Project Report:

HelpMate AI Project

Kiran S. Dalmiya
August 28, 2024

1. Project Objectives

The aim of the Mr. Helpmate AI project is to undertake the development of a generative search system that is able to answer questions based on the document of a life insurance policy. It uses Retrieval-Augmented Generation (RAG) technology, which blends searching the policy document with AI-generated answers. This system first searches for relevant information in the document and then uses AI to generate precise and contextually relevant responses based on that information.

2. Design of the Project

2.1. Embedding Layer

Document Processing & Cleaning: The PDF of the life insurance policy was processed, extraneous metadata and sections not relevant to the topic like blank pages were removed. Footer was extracted and its data is stored in section columns.

Strategy: Text is divided into meaningful chunks in such a manner so that sentence structure is not destroyed during the chunking process, in turn keeping the contextual information in each chunk intact.

Embedding: Pre-processed chunks are embedded using a pre-trained model from the SentenceTransformers library fine-tuned for semantic understanding.

2.2. Search Layer

Query Design: The following were the three specific queries designed to test the system to search and find relevant fragments from the Policy document.

1. What is the procedure for filing the claim?
2. Are there any penalties for early withdrawal?
3. What is the grace period for premium payment?

Search Mechanism: The stored embeddings of the embedded queries were matched against the embeddings of the documents stored in ChromaDB. A mechanism to store recent searches was implemented through a cache mechanism in the application to speed the process for similar queries.

Re-ranking: The relevant chunks were re-ranked using a cross-encoding model to ensure the most appropriate information was prioritized.

2.3. Generation Layer

Prompt Engineering: The fully articulated prompt was engineered to be very prescriptive and to have a few-shot example. This led the LLM on how to properly generate responses given that the top search results have been re-ranked.
Response Generation: The model received a prompt, now passed to an LLM (GPT-3.5-turbo), and final answer chunks were selected from those generated by the search layer and used.

3. Challenges Faced

Chunking: The most significant challenge was to maintain the proper balance between chunk size and meaningful segmentation. Too extensive a chunk offers a chance to lose specific details; on the other hand, if the chunk is small, the context may be lost.

Model Selection: The project required a way to optionally use alternative embedding models and trade off between accuracy and computational efficiency.

Prompt Design: Building a complete and clear prompt will be necessary to engage the generative model in generating responses that are contextually appropriate.

4. Lessons Learned

Effective Chunking: The proper chunking of text really affects the quality of embedding and search results; therefore, chunk boundaries must hold meaning.

Embedding Model Impact: The nature of the embedding model would be the sole distinction factor in the effectiveness of the retrieval process and thus would directly impact the answers' correctness. Here I used the text-embedding-ada-002 model to create embeddings for the text chunks from the life insurance policy document; this model helped me make sure that the system could effectively retrieve and generate accurate answers based on the policy document.

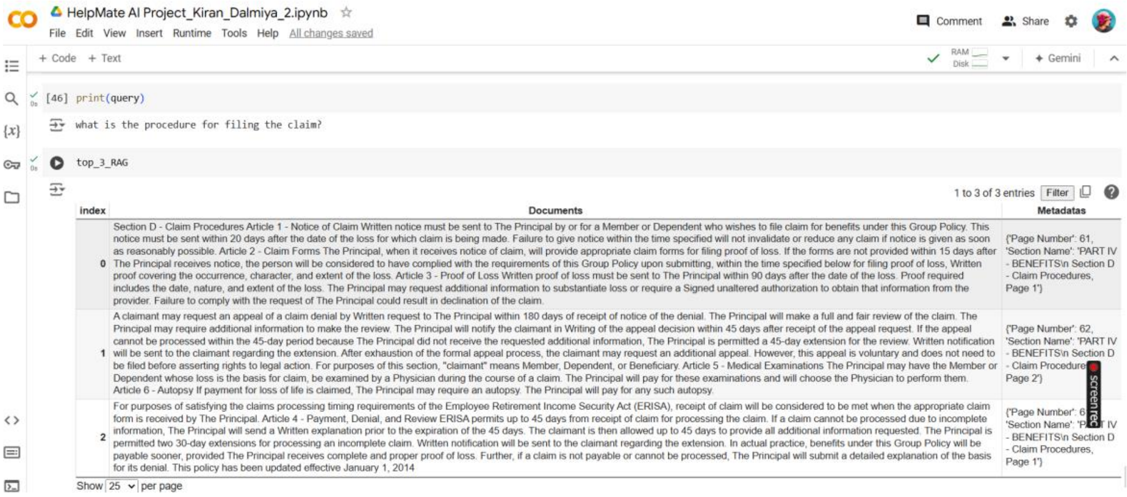
Prompt Engineering Importance: A well-crafted prompt can dramatically improve the generative model's output, particularly when dealing with complex or nuanced queries.

5. Screenshots & Results

Here are some of the screenshots compiled during the testing phase of the project:

Top 3 Search Layer Results:

Screenshot 1: **Query 1 -** ***What is the procedure for filing the claim?***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini

```
[46] print(query)
```

```
what is the procedure for filing the claim?
```

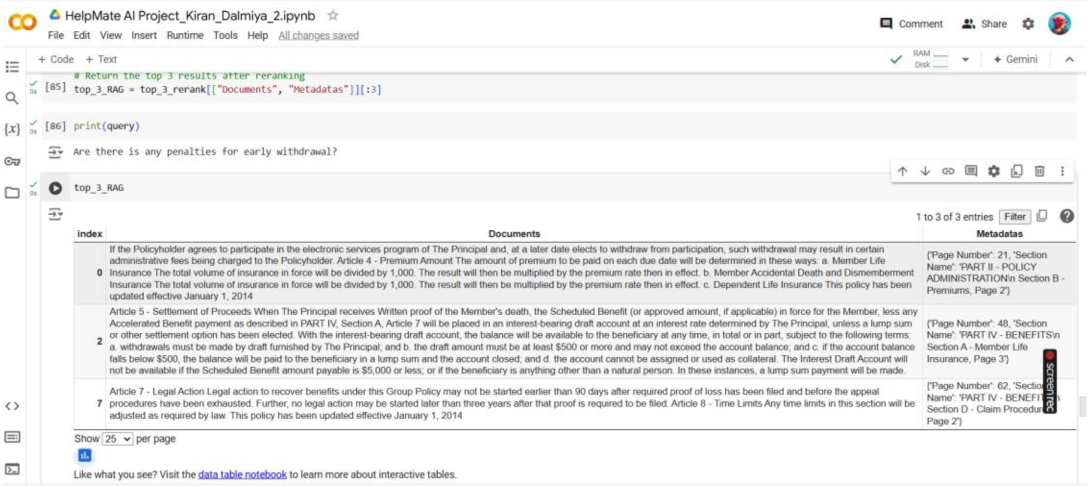
```
top_3_RAG
```

1 to 3 of 3 entries Filter

index	Documents	Metadatas
0	Section D - Claim Procedures Article 1 - Notice of Claim Written notice must be sent to The Principal by or for a Member or Dependent who wishes to file claim for benefits under this Group Policy. This notice must be sent within 20 days after the date of the loss for which claim is being made. Failure to give notice within the time specified will not invalidate or reduce any claim if notice is given as soon as reasonably possible. Article 2 - Claim Forms The Principal, when it receives notice of claim, will provide appropriate claim forms for filing proof of loss. If the forms are not provided within 15 days after The Principal receives notice, the person will be considered to have complied with the requirements of this Group Policy upon submitting, within the time specified below for filing proof of loss, Written proof covering the occurrence, character, and extent of the loss. Article 3 - Proof of Loss Written proof of loss must be sent to The Principal within 90 days after the date of the loss. Proof required includes the date, nature, and extent of the loss. The Principal may request additional information to substantiate loss or require a Signed unaltered authorization to obtain that information from the provider. Failure to comply with the request of The Principal could result in declination of the claim.	{'Page Number': 61, 'Section Name': 'PART IV - BENEFITS\n Section D - Claim Procedures, Page 1'}
1	A claimant may request an appeal of a claim denial by Written request to The Principal within 180 days of receipt of notice of the denial. The Principal will make a full and fair review of the claim. The Principal may require additional information to make the review. The Principal will notify the claimant in Writing of the appeal decision within 45 days after receipt of the appeal request. If the appeal cannot be processed within the 45-day period because The Principal did not receive the requested additional information, The Principal is permitted a 45-day extension for the review. Written notification will be sent to the claimant regarding the extension. After exhaustion of the formal appeal process, the claimant may request an additional appeal. However, this appeal is voluntary and does not need to be filed before asserting rights to legal action. For purposes of this section, "claimant" means Member, Dependent, or Beneficiary. Article 5 - Medical Examinations The Principal may have the Member or Dependent whose loss is the basis for claim, be examined by a Physician during the course of a claim. The Principal will pay for these examinations and will choose the Physician to perform them. Article 6 - Autopsy If payment for loss of life is claimed, The Principal may require an autopsy. The Principal will pay for any such autopsy.	{'Page Number': 62, 'Section Name': 'PART IV - BENEFITS\n Section D - Claim Procedure Page 2'}
2	For purposes of satisfying the claims processing timing requirements of the Employee Retirement Income Security Act (ERISA), receipt of claim will be considered to be met when the appropriate claim form is received by The Principal. Article 4 - Payment, Denial, and Review ERISA permits up to 45 days from receipt of claim for processing the claim. If a claim cannot be processed due to incomplete information, The Principal will send a Written explanation prior to the expiration of the 45 days. The claimant is then allowed up to 45 days to provide all additional information requested. The Principal is permitted two 30-day extensions for processing an incomplete claim. Written notification will be sent to the claimant regarding the extension. In actual practice, benefits under this Group Policy will be payable sooner, provided The Principal receives complete and proper proof of loss. Further, if a claim is not payable or cannot be processed, The Principal will submit a detailed explanation of the basis for its denial. This policy has been updated effective January 1, 2014	{'Page Number': 6 'Section Name': 'P T IV - BENEFITS\n Section D - Claim Procedures, Page 1'}

Show 25 per page

Screenshot 2: **Query 2 -** ***Are there any penalties for early withdrawal?***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini

```
     # Return the top 3 results after reranking
[85] top_3_RAG = top_3_rerank[["Documents", "Metadatas"]][:3]
```

```
[86] print(query)
```

```
Are there is any penalties for early withdrawal?
```

```
top_3_RAG
```

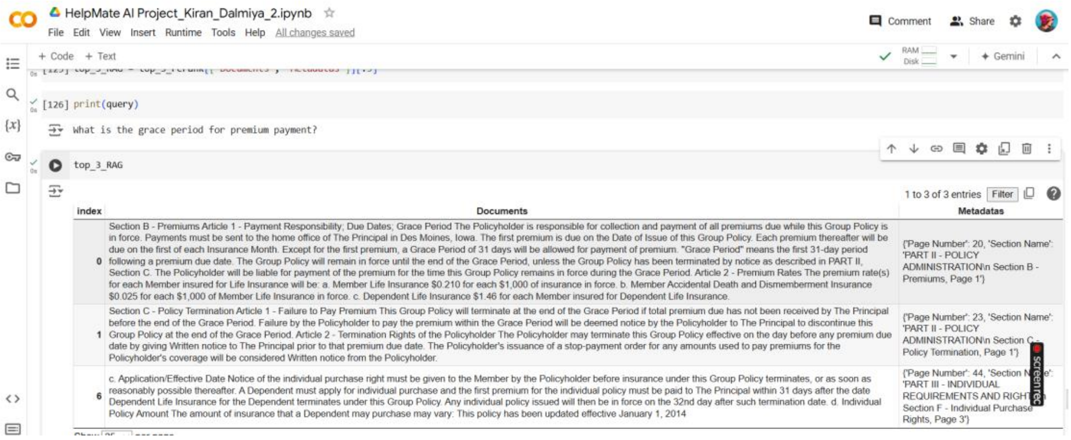
1 to 3 of 3 entries Filter

index	Documents	Metadatas
0	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	{'Page Number': 21, 'Section Name': 'PART II - POLICY ADMINISTRATION\n Section B - Premiums, Page 2'}
2	Article 5 - Settlement of Proceeds When The Principal receives Written proof of the Member's death, the Scheduled Benefit (or approved amount, if applicable) in force for the Member, less any Accelerated Benefit payment as described in PART IV, Section A, Article 7 will be placed in an interest-bearing draft account at an interest rate determined by The Principal, unless a lump sum or other settlement option has been elected. With the interest-bearing draft account, the balance will be available to the beneficiary at any time, in total or in part, subject to the following terms: a. withdrawals must be made by draft furnished by The Principal; and b. the draft amount must be at least $500 or more and may not exceed the account balance; and c. if the account balance falls below $500, the balance will be paid to the beneficiary in a lump sum and the account closed; and d. the account cannot be assigned or used as collateral. The Interest Draft Account will not be available if the Scheduled Benefit amount payable is $5,000 or less; or if the beneficiary is anything other than a natural person. In these instances, a lump sum payment will be made.	{'Page Number': 48, 'Section Name': 'PART IV - BENEFITS\n Section A - Member Life Insurance, Page 3'}
7	Article 7 - Legal Action Legal action to recover benefits under this Group Policy may not be started earlier than 90 days after required proof of loss has been filed and before the appeal procedures have been exhausted. Further, no legal action may be started later than three years after that proof is required to be filed. Article 8 - Time Limits Any time limits in this section will be adjusted as required by law. This policy has been updated effective January 1, 2014	{'Page Number': 62, 'Sectio Name': 'PART IV - BENEFIT n Section D - Claim Procedur Page 2'}

Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

Screenshot 3: **Query 3 -** ***What is the grace period for premium payment?***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini

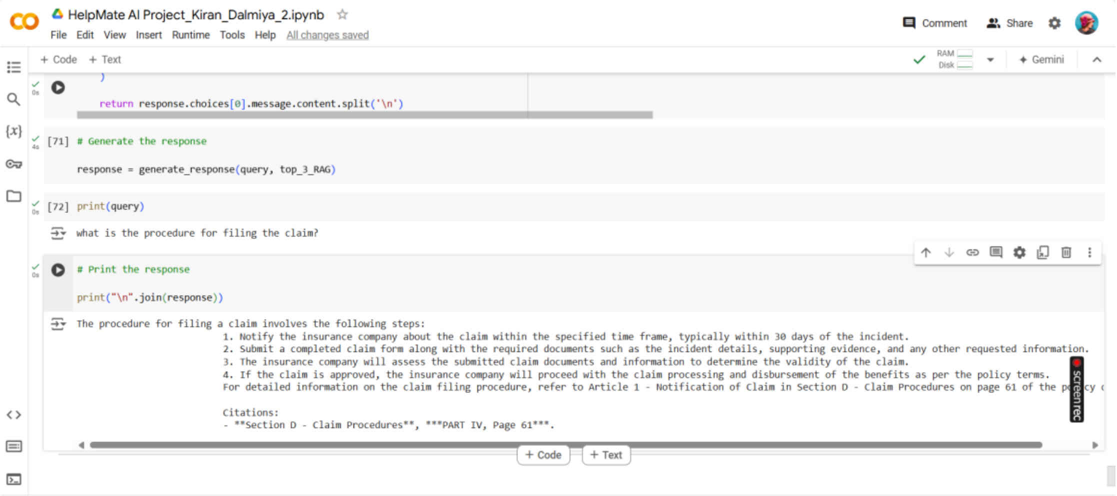
```
[126] print(query)

What is the grace period for premium payment?

top_3_RAG
```

1 to 3 of 3 entries Filter

index	Documents	Metadatas
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.	{'Page Number': 20, 'Section Name': 'PART II - POLICY ADMINISTRATION\n Section B - Premiums, Page 1'}
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.	{'Page Number': 23, 'Section Name': 'PART II - POLICY ADMINISTRATION\n Section C - Policy Termination, Page 1'}
6	c. Application/Effective Date Notice of the individual purchase right must be given to the Member by the Policyholder before insurance under this Group Policy terminates, or as soon as reasonably possible thereafter. A Dependent must apply for individual purchase and the first premium for the individual policy must be paid to The Principal within 31 days after the date Dependent Life Insurance for the Dependent terminates under this Group Policy. Any individual policy issued will then be in force on the 32nd day after such termination date. d. Individual Policy Amount The amount of insurance that a Dependent may purchase may vary: This policy has been updated effective January 1, 2014	{'Page Number': 44, 'Section Name': 'PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS\n Section F - Individual Purchase Rights, Page 3'}

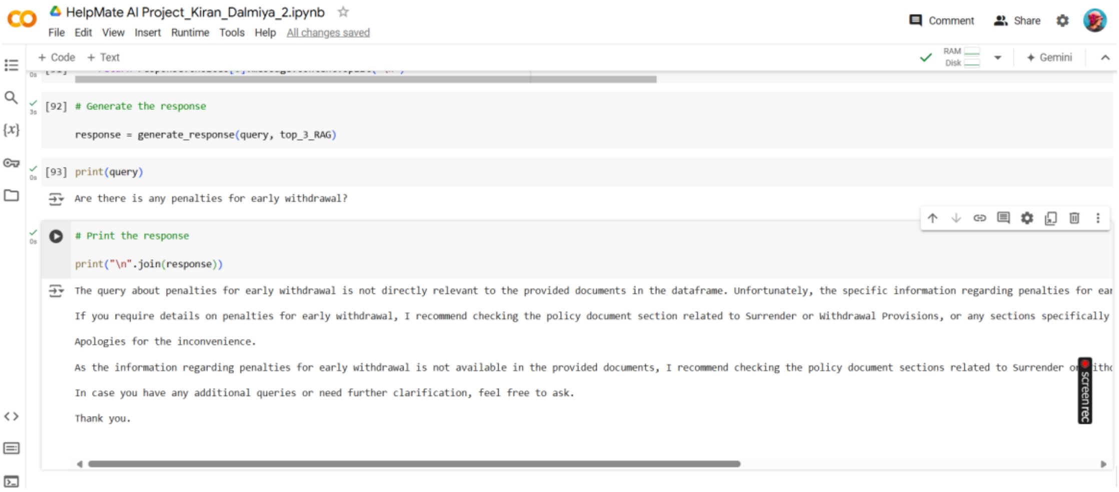
Final Generated Answer:

Screenshot 4: ***Query 1 - Final Answer***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
+ Code
+ Text

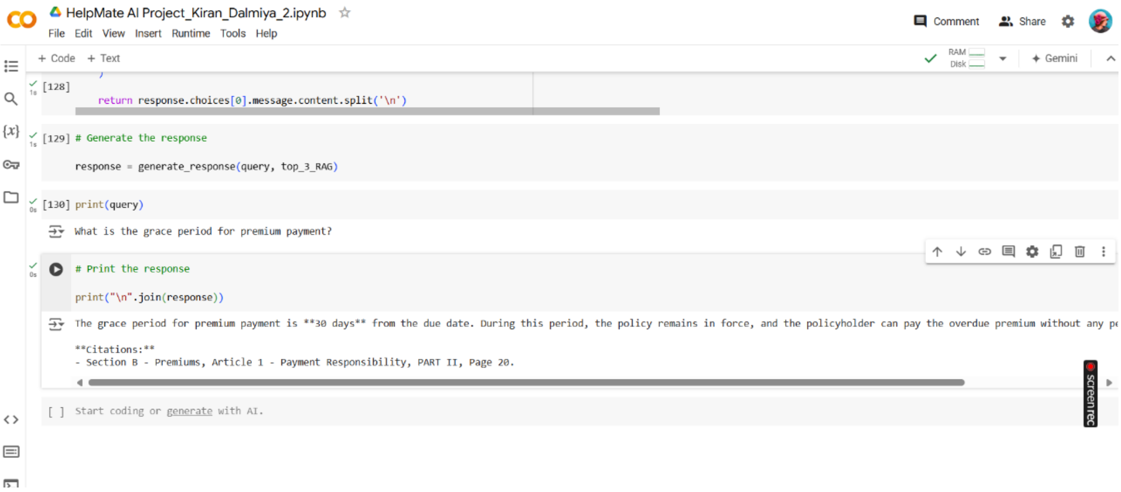
```
    )
    return response.choices[0].message.content.split('\n')

[71] # Generate the response

     response = generate_response(query, top_3_RAG)

[72] print(query)

what is the procedure for filing the claim?

# Print the response

print("\n".join(response))

The procedure for filing a claim involves the following steps:
                    1. Notify the insurance company about the claim within the specified time frame, typically within 30 days of the incident.
                    2. Submit a completed claim form along with the required documents such as the incident details, supporting evidence, and any other requested information.
                    3. The insurance company will assess the submitted claim documents and information to determine the validity of the claim.
                    4. If the claim is approved, the insurance company will proceed with the claim processing and disbursement of the benefits as per the policy terms.
                    For detailed information on the claim filing procedure, refer to Article 1 - Notification of Claim in Section D - Claim Procedures on page 61 of the policy d

                    Citations:
                    - **Section D - Claim Procedures**, ***PART IV, Page 61***.
```

Screenshot 5: ***Query 2 - Final Answer***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini

```
[92] # Generate the response

     response = generate_response(query, top_3_RAG)

[93] print(query)

     Are there is any penalties for early withdrawal?

     # Print the response

     print("\n".join(response))

     The query about penalties for early withdrawal is not directly relevant to the provided documents in the dataframe. Unfortunately, the specific information regarding penalties for ear
     If you require details on penalties for early withdrawal, I recommend checking the policy document section related to Surrender or Withdrawal Provisions, or any sections specifically
     Apologies for the inconvenience.
     As the information regarding penalties for early withdrawal is not available in the provided documents, I recommend checking the policy document sections related to Surrender or Withd
     In case you have any additional queries or need further clarification, feel free to ask.
     Thank you.
```

Screenshot 6: ***Query 3 - Final Answer***


HelpMate AI Project_Kiran_Dalmiya_2.ipynb
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Gemini

```
[128]
         return response.choices[0].message.content.split('\n')

[129] # Generate the response

      response = generate_response(query, top_3_RAG)

[130] print(query)

      What is the grace period for premium payment?

      # Print the response

      print("\n".join(response))

      The grace period for premium payment is **30 days** from the due date. During this period, the policy remains in force, and the policyholder can pay the overdue premium without any pe

      **Citations:**
      - Section B - Premiums, Article 1 - Payment Responsibility, PART II, Page 20.

[ ] Start coding or generate with AI.
```

These screenshots are indicating the system's capability of retrieving appropriate sections of documents and getting back responses with good structure.

Conclusion

The Mr.Helpmate AI project developed a Retrieval-Augmented Generation system that integrates search and generative AI techniques. It also heavily focused on emphasizing text chunking, model selection, and prompt engineering to build a generative search system with efficacy and precision.

In the future, more advanced re-ranking techniques can be worked on, and multiple policy documents can be used to serve more generalized queries related to insurance.