine - Website November 2023 - Present
- Lead the creation of an **Retrieval-Augmented Generation** pipeline over an expert's domain knowledge.
- Established a pipeline with **query optimization, reciprocal rank fusion, document re-ranking, and fact-checking** to enhance the response accuracy.
- Drove the **hallucination of the pipeline down to 8.47%**.
- Tamed the pipeline to **dodge off-topic conversations**, handle normal chit-chat differently, and **ensured that it generated non-controversial and non-harmful responses** using guardrails.

README.md Generator | Tool / GitHub March 2022 - April 2022
- Developed a user-friendly **React** application that automates the creation of README files for project repositories, enhancing the documentation process for developers with **real-time markdown rendering**.

SKILLS

Programming Languages
Python, SQL, Scala, JavaScript, Java, C/C++, R

Tools/Databases
MongoDB, PostgreSQL, MySQL, Pinecone

Frameworks and Technologies
React, Flask, FastAPI, Node.js, Express.js
NeMo-Guardrails

Machine Learning Models
llama, bge-reranker-large, jina-embeddings,
bert-base-uncased, gpt-3.5-turbo, text-davinci
U-Net, RNN, LSTM

Tools
Git/GitHub, VS Code, Databricks, DBT, Microsoft
Azure, AWS Glue, Spark, Hadoop

Languages: English, Hindi, Gujarati