Documentation of HelpMateAI

Submitted by: Dipak Sah, MLAI Cohort 60 Batch

Problem Statement

Describe the challenges posed by insurance documents, which are often lengthy, filled with complex jargon, and vary in structure, making it difficult for users to find the precise information they need.

Objective

Build a RAG based generative search system that can effectively search based on the given queries based on provided insurance policy document. Here, we have a single long life insurance policy document.

The primary goal of this project are as follows:

- Develop a semantic search system pipeline using the RAG (Embedding Layer, Search and Rank Layer, Generation Layer) pipeline for efficient document retrieval.
- Extract relevant information from PDF documents, store them in a structured format, and generate vector representations using Sentence Transformer Embedding or OpenAI embedding.
- Use different chunking strategy and embedding model, evaluate and use the best one.
- Implement a cache layer to enhance system performance by storing and retrieving previous queries and their results.

Design

The project should implement all the three layers effectively. It will be key to try out various strategies and experiments in various layers in order to build an effective search system. Let's explore what we need to do in each of the layers, and the possible experimentations that we can perform based on various choices.

1. The Embedding Layer: The PDF document needs to be effectively processed, cleaned, and chunked for the embeddings. Here, the choice of the chunking strategy will have a large impact on the final quality of the retrieved results. So, make sure that we try out various strategies and compare their performances.

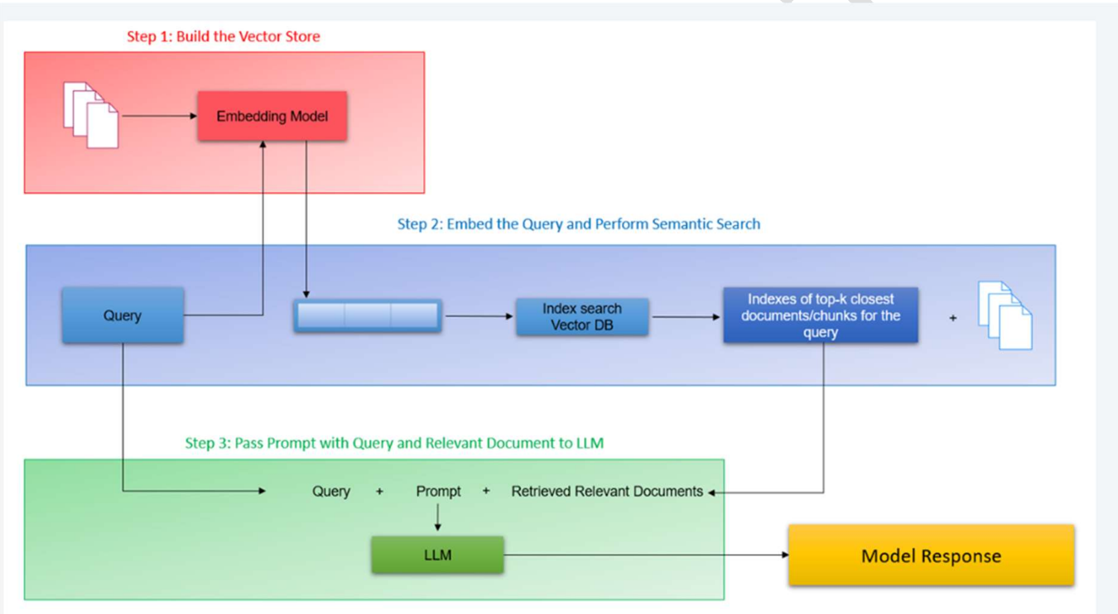
Another important aspect in the embedding layer is the choice of the embedding model. We can choose to embed chunks using the OpenAI embedding model or any model from the SentenceTransformers library on HuggingFace.

2. The Search Layer: Here, we first need to design at least 3 queries against which we will test. We need to understand and skim through the document, and accordingly come up with some queries, the answers to which can be found in the policy document.

Next, we need to embed the queries and search in ChromaDB vector database against each of these queries. Implementing a cache mechanism is also mandatory.

Finally, we need to implement the re-ranking block, and for this we can choose from a range of cross-encoding models on HuggingFace.

3. The Generation Layer: In the generation layer, the final prompt that we design is the major component. Make sure that the prompt is exhaustive in its instructions, and the relevant information is correctly passed to the prompt. We may also choose to provide some few-shot examples in an attempt to improve the LLM output.



Retrieval Augmented Generation - System Design

Implementation

Used Google Colab for development and leveraged libraries such as pdfplumber, tiktoken, openai, chromaDB, and sentence-transformers for document processing, embedding, and caching.

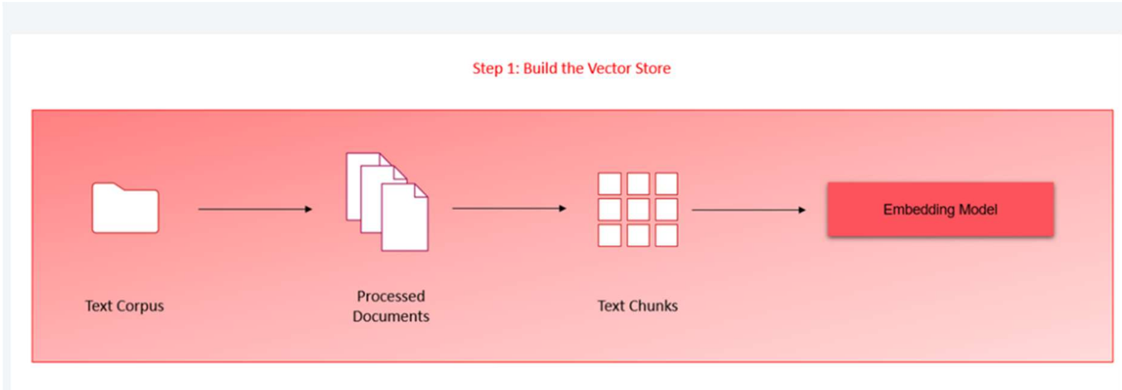
Embedding Layer

We used the PDFPlumber library to extract the text from PDF document to make a big corpus. Later we have used different chunking strategy to chunk the text before using embedding.

We used below chunking strategy:

1. Fixed length with no overlap (chunk size = 1024)
2. Fixed length with overlap (chunk size = 1024 & overlap size = 256)
3. Page level

4. Section Level (sections are manually identified by skimming through the documents)



We have used below embedding models:

1. Sentence Transformer - **all-MiniLM-L6-v2**
2. Openai - **text-embedding-ada-002**

Totally 8 different collections are created in chromaDB.

- openai_collection_fixed_length_zero_overlap_chunks
- openai_collection_fixed_length_with_overlap_chunks
- openai_collection_page_chunks
- openai_collection_section_chunks
- transformer_collection_fixed_length_with_overlap_chunks
- transformer_collection_fixed_length_zero_overlap_chunks
- transformer_collection_page_chunks
- transformer_collection_section_chunks

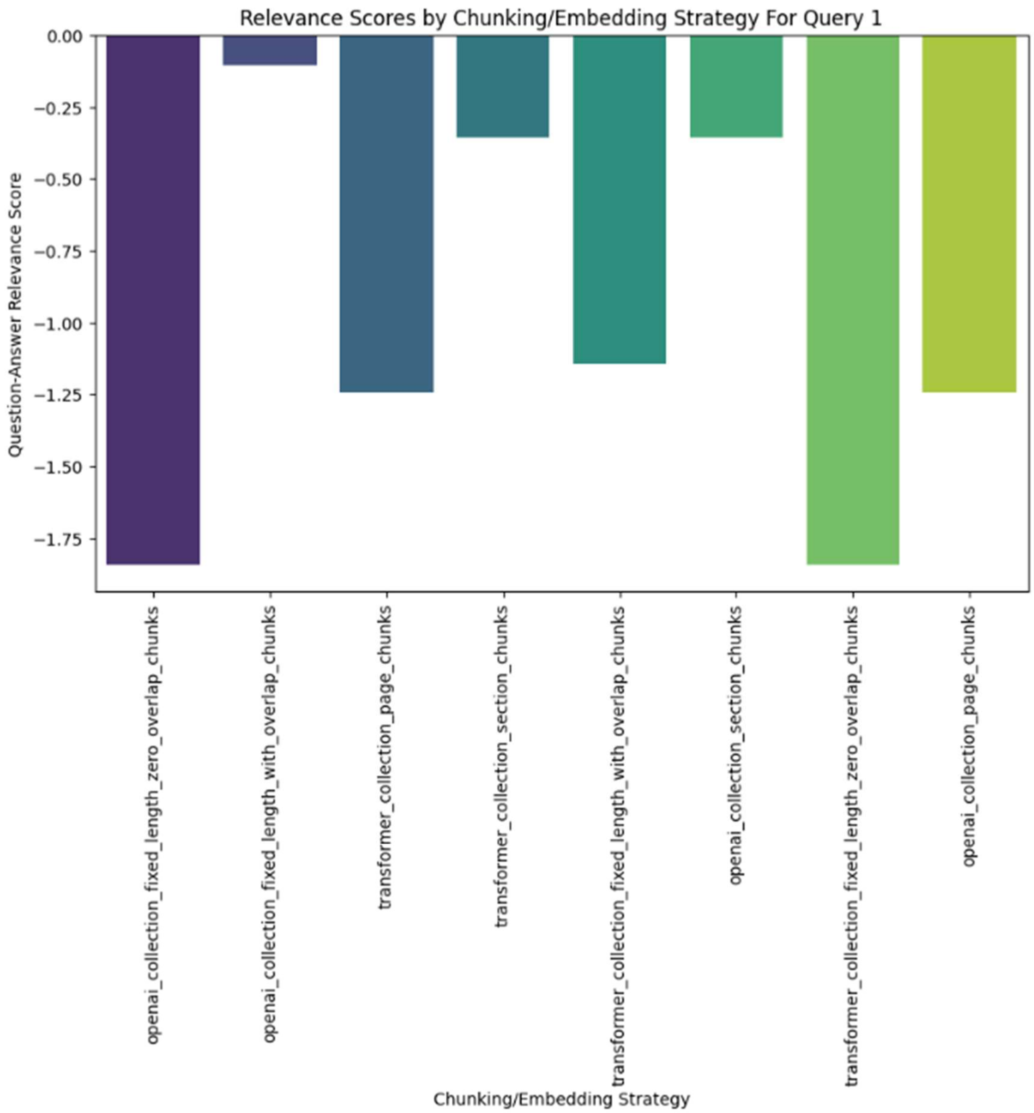
Evaluation of Chunking and Embedding strategy

We will use cross encoder to calculate relevancy between question and answer retrieved from differents embedded chunks we created mentioned above

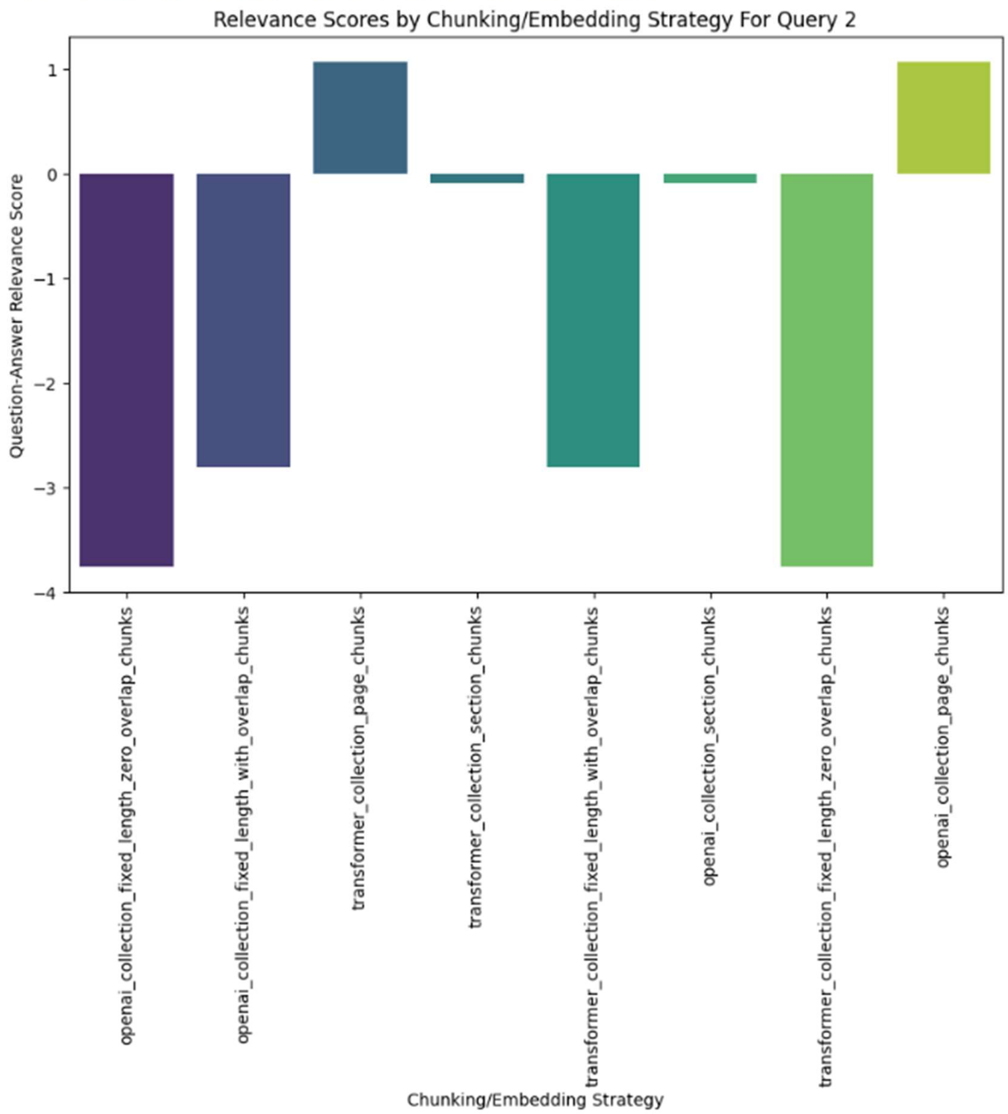
Below questions have been used to evaluate the relevance.

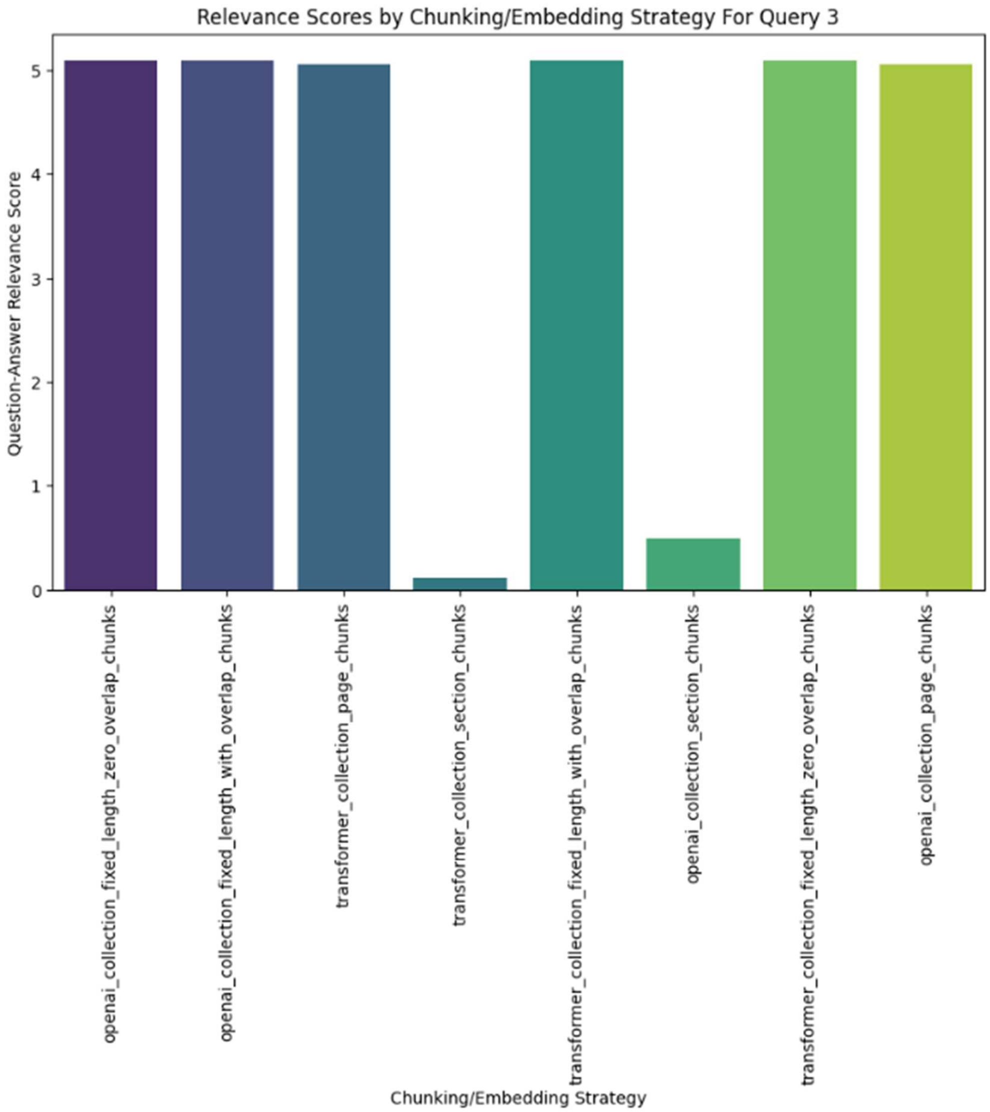
```
# We have chosen these questions to evaluate our model
questions = ["What are the key benefits provided under this life insurance policy ?",
             "What happens if the premium is not paid on time ?",
             "What are the proofs required for good health ?"]
```

Below are screenshots of the bar plot for all queries which shows relevance score between query and the answer retrieved from respective vector data.


Relevance Scores by Chunking/Embedding Strategy For Query 1
Question-Answer Relevance Score
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
openai_collection_fixed_length_zero_overlap_chunks
openai_collection_fixed_length_with_overlap_chunks
transformer_collection_page_chunks
transformer_collection_section_chunks
transformer_collection_fixed_length_with_overlap_chunks
openai_collection_section_chunks
transformer_collection_fixed_length_zero_overlap_chunks
openai_collection_page_chunks
Chunking/Embedding Strategy

```
sns.barplot(x='Collection', y='Relevance', data=df,palette='viridis')
```





[58]

```
# Checking mean score of relevance for each collections
relevance_df.groupby('Collection')['Relevance'].mean().sort_values(ascending=False)
```

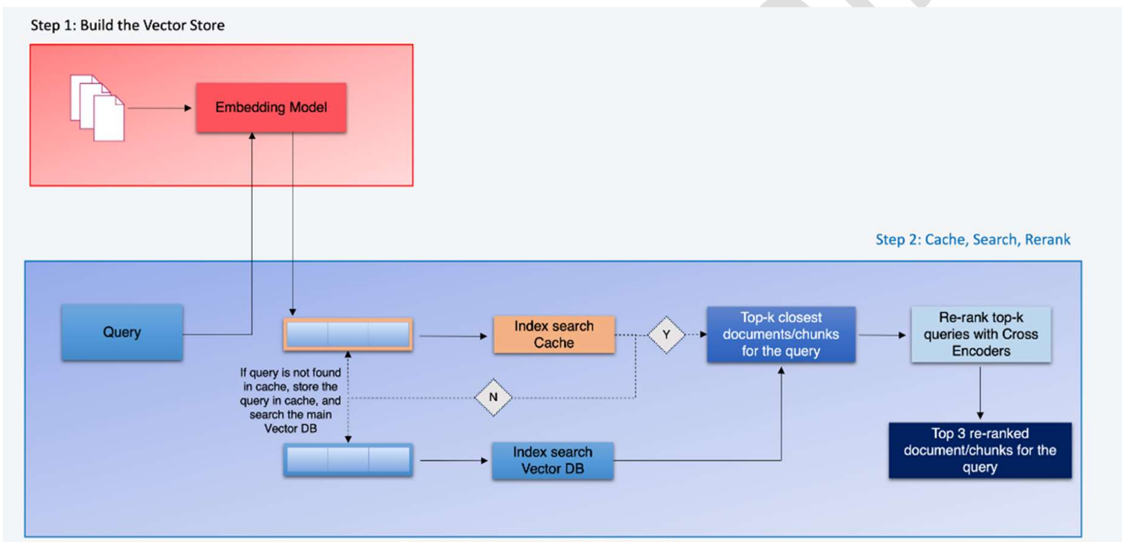
Collection	Relevance
openai_collection_page_chunks	1.625861
transformer_collection_page_chunks	1.625861
openai_collection_fixed_length_with_overlap_chunks	0.726226
transformer_collection_fixed_length_with_overlap_chunks	0.380529
openai_collection_section_chunks	0.014220
transformer_collection_section_chunks	-0.113132
openai_collection_fixed_length_zero_overlap_chunks	-0.170578
transformer_collection_fixed_length_zero_overlap_chunks	-0.170578

Above results shows that **page level chunking** is better. As such it does not tell which embedding is good but when you look overall, **OpenAI embedding** looks better. So we continue our development with Page level chunking and using openai embedding.

Search Layer

Creating a cache collection within the vector database improves the scalability and performance of the RAG system by storing and retrieving previous queries and their responses, especially when dealing with large documents and multiple users.

This cache layer reduces the need for repeated semantic similarity searches, leading to faster response times. The cache stores the semantic meaning of queries, allowing the system to bypass the bottleneck of semantic searches for previously seen queries. As a result, users experience quicker and more efficient application performance.

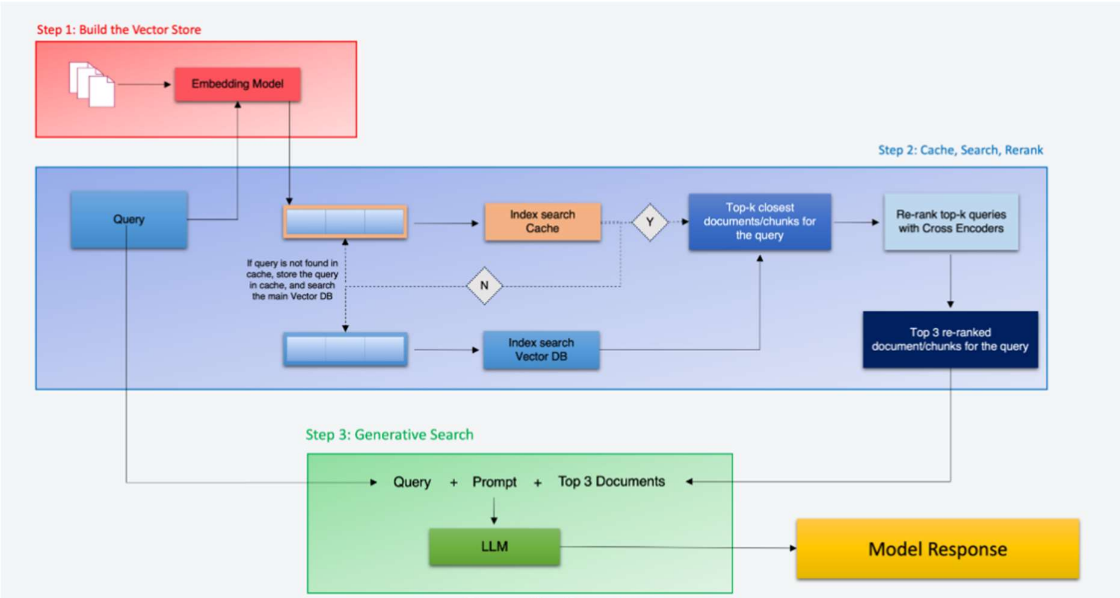


Semantic Search with Re-Ranking

So far, we have implemented a semantic search layer with a cache. Now, the focus is on enhancing performance with a re-ranking layer, which sorts the top K results based on their relevance to the query. This stage aims to improve accuracy and relevance, reduce irrelevant information, and provide more personalized search results. The project will utilize cross-encoder models for re-ranking due to their ability to accurately measure semantic similarity between text sequences. The next step is to incorporate a generative AI model for more comprehensive responses.

Generation Layer

In this stage of the generative search application, the generation layer uses a large language model (LLM) to enhance the system's output. This layer takes the top K documents retrieved by the semantic search layer, along with the user's query and system prompt, to generate a relevant response. The LLM processes these inputs to produce an accurate and contextually appropriate answer. The system prompt can be adjusted based on the domain for optimal performance.



Generative Search

❖ **Role**	You are a helpful assistant in the insurance domain who can effectively answer user queries about insurance policies and documents.
❖ **Context**	You have a question asked by the user in '{query}' and you have some search results from a corpus of insurance documents in the dataframe '{top_3_RAG}'. These search results are essentially one page of an insurance document that may be relevant to the user query. The column 'documents' inside this dataframe contains the actual text from the policy document and the column 'metadata' contains the policy name and source page. The text inside the document may also contain tables in the format of a list of lists where each of the nested lists indicates a row.
❖ **Task**	Use the documents in '{top_3_RAG}' to answer the query '{query}'. Frame an informative answer and also, use the dataframe to return the relevant policy names and page numbers as citations.
❖ **Guidelines**	Follow the guidelines below when performing the task. 1. Try to provide relevant/accurate numbers if available. 2. You don't have to necessarily use all the information in the dataframe. Only choose information that is relevant. 3. If the document text has tables with relevant information, please reformat the table and return the final information in a tabular in format. 4. Use the Metadatas columns in the dataframe to retrieve and cite the policy name(s) and page numbers(s) as citation. 5. If you can't provide the complete answer, please also provide any information that will help the user to search specific sections in the relevant cited documents. 6. You are a customer facing assistant, so do not provide any information on internal workings, just answer the query directly.
❖ **Output Format**	The generated response should answer the query directly addressing the user and avoiding additional information. If you think that the query is not relevant to the domain, reply that the query is irrelevant. Provide the final response as a well-formatted and easily readable text along with the citation. Provide your complete response first with all information, and then provide the citations.

RAG Layer Prompt

Output from Search & Generation Layer

We have drafted 3 questions based on the policy documents. Below are the screenshots of the response from search layer.

```
# checking for test question 1
query = test_questions[0]
print('\nQuery :', query)
retrieve_top3withrerank(query,search(query, cache_collection, openai_collection_page_chunks))
```

```
Query : What are the policyholder eligibility requirements ?
Found in cache!

The top 3 results/chunks retrieved from the search layer are below:
```

1 to 3 of 3 entries Filter

index	Documents	Metadata
4	TABLE OF CONTENTS PART I - DEFINITIONS PART II - POLICY ADMINISTRATION Section A – Contract Entire Contract Article 1 Policy Changes Article 2 Policyholder Eligibility Requirements Article 3 Policy Incontestability Article 4 Individual Incontestability Article 5 Information to be Furnished Article 6 Certificates Article 7 Assignments Article 8 Dependent Rights Article 9 Policy Interpretation Article 10 Electronic Transactions Article 11 Section B – Premium Payment Responsibility; Due Dates; Grace Period Article 1 Premium Rates Article 2 Premium Rate Changes Article 3 Premium Amount Article 4 Contributions from Members Article 5 Section C - Policy Termination Failure to Pay Premium Article 1 Termination Rights of the Policyholder Article 2 Termination Rights of The Principal Article 3 Policyholder Responsibility to Members Article 4 Section D - Policy Renewal Renewal Article 1 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS This policy has been updated effective January 1, 2014 GC 6001 TABLE OF CONTENTS, PAGE 1	{'Page No.': 'Page 6', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
0	a. be actively engaged in business for profit within the meaning of the Internal Revenue Code, or be established as a legitimate nonprofit corporation within the meaning of the Internal Revenue Code; and b. make at least the level of premium contributions required for insurance on its eligible Members. The Policyholder must: (1) contribute at least 50% of the required premium for all Members (including disabled Members, if any); and c. if the Member is to contribute part of the premium, maintain the following participation percentages with respect to eligible employees and Dependents, excluding those for whom Proof of Good Health is not satisfactory to The Principal: (1) Employees: - at least 75% of all eligible employees must enroll; (2) Dependents: - maintain a Dependent participation of at least 75% of eligible Dependents; and d. if the Member is to contribute no part of the premium, 100% of eligible employees and Dependents must enroll. Article 4 - Policy Incontestability In the absence of fraud, after this Group Policy has been in force two years, The Principal may not contest its validity except for nonpayment of premium. Article 5 - Individual Incontestability All statements made by any individual insured under this Group Policy will be representations and not warranties. In the absence of fraud, these statements may not be used to contest an insured person's insurance unless: a. the insured person's insurance has been in force for less than two years during the insured's lifetime; and b. the statement is in Written form Signed by the insured person; and This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6003 Section A - Contract, Page 2	{'Page No.': 'Page 17', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
2	PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS Section A - Eligibility Article 1 - Member Life Insurance A person will be eligible for Member Life Insurance on the date the person completes 30 consecutive days of continuous Active Work with the Policyholder as a Member. In no circumstance will a person be eligible for Member Life Insurance under this Group Policy if the person is eligible under any other Group Term Life Insurance policy underwritten by The Principal. Article 2 - Member Accidental Death and Dismemberment Insurance A person will be eligible for Member Accidental Death and Dismemberment Insurance on the latest of: a. the date the person is eligible for Member Life Insurance; or b. the date the person enters a class for which Member Accidental Death and Dismemberment Insurance is provided under this Group Policy; or c. the date Member Accidental Death and Dismemberment Insurance is added to this Group Policy. Article 3 - Dependent Life Insurance A person will be eligible for Dependent Life Insurance on the latest of: a. the date the person is eligible for Member Life Insurance; or b. the date the person first acquires a Dependent; or c. the date the person enters a class for which Dependent Life Insurance is provided under this Group Policy; or d. the date Dependent Life Insurance is added to this Group Policy. This policy has been updated effective January 1, 2014 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS GC 6006 Section A - Eligibility, Page 1	{'Page No.': 'Page 26', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}

```
# checking for test question 2
query = test_questions[1]
print('\nQuery :', query)
retrieve_top3withrerank(query,search(query, cache_collection, openai_collection_page_chunks))
```

```
Query : What is the grace period for the policy premium payments?
Not found in cache. Found in main collection.

The top 3 results/chunks retrieved from the search layer are below:
```

1 to 3 of 3 entries Filter

index	Documents	Metadata
0	Section B - Premiums Article 1 - Payment Responsibility; Due Dates; Grace Period The Policyholder is responsible for collection and payment of all premiums due while this Group Policy is in force. Payments must be sent to the home office of The Principal in Des Moines, Iowa. The first premium is due on the Date of Issue of this Group Policy. Each premium thereafter will be due on the first of each Insurance Month. Except for the first premium, a Grace Period of 31 days will be allowed for payment of premium. "Grace Period" means the first 31-day period following a premium due date. The Group Policy will remain in force until the end of the Grace Period, unless the Group Policy has been terminated by notice as described in PART II, Section C. The Policyholder will be liable for payment of the premium for the time this Group Policy remains in force during the Grace Period. Article 2 - Premium Rates The premium rate(s) for each Member insured for Life Insurance will be: a. Member Life Insurance $0.210 for each $1,000 of insurance in force. b. Member Accidental Death and Dismemberment Insurance $0.025 for each $1,000 of Member Life Insurance in force. c. Dependent Life Insurance $1.46 for each Member insured for Dependent Life Insurance. If the Policyholder has at least two other eligible group insurance policies underwritten by The Principal, as determined by The Principal, the Policyholder may be eligible for a multiple policy discount. Article 3 - Premium Rate Changes The Principal may change a premium rate: a. on any premium due date, if the initial premium rate has then been in force 24 months or more and if Written notice is given to the Policyholder at least 31 days before the date of change; or This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6004 Section B - Premiums, Page 1	{'Page No.': 'Page 20', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
1	Section C - Policy Termination Article 1 - Failure to Pay Premium This Group Policy will terminate at the end of the Grace Period if total premium due has not been received by The Principal before the end of the Grace Period. Failure by the Policyholder to pay the premium within the Grace Period will be deemed notice by the Policyholder to The Principal to discontinue this Group Policy at the end of the Grace Period. Article 2 - Termination Rights of the Policyholder The Policyholder may terminate this Group Policy effective on the day before any premium due date by giving Written notice to The Principal prior to that premium due date. The Policyholder's issuance of a stop-payment order for any amounts used to pay premiums for the Policyholder's coverage will be considered Written notice from the Policyholder. Article 3 - Termination Rights of The Principal The Principal may nonrenew or terminate this Group Policy by giving the Policyholder 31 days advance notice in Writing, if the Policyholder: a. ceases to be actively engaged in business for profit within the meaning of the Internal Revenue Code, or be established as a legitimate nonprofit corporation within the meaning of the Internal Revenue Code; or b. fails to maintain the participation percentages requirements of PART II, Section A with respect to eligible employees, excluding those for whom Proof of Good Health is not satisfactory to The Principal; or c. fails to maintain three or more insured employees under this Group Policy; or d. fails to pay premium in accordance with the requirements of PART II, Section B; or e. has performed an act or practice that constitutes fraud or has made an intentional misrepresentation of material fact under the terms of this Group Policy; or f. does not promptly provide The Principal with information that is reasonably required; or g. fails to perform any of its obligations that relate to this Group Policy. This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6005 Section C - Policy Termination, Page 1	{'Page No.': 'Page 23', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
2	T he Principal may terminate the Policyholder's coverage on any premium due date if the Policyholder relocates to a state where this Group Policy is not marketed, by giving the Policyholder 31 days advanced notice in Writing. Article 4 - Policyholder Responsibility to Members If this Group Policy terminates for any reason, the Policyholder must: a. notify each Member of the effective date of the termination; and b. refund or otherwise account to each Member all contributions received or withheld from Members for premiums not actually paid to The Principal. This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6005 Section C - Policy Termination, Page 2	{'Page No.': 'Page 24', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}

```
# checking for test question 3
query = test_questions[2]
print('\nQuery :', query)
retrieve_top3withrerank(query,search(query, cache_collection, openai_collection_page_chunks))
```

```
Query : How often premium rate changes?
Not found in cache. Found in main collection.

The top 3 results/chunks retrieved from the search layer are below:
```

1 to 3 of 3 entries Filter

index	Documents	Metadata
1	Section B - Premiums Article 1 - Payment Responsibility; Due Dates; Grace Period The Policyholder is responsible for collection and payment of all premiums due while this Group Policy is in force. Payments must be sent to the home office of The Principal in Des Moines, Iowa. The first premium is due on the Date of Issue of this Group Policy. Each premium thereafter will be due on the first of each Insurance Month. Except for the first premium, a Grace Period of 31 days will be allowed for payment of premium. "Grace Period" means the first 31-day period following a premium due date. The Group Policy will remain in force until the end of the Grace Period, unless the Group Policy has been terminated by notice as described in PART II, Section C. The Policyholder will be liable for payment of the premium for the time this Group Policy remains in force during the Grace Period. Article 2 - Premium Rates The premium rate(s) for each Member insured for Life Insurance will be: a. Member Life Insurance $0.210 for each $1,000 of insurance in force. b. Member Accidental Death and Dismemberment Insurance $0.025 for each $1,000 of Member Life Insurance in force. c. Dependent Life Insurance $1.46 for each Member insured for Dependent Life Insurance. If the Policyholder has at least two other eligible group insurance policies underwritten by The Principal, as determined by The Principal, the Policyholder may be eligible for a multiple policy discount. Article 3 - Premium Rate Changes The Principal may change a premium rate: a. on any premium due date, if the initial premium rate has then been in force 24 months or more and if Written notice is given to the Policyholder at least 31 days before the date of change; or This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6004 Section B - Premiums, Page 1	{'Page No.': 'Page 20', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
0	b . on any date the definition of Member or Dependent is changed; and c. on any date the Policyholder's business, as specified on the Policyholder application, is changed; and d. on any date that a schedule of insurance or class of insured Members is changed; and e. on any premium due date, if the Policyholder has been receiving a multiple policy discount rate and the Policyholder drops below the minimum number of coverages to receive such discount rate; and f. on any date the premium contribution required of Members is changed; and g. with respect to Member Life Insurance, on any Policy Anniversary, if the average age, average Scheduled Benefit amount, or the male/female distribution for then insured Members has changed since the last Policy Anniversary; and h. on any Policy Anniversary, if the volume of insurance for then insured Members has increased or decreased by more than 25% since the last Policy Anniversary. If the Policyholder has other group insurance with The Principal, and if life coverage is initially added on a date other than the Policy Anniversary and it is more than six months before the next Policy Anniversary, The Principal reserves the right to change the premium rate on the next Policy Anniversary. Written notice will be given to the Policyholder at least 31 days before the date of change. If the Policyholder agrees to participate in the electronic services program of The Principal and, at a later date elects to withdraw from participation, such withdrawal may result in certain administrative fees being charged to the Policyholder. Article 4 - Premium Amount The amount of premium to be paid on each due date will be determined in these ways: a. Member Life Insurance The total volume of insurance in force will be divided by 1,000. The result will then be multiplied by the premium rate then in effect. b. Member Accidental Death and Dismemberment Insurance The total volume of insurance in force will be divided by 1,000. The result will then be multiplied by the premium rate then in effect. c. Dependent Life Insurance This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6004 Section B - Premiums, Page 2	{'Page No.': 'Page 21', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}
2	The number of Members insured for Dependent Life Insurance will be multiplied by the premium rate then in effect. To ensure accurate premium calculations, the Policyholder is responsible for reporting to The Principal, the following information during the stated time periods: a. Members who are eligible to become insured are to be reported during the month prior to or during the month that coverage becomes effective. b. Members whose coverage has terminated are to be reported within a month of the date coverage terminated. c. Changes in Member insurance class are to be reported within a month of the date that the change in insurance class took place. If a Member is added or a present Member's insurance is increased or terminated on other than the first of an Insurance Month, premium for that Member will be adjusted and applied as if the change were to take place on the first of the next following Insurance Month. Article 5 - Contributions from Members Members are not required to contribute a part of the premium for their Member insurance under this Group Policy. Members are required to contribute a part of the premium for their Dependent's insurance under this Group Policy. This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6004 Section B - Premiums, Page 3	{'Page No.': 'Page 22', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}

Show 25 per page

Below are the screenshots of the response from generation layer.

```
[97] # Question 1
search_from_pdf(test_questions[0])
```

```
Query:
 What are the policyholder eligibility requirements ?
Found in cache!

The top 3 results/chunks retrieved from the search layer are below:

 Top 3 results from search layer :
```

	Documents	Metadata
4	TABLE OF CONTENTS PART I - DEFINITIONS PART II...	{'Page No.': 'Page 6', 'Policy_Name': 'Princip...
0	a. be actively engaged in business for profit ...	{'Page No.': 'Page 17', 'Policy_Name': 'Princi...
2	PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS ...	{'Page No.': 'Page 26', 'Policy_Name': 'Princi...

```
Answer:
The policyholder eligibility requirements typically include being actively engaged in a business for profit. Further details may be available in the
specific sections of the policy documents provided.

Here is the information extracted from the relevant policy documents:

| Policy Name      | Page Number | Eligibility Requirement                           |
|------------------|-------------|---------------------------------------------------|
| Principal Policy | Page 17     | Actively engaged in business for profit           |
| Principal Policy | Page 26     | Individual requirements and rights information   |

Please refer to the Principal Policy document on Page 17 for details on the eligibility requirement of being actively engaged in a business for
profit. Additionally, Page 26 may contain further information on individual requirements and rights.

Citations:
- Policy Name: Principal Policy
- Page Numbers: Page 17, Page 26
```

```
[101] # Question 2
search_from_pdf(test_questions[1])
```

```
Query:
 What is the grace period for the policy premium payments?
Found in cache!

The top 3 results/chunks retrieved from the search layer are below:

 Top 3 results from search layer :
```

	Documents	Metadata
0	Section B - Premiums Article 1 - Payment Respo...	{'Page No.': 'Page 20', 'Policy_Name': 'Princi...
1	Section C - Policy Termination Article 1 - Fai...	{'Page No.': 'Page 23', 'Policy_Name': 'Princi...
2	T he Principal may terminate the Policyholder'...	{'Page No.': 'Page 24', 'Policy_Name': 'Princi...

```
Answer:
The grace period for policy premium payments is 30 days from the due date as per the insurance document found in Section B - Premiums, Article 1. The
policy name is mentioned as "Principal Policy" and the relevant page number is Page 20.

**Response:**
The grace period for policy premium payments is 30 days from the due date as per the "Principal Policy" document on Page 20.

**Citations:**
- Policy Name: Principal Policy
- Page Number: Page 20
```

```
# Question 3
search_from_pdf(test_questions[2])
```

```
Query:
 How often premium rate changes?
Found in cache!

The top 3 results/chunks retrieved from the search layer are below:

 Top 3 results from search layer :
```

	Documents	Metadata
1	Section B - Premiums Article 1 - Payment Respo...	{'Page No.': 'Page 20', 'Policy_Name': 'Princi...
0	b . on any date the definition of Member or De...	{'Page No.': 'Page 21', 'Policy_Name': 'Princi...
2	The number of Members insured for Dependent Li...	{'Page No.': 'Page 22', 'Policy_Name': 'Princi...

```
Answer:
The frequency of premium rate changes varies depending on the insurance policy. To determine how often premium rate changes occur, you should refer to
the specific policy document provided to you. Look for sections related to premium adjustments, rate changes, or renewal terms within the policy
document. These sections typically outline the conditions under which premium rates may change and the frequency of such changes.

If you are unable to find the specific information within the document, you can search for keywords like "premium rate changes," "rate adjustment
frequency," or "premium renewal terms" to locate the relevant sections.

Citations:
- Policy Name: Principal Life Insurance Policy
- Page Number: Page 20

Please refer to the "Section B - Premiums" in the Principal Life Insurance Policy document on Page 20 for detailed information regarding how often
premium rate changes occur.
```

Challenges

These challenges highlight the intricacies involved in managing and optimizing a semantic search system for complex documents such as insurance policies.

1. **Data Quality and Preprocessing**:

- Extracting relevant information from insurance documents, which often contain complex text structures, poses significant challenges.

2. **Effective Chunking Strategies**:
    - Determining the optimal chunk size and overlap to capture meaningful context without losing coherence is difficult and requires careful consideration.

3. **Performance and Query Management**:
    - Enhancing system performance to handle an increased number of documents or users by implementing vector databases and scaling up compute units.
    - Optimizing cache storage to efficiently store and retrieve queries and results, ensuring that the system remains responsive and efficient.

Learnings

These points encapsulate the key learnings from this project.

- **Efficient Document Processing**: Utilizing libraries like pdfplumber for efficient PDF processing and employing suitable data structures for storage is crucial.
- **Semantic Search Optimization**: Fine-tuning semantic search parameters and thresholds is essential for achieving optimal results.
- **Cache Management**: Implementing an effective cache management strategy balances storage and retrieval efficiency.
- **Thorough PDF Analysis and Domain Adaptation**: Robust extraction logic and adapting models to specific domains enhance accuracy, while hybrid retrieval and cost-effective architecture (mix of APIs and local models) further improve performance.

Summary

Mr. HelpMate AI successfully demonstrates the potential of semantic search and AI-powered question answering to transform how users navigate insurance documents.

The project highlighted the value of PDF preprocessing, embedding-based search, and language model fine-tuning.

Link to Github Repo: [dipaksah20/HelpMateAI: RAG based Assistant to help with queries on insurance document]