

Hansal Shah

 [GitHub](#)  [LinkedIn](#)  [Portfolio](#)  [Google Scholar](#)  hms262@cornell.edu  607-262-1984

EDUCATION

Cornell University

Masters of Engineering in Computer Science: GPA 4.14/4.3

Ithaca, New York

August 2023 - May 2024

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

Pandit Deendayal Energy University

Bachelors in Computer Science & Engineering: CGPA - 9.99/10 - **Gold Medalist**

Gujarat, India

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, model 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

R2 Decide - Chatbot | Verticle Search Engine - [Discord Bot](#)

November 2023 - Present

- Leading the creation of a **Retrieval-Augmented Generation** system with a fine-tuned **LLAMA 2** model, using **PEFT/ QLoRA**, for personalized, expert responses.
- Testing a **Conversational Search Engine** that guides the conversation and addresses the needs of the user.
- Established a pipeline with **query optimization**, **reciprocal rank fusion**, **document reranking**, and **fact-checking** to enhance response accuracy.

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