Semantic Spotter Project- Build a RAG System

1.Project Goal

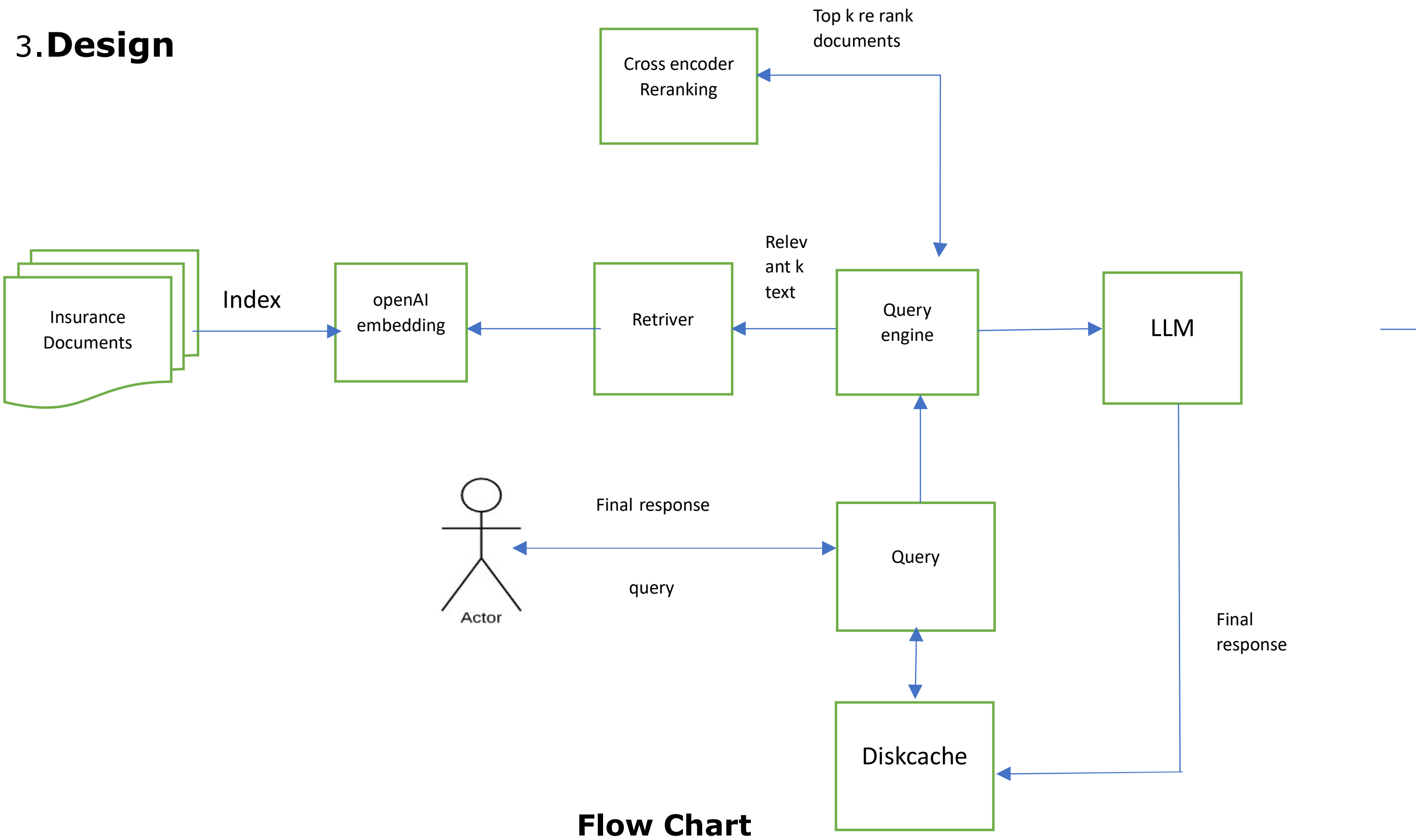
Build a project in the insurance domain. The goal of the project will be to build a robust generative search system capable of effectively and accurately answering questions from various policy documents. Using LlamaIndex to build the generative search application.

2. Data Source

Seven HDFC insurance documents in Pdf format provides inside a single folder.

a. HDFC-Life-Easy-Health-101N110V03-Policy-Bond-Single-Pay.pdf
b. HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf
c. HDFC-Life-Group-Term-Life-Policy.pdf
d. HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document.pdf
e. HDFC-Life-Sanchay-Plus-Life-Long-Income-Option-101N134V19-Policy-Document.pdf
f. HDFC-Life-Smart-Pension-Plan-Policy-Document-Online.pdf
g. HDFC-Surgicare-Plan-101N043V01.pdf

3.Design



Flow Chart

Descriptions about the Architecture:

1. Documents: List of seven HDFC insurance documents provides inside a single folder.

2. Open API embedding: OpenAPI embedding as Vector DB for indexing insurance documents in the form of embedding.

3. Query Engine: We are using Query Engine Module of Llammaindex for performing semantic Search. Query Engine will use internally Retriever and SentenceTransformerRerank- model="cross-encoder/ms-marco-MiniLM-L-2-v2 retrieve top-k relevant nodes from embedding.

4. LLM: top k-documents along with user query will be passed to LLM to generate the accurate response.

5. Caching:" Caching is being used to improve the read operation. Recent similar search will be store in Caching and user query first will be served from Cache. If user query not found in cache, then query will be forwarded to query engine and then LLM to generate the response.

6. Meta data : Along with Response we are also returning docs reference and similarly score to improve the user confidence towards the implemented RAG system.

7. SentenceTransformerRerank- model="cross-encoder/ms-marco-MiniLM-L-2-v2 Is being used to rerank the query based on semantic score.

8. Evaluation- LLM-gpt4 is used for evaluation on matrices relevancy ,faithfulness and correctness.

4. Solution Strategy

- Build a solution which should solve the following requirements:

- Users would get responses from insurance policy knowledge base.

- If user want to perform a query system must be able to response to query accurately.

- If they want to refer to the original page from which the bot is responding, the bot should provide a citation as well.

5. Tools used

- LlamaIndex has been used due to its powerful query engine, fast data processing ,easier and faster implementation using fewer lines of code.

SimpleDirectoryReader is used to read the documents.

Vectorstoreindex is used to create index.

-SentenceTransformerRerank - model="cross-encoder/ms-marco-MiniLM-L-2-v2" is used to Rerank.

-Diskcache

- openAI API key

-LLM- gpt-4 for evaluation

6. Why LlamaIndex ?

LlamaIndex is an innovative data framework specially designed to support LLM-based RAG framework application development. It offers an advanced framework that empowers developers to integrate diverse data sources with large language models. LlamaIndex includes a variety of file formats, such as PDFs and PowerPoints, as well as applications like Notion and Slack and even databases like Postgres and MongoDB.

The framework brings an array of connectors that assist in data ingestion, facilitating a seamless interaction with LLMs. Moreover, LlamaIndex boasts an efficient data retrieval and query interface.

LlamaIndex enables developers to input any LLM prompt and, in return, receive an output that is both context-rich and knowledge-augmementation.

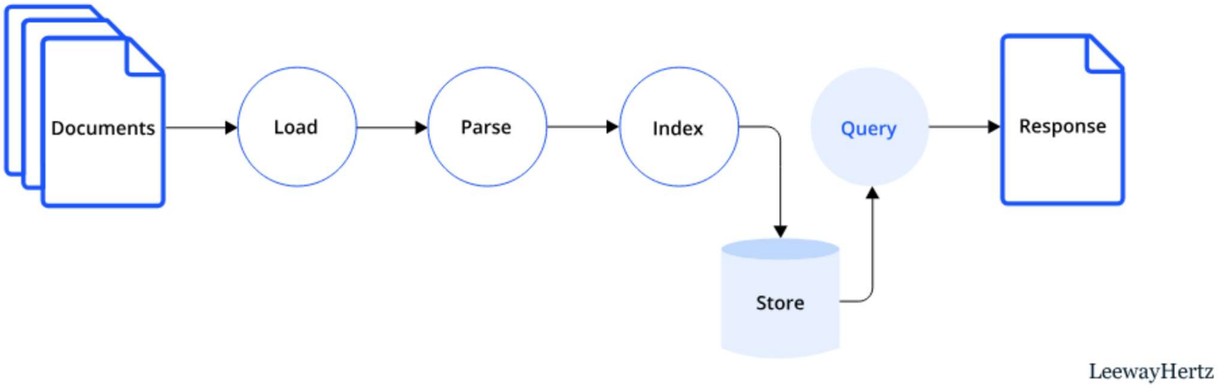
Key Feature of LlamaIndex:

● Data connectors allow ingestion from various data sources and formats.

● It can synthesize data from multiple documents or heterogeneous data sources.

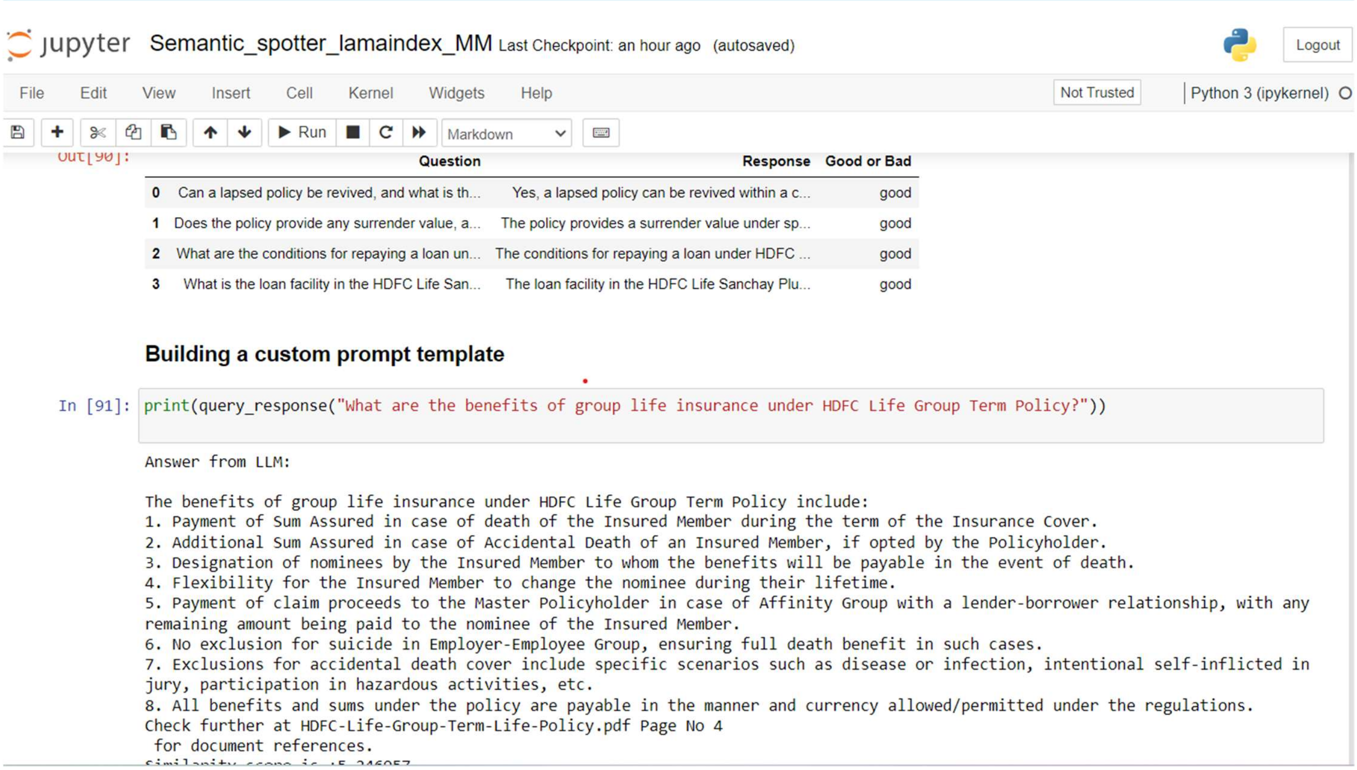
● It provides numerous integrations with vector stores, ChatGPT plugins, tracing tools,

LangChain, and more.



7. Generative Search Response from Insurance documents :

We have attached custom query generative search results.



Out[90]:

	Question	Response	Good or Bad
0	Can a lapsed policy be revived, and what is th...	Yes, a lapsed policy can be revived within a c...	good
1	Does the policy provide any surrender value, a...	The policy provides a surrender value under sp...	good
2	What are the conditions for repaying a loan un...	The conditions for repaying a loan under HDFC ...	good
3	What is the loan facility in the HDFC Life San...	The loan facility in the HDFC Life Sanchay Plu...	good

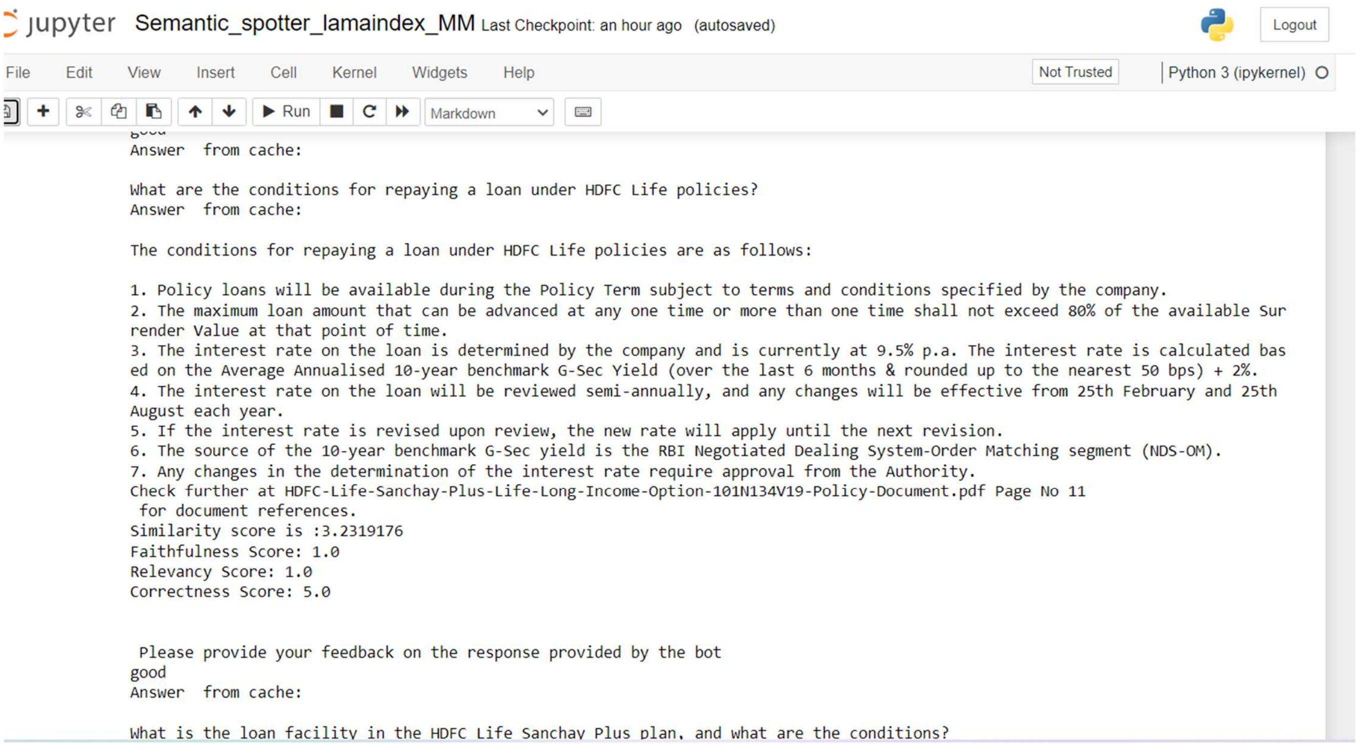
Building a custom prompt template

```
In [91]: print(query_response("What are the benefits of group life insurance under HDFC Life Group Term Policy?"))
```

```
Answer from LLM:

The benefits of group life insurance under HDFC Life Group Term Policy include:
1. Payment of Sum Assured in case of death of the Insured Member during the term of the Insurance Cover.
2. Additional Sum Assured in case of Accidental Death of an Insured Member, if opted by the Policyholder.
3. Designation of nominees by the Insured Member to whom the benefits will be payable in the event of death.
4. Flexibility for the Insured Member to change the nominee during their lifetime.
5. Payment of claim proceeds to the Master Policyholder in case of Affinity Group with a lender-borrower relationship, with any
remaining amount being paid to the nominee of the Insured Member.
6. No exclusion for suicide in Employer-Employee Group, ensuring full death benefit in such cases.
7. Exclusions for accidental death cover include specific scenarios such as disease or infection, intentional self-inflicted in
jury, participation in hazardous activities, etc.
8. All benefits and sums under the policy are payable in the manner and currency allowed/permitted under the regulations.
Check further at HDFC-Life-Group-Term-Life-Policy.pdf Page No 4
 for document references.
Similarity score is :5.246057
```

8. Multiple Query Response



```
good
Answer  from cache:

What are the conditions for repaying a loan under HDFC Life policies?
Answer  from cache:

The conditions for repaying a loan under HDFC Life policies are as follows:

1. Policy loans will be available during the Policy Term subject to terms and conditions specified by the company.
2. The maximum loan amount that can be advanced at any one time or more than one time shall not exceed 80% of the available Sur
render Value at that point of time.
3. The interest rate on the loan is determined by the company and is currently at 9.5% p.a. The interest rate is calculated bas
ed on the Average Annualised 10-year benchmark G-Sec Yield (over the last 6 months & rounded up to the nearest 50 bps) + 2%.
4. The interest rate on the loan will be reviewed semi-annually, and any changes will be effective from 25th February and 25th
August each year.
5. If the interest rate is revised upon review, the new rate will apply until the next revision.
6. The source of the 10-year benchmark G-Sec yield is the RBI Negotiated Dealing System-Order Matching segment (NDS-OM).
7. Any changes in the determination of the interest rate require approval from the Authority.
Check further at HDFC-Life-Sanchay-Plus-Life-Long-Income-Option-101N134V19-Policy-Document.pdf Page No 11
 for document references.
Similarity score is :3.2319176
Faithfulness Score: 1.0
Relevancy Score: 1.0
Correctness Score: 5.0


 Please provide your feedback on the response provided by the bot
good
Answer  from cache:

What is the loan facility in the HDFC Life Sanchay Plus plan, and what are the conditions?
```

9. Challenges faced:

- Faced compatibility issue while importing RAGAS for evaluation.
- Compatibility issue while using gptcache

10. Alternative Solutions:

-diskcache is used instead of gptcache

- instead of importing RAGAS we imported

```
 from llama_index.core.evaluation import (
    CorrectnessEvaluator,
    FaithfulnessEvaluator,
    RelevancyEvaluator,
)
```

11. Alternative option

-reranking could be done with Cohere rerank

BY: Madhusmita Ghosh