VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

Artificial Intelligence– BCE4306L

Slot: C2

Winter Semester – 2023-24

Digital Assignment-2

Sairudra More **21BCE1389**

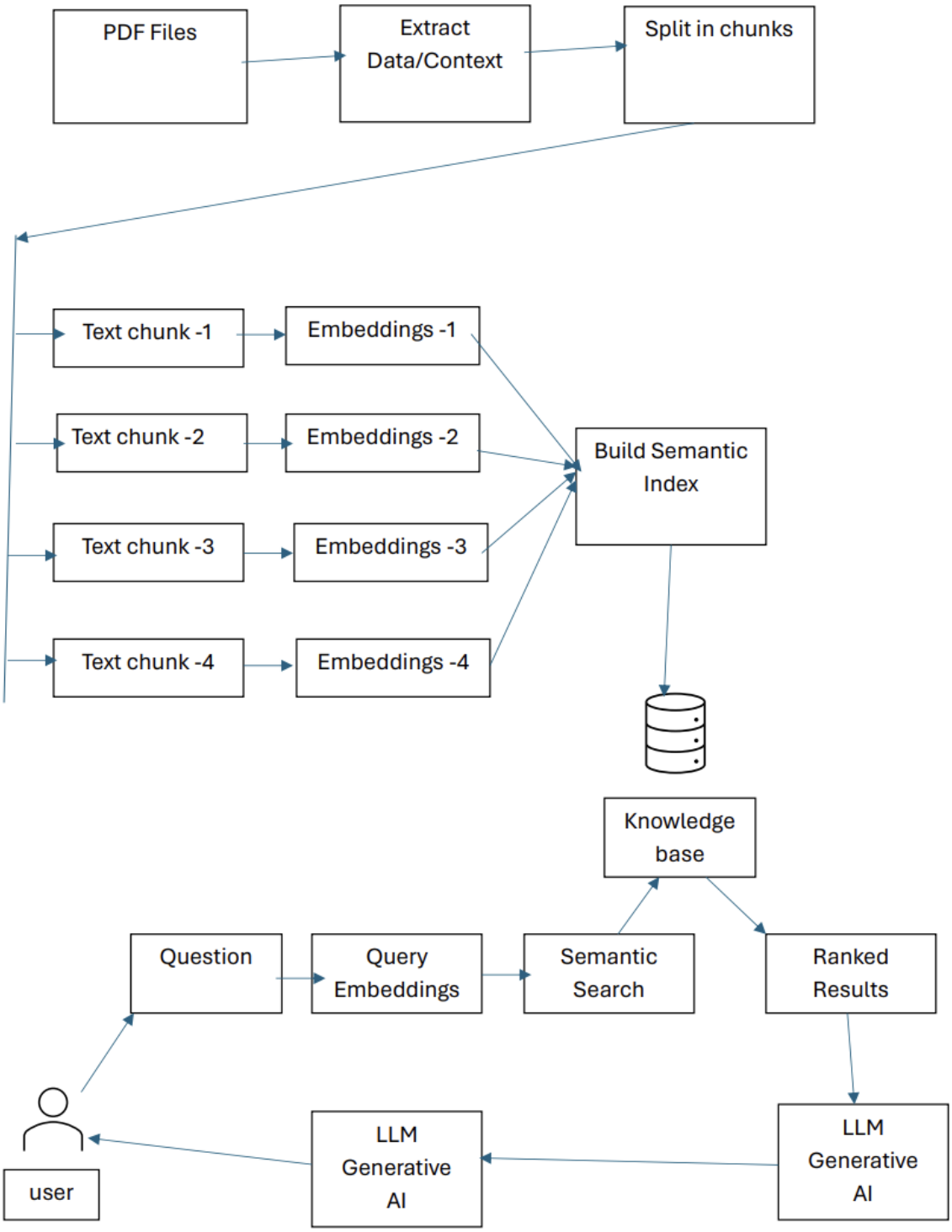
Vrinda Sanivarapu **21BCE1370**

Submitted to

Dr. Christopher Columbus C

SCOPE

Working – Flow Diagram



Explanation:

The diagram depicts a process for extracting data and using it to answer questions with a large language model (LLM). Here's a breakdown of the process:

1. **Extracting data from files:** The first step involves extracting data from multiple files. This could involve things like text, tables, or images.

2. **Splitting the data into chunks:** The data is then split up into smaller, more manageable chunks.

3. **Overlapping the chunks:** The chunks are overlapped to create an illusion that all the data are connected to 1 another.

4. **Generating Embeddings:** Each chunk of text data is then passed through a process that generates embeddings. An embedding is a mathematical representation of a piece of text that captures its meaning.
   formula: (word-topic distribution × topic importance × topic-document distribution)

5. **Building a Semantic Index:** The embeddings are used to build a semantic index. A semantic index is a

data structure that allows you to efficiently search for similar text.

6. **Querying the System:** Once the semantic index is built, you can query the system with a question.

7. **Matching the query to the Semantic Embeddings:** The system then tries to match the query to the semantic embeddings in the index.

8. **Finding similar questions:** The system finds questions in the index that are like the query.

9. **Finding Answers Using the Knowledge Base:** The system then uses a knowledge base to find answers to similar questions. A knowledge base is a collection of information that can be used to answer questions.

10. **Using an LLM to Generate Text:** Finally, the system uses an LLM to generate text that answers the query. An LLM is a large language model, which is a type of artificial intelligence that can generate text.

Overall, this diagram depicts a system that can extract data from files and process that data to create a semantic index, and then use that index to answer questions using a large language model.

How it is better than other LLMs?

1st) It utilizes vectors/embeddings instead of keyword search, thus providing faster and more accurate results than other LLM models**.**

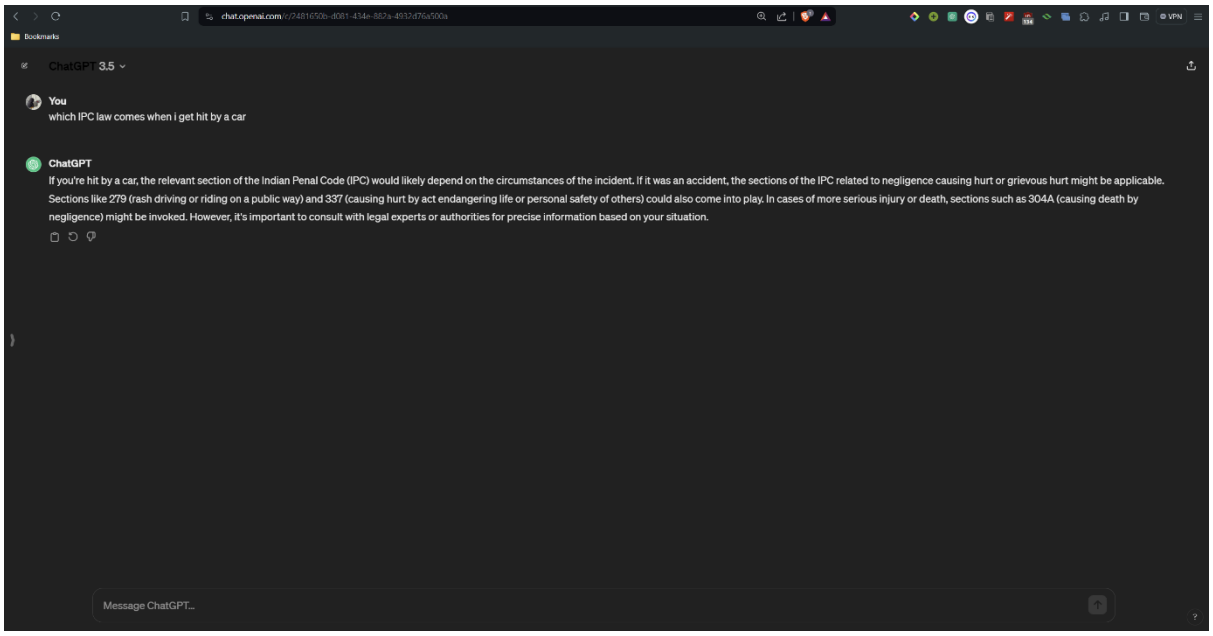
2nd) It utilizes the IPC document from the Government of India and doesn't engage in web crawling like other LLMs, ensuring that the results produced are 100% accurate.

3rd) All the Law Bots are paid services, hence most of the general public doesn't have the opportunity to access them.

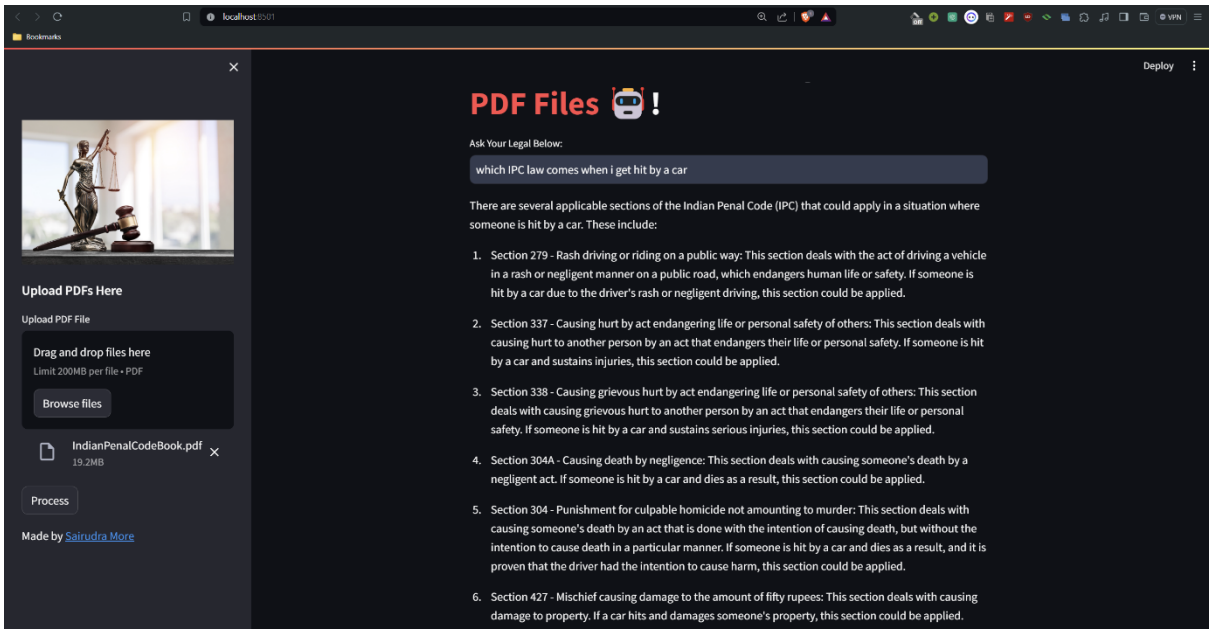
4th) The LawBot is also capable of accepting multiple documents as inputs. Hence, it possesses the necessary knowledge with which it can work, rather than users specifying default data. Thus, users can focus on the main agenda.

5th) How it is better than the standard LLMs like ChatGPT, Gemini, etc
Here is an example:

ChatGPT:



LAWBOT



As you can see, standard GPTs provide very short and less accurate results, while LAWBOT is strong in its knowledge and is able to answer questions better for the same question as input.