

LLM Assignment

Topic : SU Chatbot

Team Members: Deepak Kumar, Manu Kumar, Sahil Kumar, Sanjay V P

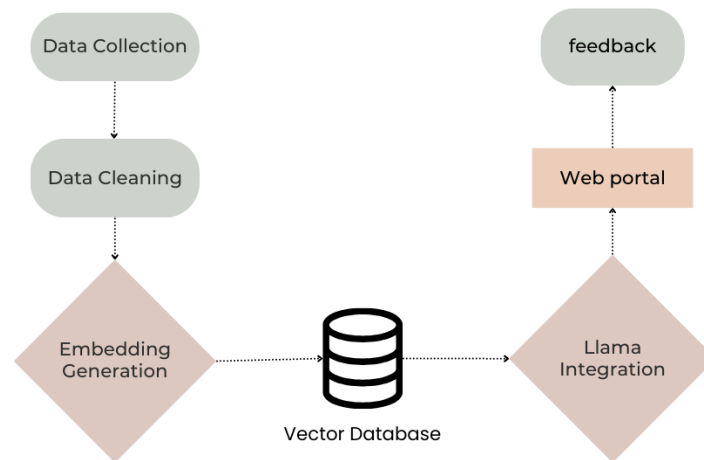
Objective

To develop a user-friendly RAG (Retrieval-Augmented Generation) system that combines an Information Retrieval (IR) module with a Generative AI model. The system will be extended to include a web portal where users can ask questions about joining Sitare University and receive detailed responses. The portal will also include an option for users to provide feedback.

Individual Contributions

- **Deepak Kumar:**
 - Developed the IR pipeline:
 - Converted text into embeddings using a sentence transformer model.
 - Set up pgVector with PostgreSQL and stored embeddings in the database.
 - Implemented cosine similarity checks using pg_vector.
- **Sahil Kumar:**
 - Transformed the IR pipeline into a RAG system by integrating *LlaMA 3.1 (70B Versatile)*.
 - Performed data cleaning to enhance response accuracy.
- **Manu Kumar:**
 - Collected and curated data for the chatbot.
 - Tested the model for accuracy with over 250 queries.
- **Sanjay V P:**
 - Integrated the system with a web application.
 - Worked on collecting user feedback.
 - Created the report for the LLM Assignment.

Process flow



- **Data Collection:**

- The initial dataset was provided by Shravan (third-year student).
- Attempted to collect data from seniors and juniors using Google Forms but received only limited responses.
- Manually generated 200+ question-and-answer pairs to create a more robust dataset.
- [Link to the dataset](#) (includes google form responses and manually curated data).

- **Data Cleaning:**

- Removed irrelevant answers and replaced them with more suitable responses.
- Eliminated duplicate entries in the dataset.
- Tried using NLTK and Regular Expression for text preprocessing, since it reduces accuracy we leave it.

- **IR pipeline creation**

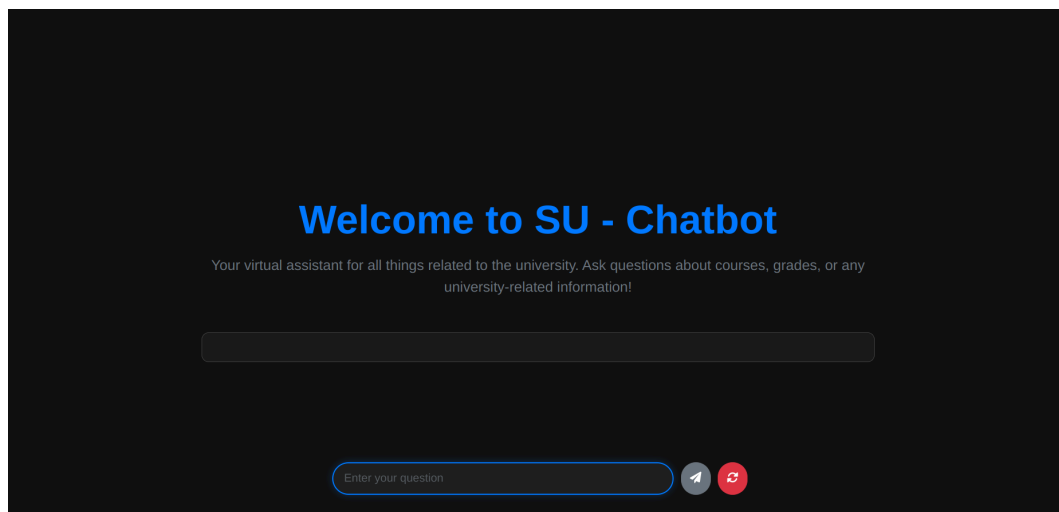
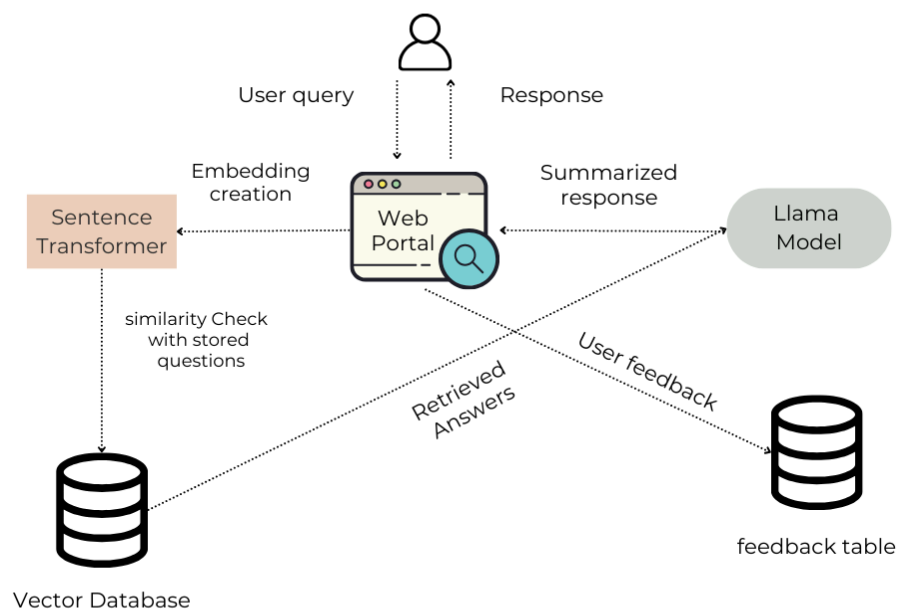
- We experimented with multiple models before selecting the most accurate one, with optimal dimensions for embeddings.
 - **all-minilm-l6-v2**
 - This is a lightweight version of the MiniLM model.
 - Ideal for applications where speed is a priority.
 - **all-mpnet-base-v2**
 - It is designed for general-purpose sentence embedding tasks.

- **multi-qa-mpnet-base-dot-v1**
 - This is a version of MPNet specifically fine-tuned for multi-domain question answering (QA) tasks.
 - Since it works well with retrieval systems that need to answer a wide range of questions accurately and efficiently, it is suitable for our use case.
 - Used **multi-qa-mpnet-base-dot-v1** to convert user queries into embeddings and it is optimized for question-answer retrieval .
 - Stored questions, answers, and embeddings in a PostgreSQL table using pgVector extension.
 - Leveraged **pg_vector** to compute cosine similarity between user queries and the stored data.
- **Generative AI model integration:**
 - Integrated **LLaMA 3.1 (70B Versatile)** to generate concise and accurate summaries based on retrieved information.
 - [Code](#) for the whole pipeline.
- **Web Portal Development**
 - Developed an intuitive, web-based interface using Flask, along with HTML, CSS, and JavaScript, to efficiently address user queries and gather feedback. The interface ensures a seamless user experience.
 - Stored user feedback in a PostgreSQL table for analysis.
 - [Link for](#) the code of integrated web portal.
- **Accuracy and Relevance:**
 - Manually verified responses to 250 queries spanning 12 different topics.
 - Out of 250 queries:
 - Accurate responses were retrieved for 200+ queries.
 - 30+ queries lacked relevant answers in the dataset.
 - 10+ queries returned irrelevant responses.
 - Achieved approximately 80% accuracy.
- **User Feedback:**
 - Collected feedback from seniors.
 - [Link to the](#) collected feedback.

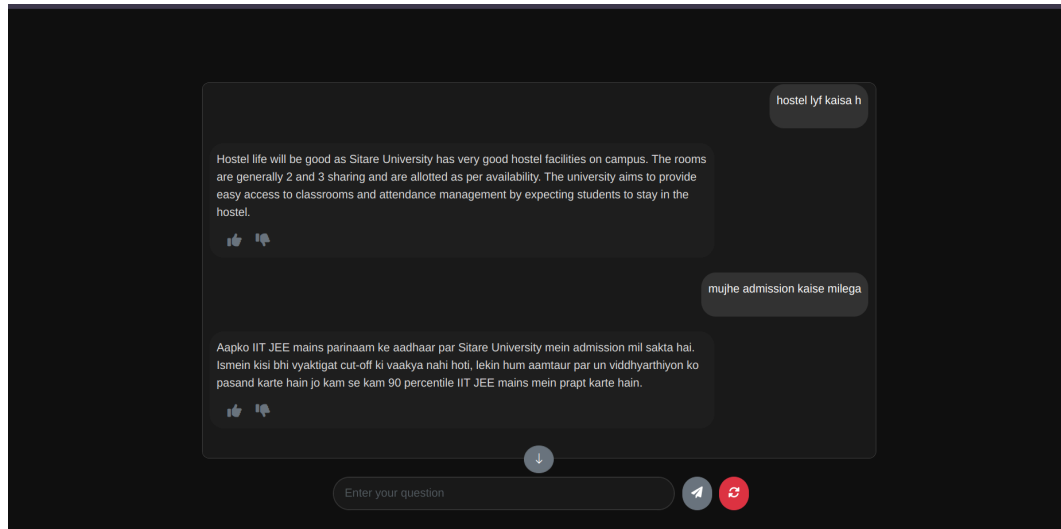
Challenges Faced

- **Data Collection:** Collecting an extensive dataset was a time-intensive process.
- **Manual Verification:** Verifying the accuracy of 250+ queries manually was labor-intensive.
- **Compatibility Issues:** Resolving compatibility issues related to embedding integration with pg_vector consumed significant time.

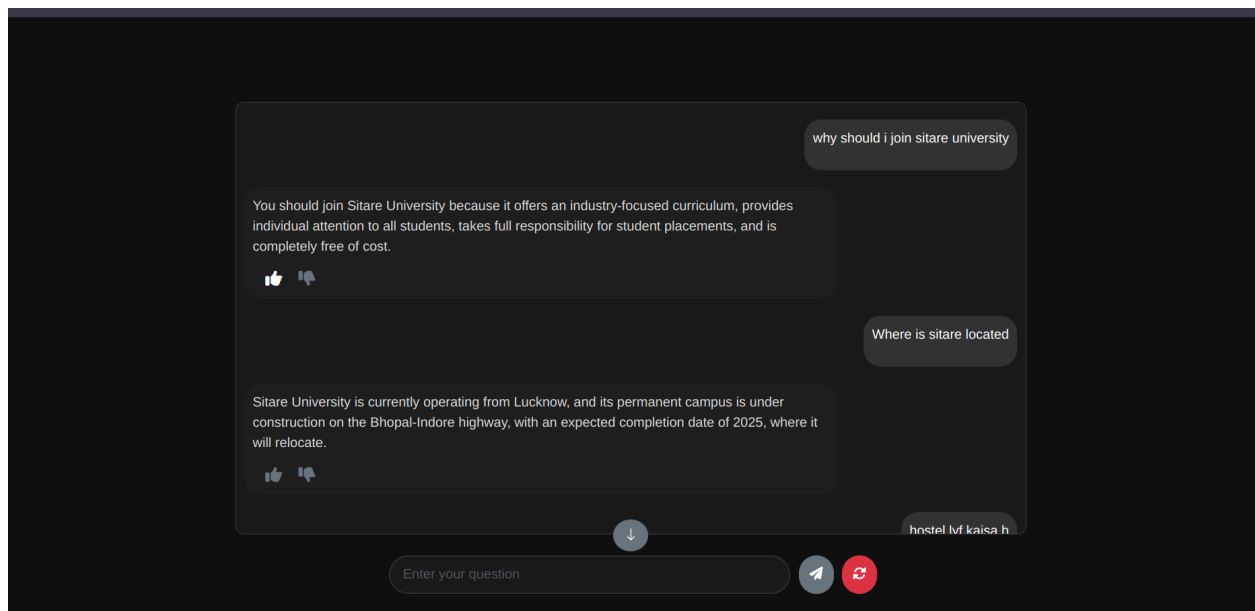
Website flowchart



This is the opening page of our Chatbot.



We tested queries in both Hindi and English, and the chatbot is working flawlessly with Hindi inputs.



Users can provide feedback by clicking on the thumbs icons.

id	message_id	action	question	
16	10	like	mess food	Our mess provides hygienic food of very good quality.
17	2	like	hiiii	Good day, I'd be happy to help.
18	6	like	who is amit singhal?	Hello, Amit Singhal is the founder of Sitare University and its founder.
19	2	like	yhnuij	We're currently in the process of collecting data to provide a comprehens
21	10	like	ikjdkh	It seems like you've entered a random combination of characters. Could yo
22	11	like	how many students are there in the university	There are 221 students at Sitare University as of 2024.
23	18	unlike	What course are there in 3rd year	In the 3rd year at Sitare University, students take courses in:
				1. Frontend
				2. Systems
				3. More Machine Learning
				Additionally, there are Hands-on Projects and Problem Solving as part of
28	10	like		Good morning. I'd be happy to assist you today. How can I help?
31	8	like		Good morning. I'd be happy to assist you today.
32	6	like	how t get admission in sitare univesirly	To get admission in Sitare University, you need to appear for the IIT JEE
34	6	like	hjoikjhjhgfgfgrfkhjghgghjkjghj;	I'm here to help with your inquiries about Sitare University. It seems yo
35	4	like	sitare	Sitare refers to Sitare University, a university focused on Computer Sci
36	2	like	What is Sitare university?	Sitare University is a unique initiative that provides high-quality Compl
37	8	like	Who are you whose knows very well about the sitare	I am part of the team associated with Sitare University, which is a uniqu
38	10	unlike	What if Amit singhal will no longer be able to support Sitare university?	We're currently in the process of collecting data to provide a comprehens
39	21	like	What if I am not able to secure 90 percentile but income is lesser than 3 lakhs	You may be considered for admission at Sitare University, and since your
40	19	like	How much fees will be charged from the students?	For students who meet the criteria of securing a minimum of 90 percentile
41	21	unlike	How bigger classes are divided in the multiple sections	We're currently in the process of collecting data to provide a comprehens
42	21	like	Sitare ke baare mai mujhe batao	Sitare ek university hai jo poor parivaaron ke hoshiaar vidhyarthiyon ko
				Sitare University ka lakshya hain high-quality computer science education
				University ki website par visit karke aap aur adhik jaankari prapt kar s
45	2	unlike	why do we need siare	We're currently in the process of collecting data to provide a comprehens
46	2	like	why should i join sitare university	You should join Sitare University because it offers an industry-focused d

User feedback is stored in a database table, allowing us to analyze it and continuously improve the chatbot's performance.

Links

- Here is the link for the whole folder where we have uploaded our whole code, dataset and feedback. [Source](#)