Semantic Spotter Llamaindex Project

Build an RAG System

- **Created By Nishu Kumari**

1. Project Goal

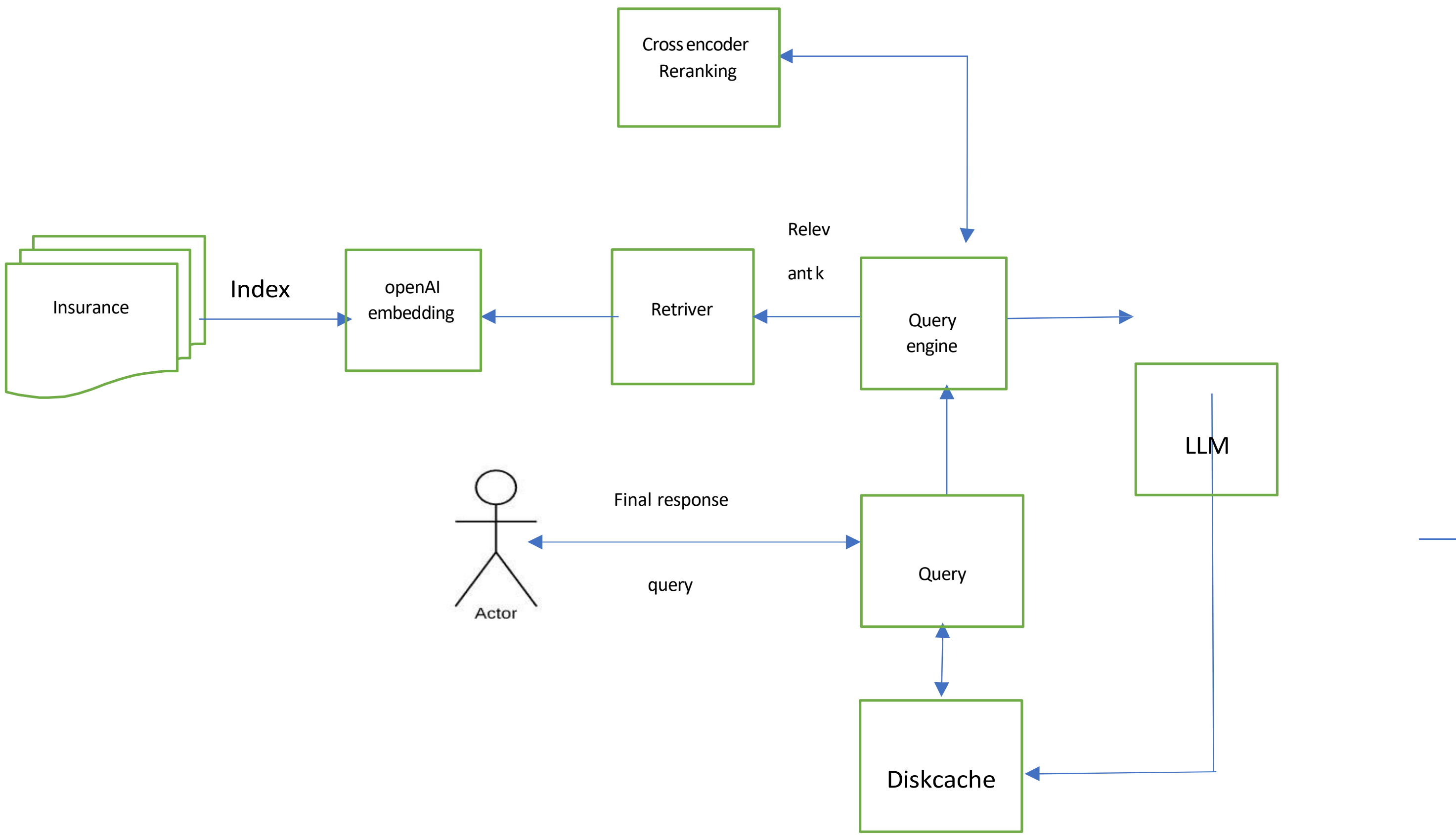
Build a project in the insurance domain. The goal of the project will be to build a robust generative search system capable of effectively and accurately answering questions from various policy documents. Using LlamaIndex to build the generative search application.

2. Data Source

Seven HDFC insurance documents in Pdf format are provided 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



Final response

Flow Chart

Descriptions about 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

- Build a solution which should solve the following requirements:
- Users would get responses from insurance policy knowledge base.
- If user wants 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 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

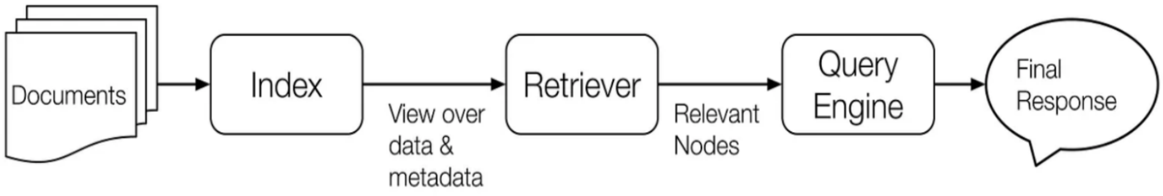
6. Why LlamaIndex?

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 seamless interaction with LLMs. Moreover, LlamaIndex boasts an efficient data retrieval and query interface.

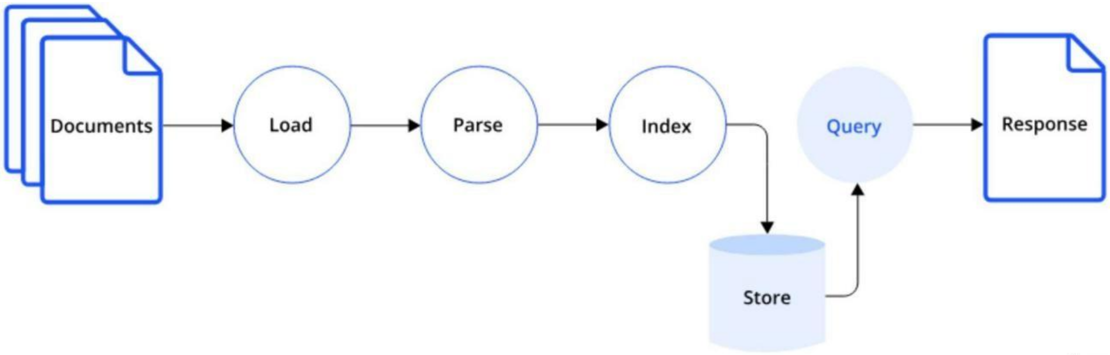
LlamaIndex enables developers to input any LLM prompt and, in return, receive an output that is both context-rich and knowledge-augmentation.

Core Components Of LlamaIndex



Key Feature of LlamaIndex:

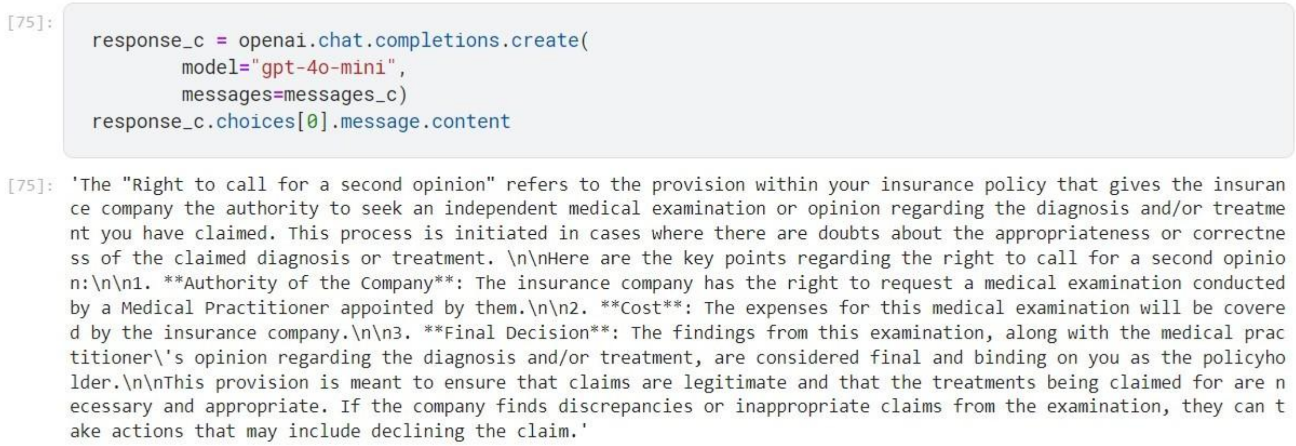
- Data connectors allow ingestion from various data sources and formats.
- It can synthesize data from multiple documents or heterogeneous data sources.
- It provides numerous integrations with vector stores, ChatGPT plugins, tracing tools, LangChain, and more.



LeewayHertz

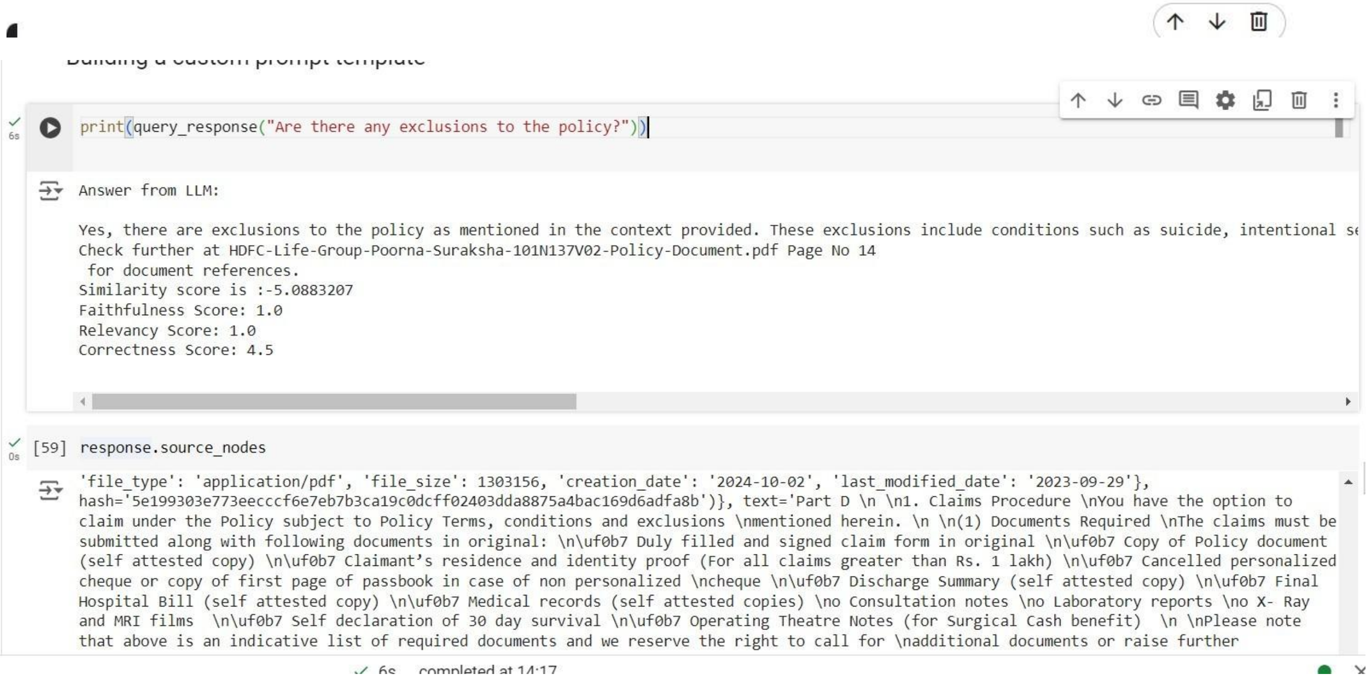
7. **Generative Search Response from Insurance documents:**

We have attached custom query generative search results.



```
[75]:
response_c = openai.chat.completions.create(
        model="gpt-4o-mini",
        messages=messages_c)
response_c.choices[0].message.content
```

```
[75]: 'The "Right to call for a second opinion" refers to the provision within your insurance policy that gives the insuran
ce company the authority to seek an independent medical examination or opinion regarding the diagnosis and/or treatme
nt you have claimed. This process is initiated in cases where there are doubts about the appropriateness or correctne
ss of the claimed diagnosis or treatment. \n\nHere are the key points regarding the right to call for a second opinio
n:\n\n1. **Authority of the Company**: The insurance company has the right to request a medical examination conducted
by a Medical Practitioner appointed by them.\n\n2. **Cost**: The expenses for this medical examination will be covere
d by the insurance company.\n\n3. **Final Decision**: The findings from this examination, along with the medical prac
titioner\'s opinion regarding the diagnosis and/or treatment, are considered final and binding on you as the policyho
lder.\n\nThis provision is meant to ensure that claims are legitimate and that the treatments being claimed for are n
ecessary and appropriate. If the company finds discrepancies or inappropriate claims from the examination, they can t
ake actions that may include declining the claim.'
```



Building a custom prompt template

```
print(query_response("Are there any exclusions to the policy?"))
```

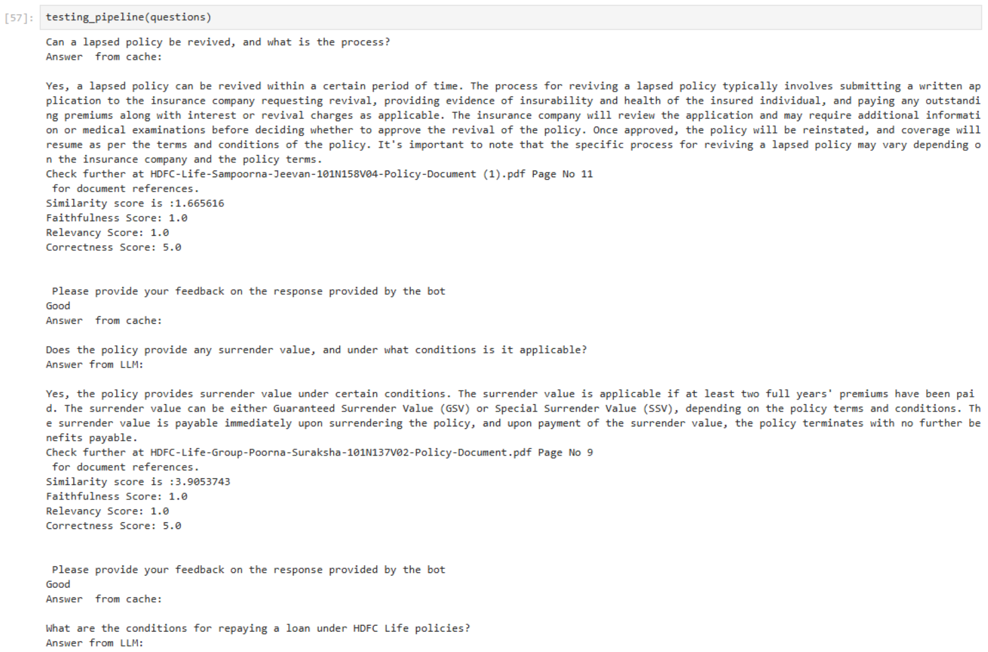
```
Answer from LLM:

Yes, there are exclusions to the policy as mentioned in the context provided. These exclusions include conditions such as suicide, intentional se
Check further at HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf Page No 14
 for document references.
Similarity score is :-5.0883207
Faithfulness Score: 1.0
Relevancy Score: 1.0
Correctness Score: 4.5
```

```
[59] response.source_nodes
```

```
'file_type': 'application/pdf', 'file_size': 1303156, 'creation_date': '2024-10-02', 'last_modified_date': '2023-09-29'},
hash='5e199303e773eecccf6e7eb7b3ca19c0dcff02403dda8875a4bac169d6adfa8b')}, text='Part D \n \n1. Claims Procedure \nYou have the option to
claim under the Policy subject to Policy Terms, conditions and exclusions \nmentioned herein. \n \n(1) Documents Required \nThe claims must be
submitted along with following documents in original: \n\uf0b7 Duly filled and signed claim form in original \n\uf0b7 Copy of Policy document
(self attested copy) \n\uf0b7 Claimant’s residence and identity proof (For all claims greater than Rs. 1 lakh) \n\uf0b7 Cancelled personalized
cheque or copy of first page of passbook in case of non personalized \ncheque \n\uf0b7 Discharge Summary (self attested copy) \n\uf0b7 Final
Hospital Bill (self attested copy) \n\uf0b7 Medical records (self attested copies) \no Consultation notes \no Laboratory reports \no X- Ray
and MRI films  \n\uf0b7 Self declaration of 30 day survival \n\uf0b7 Operating Theatre Notes (for Surgical Cash benefit)  \n \nPlease note
that above is an indicative list of required documents and we reserve the right to call for \nadditional documents or raise further
```

8. Multiple Query Response



```
[57]: testing_pipeline(questions)

Can a lapsed policy be revived, and what is the process?
Answer  from cache:

Yes, a lapsed policy can be revived within a certain period of time. The process for reviving a lapsed policy typically involves submitting a written ap
plication to the insurance company requesting revival, providing evidence of insurability and health of the insured individual, and paying any outstandi
ng premiums along with interest or revival charges as applicable. The insurance company will review the application and may require additional informati
on or medical examinations before deciding whether to approve the revival of the policy. Once approved, the policy will be reinstated, and coverage will
resume as per the terms and conditions of the policy. It's important to note that the specific process for reviving a lapsed policy may vary depending o
n the insurance company and the policy terms.
Check further at HDFC-Life-Sampoorna-Jeevan-101N158V04-Policy-Document (1).pdf Page No 11
 for document references.
Similarity score is :1.665616
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:

Does the policy provide any surrender value, and under what conditions is it applicable?
Answer from LLM:

Yes, the policy provides surrender value under certain conditions. The surrender value is applicable if at least two full years' premiums have been pai
d. The surrender value can be either Guaranteed Surrender Value (GSV) or Special Surrender Value (SSV), depending on the policy terms and conditions. Th
e surrender value is payable immediately upon surrendering the policy, and upon payment of the surrender value, the policy terminates with no further be
nefits payable.
Check further at HDFC-Life-Group-Poorna-Suraksha-101N137V02-Policy-Document.pdf Page No 9
 for document references.
Similarity score is :3.9053743
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 are the conditions for repaying a loan under HDFC Life policies?
Answer from LLM:
```

Challenges faced:

- Faced compatibility issue while importing RAGAS for evaluation.
- Compatibility issue while using gptcache
- Performance Bottlenecks
- Dependency Conflicts

9. Alternative Solutions:

-diskcache is used instead of gptcache

- instead of importing RAGAS we imported from

```
llama_index.core.evaluation import (

    CorrectnessEvaluator,

    FaithfulnessEvaluator,

    RelevancyEvaluator,

)
```

10.Alternative option

-reranking could be done with Cohere rerank.