

HelpMate AI

Semantic Search and Question Answering for Insurance Documents

Prepared by: Mohit Kumar Dubey

Problem Statement

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 you need to do in each of the layers, and the possible experimentations that you can perform based on various choices.

- **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 you try out various strategies and compare their performances. Another important aspect in the embedding layer is the choice of the embedding model. You can choose to embed your chunks using the OpenAI embedding model or any model from the SentenceTransformers library on HuggingFace.
- **The Search Layer:** Here, you first need to design at least 3 queries against which you will test your system. You 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, you need to embed the queries and search your ChromaDB vector database against each of these queries. Implementing a cache mechanism is also mandatory. Finally, you need to implement the re-ranking block, and for this you can choose from a range of cross-encoding models on HuggingFace.
- **The Generation Layer:** In the generation layer, the final prompt that you 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. You may also choose to provide some few-shot examples in an attempt to improve the LLM output. Make sure you try out different strategies and models in each layer, and you might be surprised with the variety of top answers retrieved and generated by your system.

Solution

Mr. HelpMate AI addresses this challenge by combining semantic search with powerful language models to deliver accurate and context-aware answers to user queries from within insurance documents.

Core Features:

- PDF processing and text extraction
- Semantic search to retrieve relevant insurance document sections
- AI-powered question answering to provide direct answers and citations.

Benefits:

Streamlines information retrieval, saves time for policyholders and agents, and potentially enhances decision-making based on policy terms.

Problem Statement:

Insurance documents often present significant challenges due to their length, complex terminology, and diverse structural formats. This complexity hinders users' ability to efficiently locate and understand the specific information they require.

Objectives:

- **Text Extraction:** Develop a system capable of accurately and efficiently extracting text from insurance PDFs, overcoming challenges such as varying layouts, formatting, and OCR errors.
- **Semantic Search:** Implement a semantic search approach that can effectively retrieve relevant information from insurance documents by understanding the underlying meaning and context of queries, rather than relying solely on keyword matching.
- **Question-Answering:** Create a question-answering component that leverages natural language processing to provide precise answers and corresponding citations from insurance documents, addressing user queries in a natural and informative manner.

Target Audience:

The target audience for this research includes insurance policyholders, insurance agents, underwriters, and claims adjusters.

Scope and Requirements

Features

- PDF Preprocessing: (pdfplumber library)
- Semantic Search: ChromaDB, Gemini/ OpenAI embeddings
- Question Answering: LLM (Gemini/ OpenAI language models)
- Potential Additional Features: Table extraction and reformatting, answer summarization

Technology Stack

- Python
- pdfplumber: PDF parsing
- Gemini/ OpenAI: Text embeddings, language model access
- ChromaDB: Semantic search database with embedding storage
- Sentence Transformers: Cross-encoder for result re-ranking (optional)
- Other Libraries: pandas for JSON parsing, pandas for DataFrame manipulations

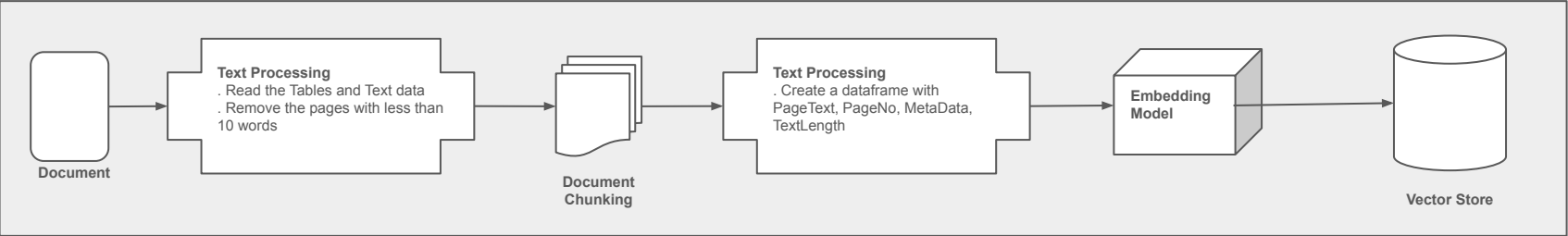
System Design

It consist of three layers

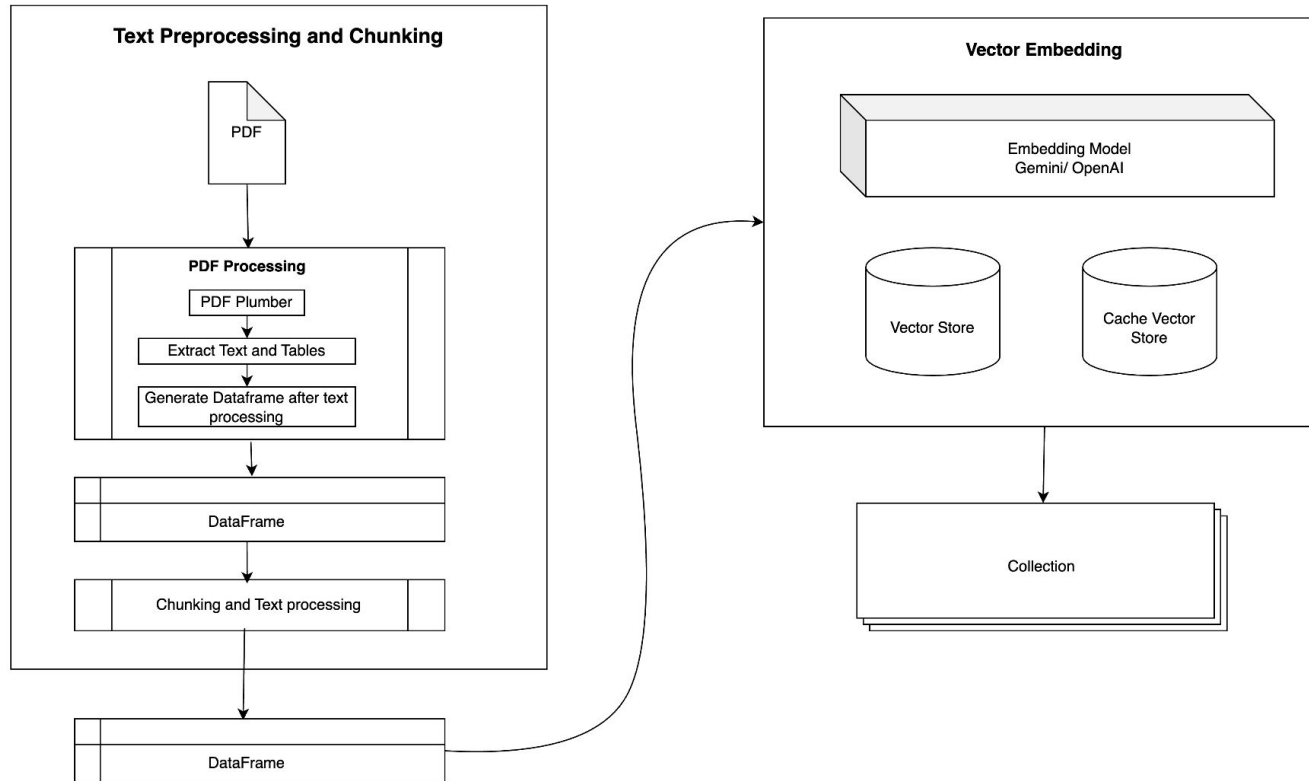
- **Embedding layer:** Text Processing with Embedding and storing to Vector Store
- **Semantic Search and Ranking layer:** Query Embedding and searching in Vector cache and Vector store
- **Generative Search layer:** Generate Search layer - Generate response from LLM based on query prompt and search results

Embedding layer

Text Processing with Embedding and storing to Vector Store

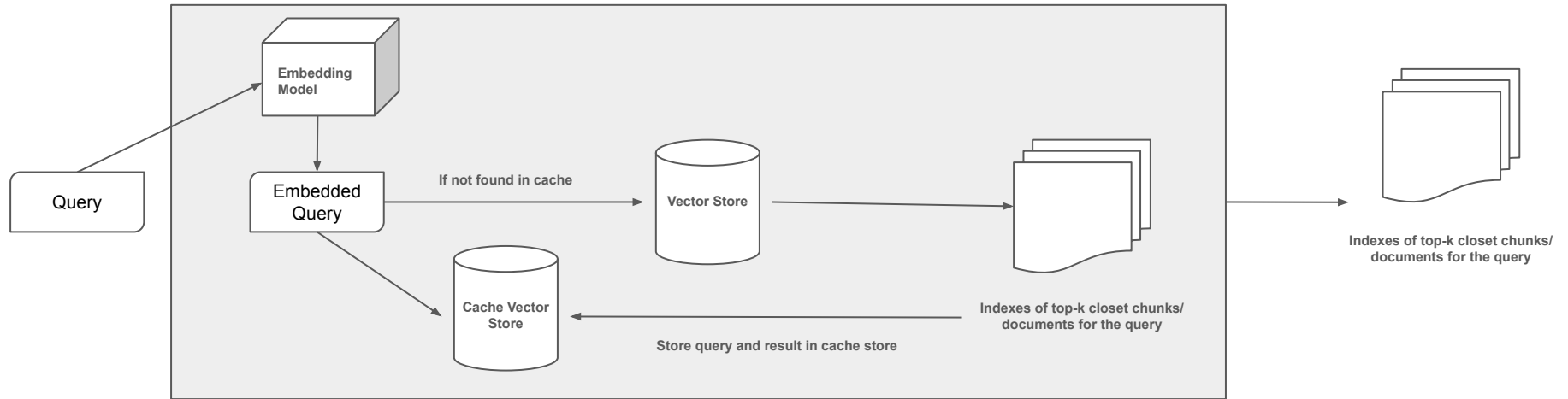


Text Processing with Embedding and storing to Vector Store



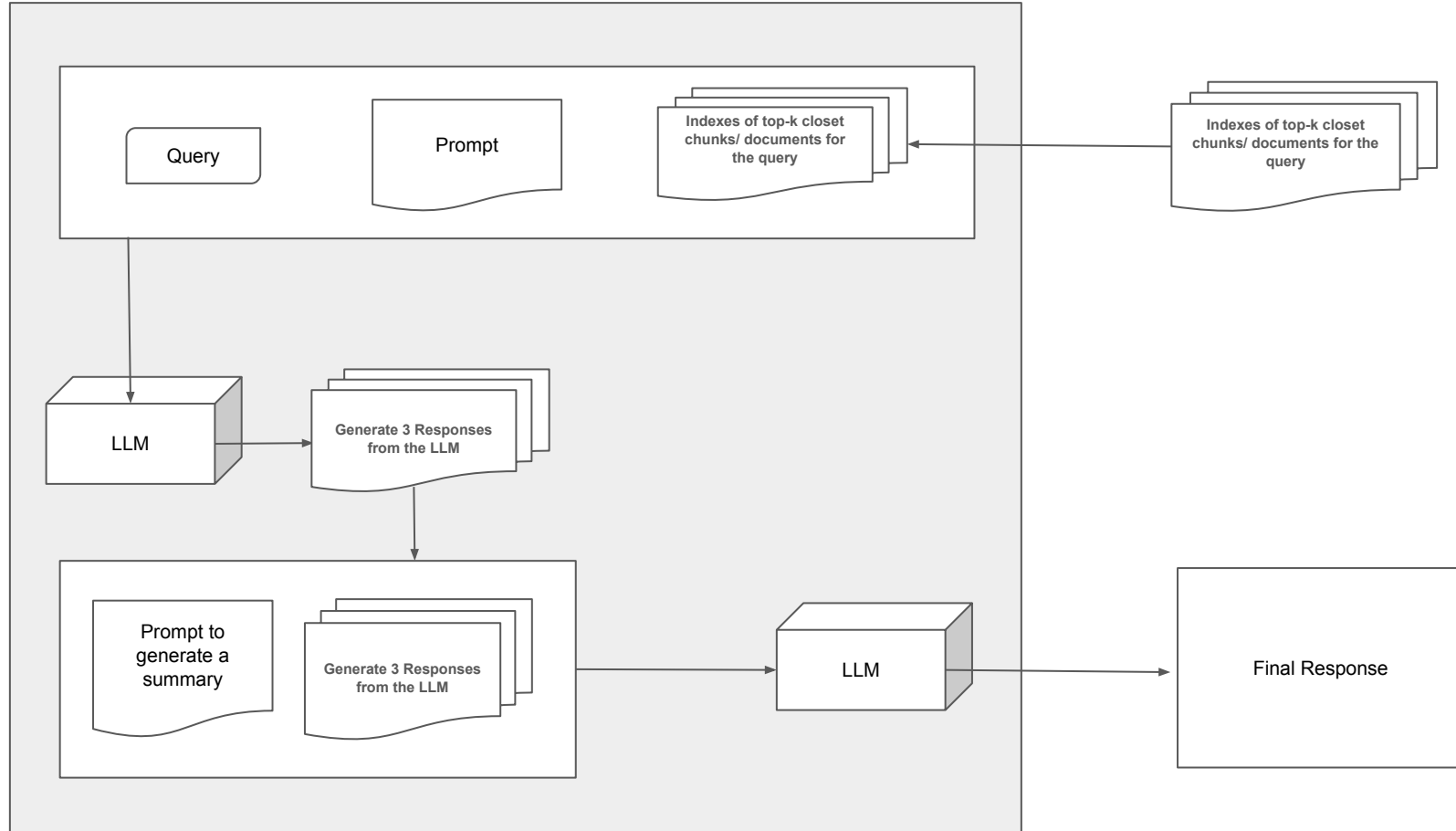
Semantic Search and Ranking layer

Query Embedding and searching in Vector cache and Vector store



Generative Search layer

Generate Search layer - Generate response from LLM based on query prompt and search results



Challenges and Lessons learned

Challenges

- Complex PDF Structures: Handling diverse PDF formats.
- Insurance-Specific Terminology: Adapting models to insurance vocabulary.
- Answer Granularity: Finding the ideal balance in result length.
- API Costs: Optimizing usage for cost-effectiveness.

Lessons Learned

- Thorough PDF Analysis: Pays off in robust extraction logic.
- Domain Adaptation: Crucial for language model accuracy.
- Hybrid Retrieval: Can enhance certain query types.
- Cost-Effective Architecture: Exploring a mix of APIs and local models.

Conclusion

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.

Future Work

- Enhanced User Interface
- Multi-Document QA
- Policy Comparison

Query and Results

#1 What are the defaulter cases in Policy?

- Semantic Search Result

```
[86]: query = "What are the defaulter cases in this policy?"
      final_response, search_results = gemini_pipeline_with_original_results(query)
```

```
[87]: search_results
```

[87]:	Documents	Metadatas
		{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 50'}
2	Failure to give Written proof within the time specified will not invalidate or reduce any claim if Written proof is given as soon as reasonably possible. Further, if a death benefit is paid under this section of the Group Policy, it will be in place of all other Member Life Insurance benefits provided under this Group Policy. f Termination (Premium Waiver) This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6013 Section A - Member Life Insurance, Page 5	
4	The naming of a new beneficiary in an application for individual purchase under PART III, Section F, Article 1, will be treated as a beneficiary change request under this Group Policy. b If a Member's terminated insurance is reinstated, his or her beneficiary will be as recorded on the date of termination. If a Member is insured under this Group Policy on its Date of Issue and this insurance replaces insurance in force on the day immediately before the Date of Issue, the beneficiary named in such replaced insurance and recorded by the Policyholder or The Principal will be the beneficiary under this Group Policy until a new beneficiary is named. Article 4 - Facility of Payment If any of the below occur, benefits will be paid as stated. All such payments will discharge The Principal to the full extent of those payments. a If a beneficiary is found guilty of the Member's death, such beneficiary may be disqualified from receiving any benefit due. Payment may then be made to any contingent beneficiary or to the executor or administrator of the Member's estate. b Any benefit due a beneficiary who dies before the Member's death will be paid in equal shares to the Member's surviving beneficiaries. This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6013 Section A - Member Life Insurance, Page 2	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 47'}
6	(3) The policy will be on one of the forms, other than term insurance, then issued by The Principal to persons in the risk class to which the Member belongs on the individual policy's effective date. (4) Premium will be based on the Member's age and the standard rate of The Principal for the policy form to be issued. b Purchase Qualification A Member will qualify for individual purchase if insurance under this Group Policy terminates and: (1) the Member's total Life Insurance, or any portion of it, terminates because he or she ends Active Work or ceases to be in a class eligible for insurance; or (2) after the Member has been continuously insured under this Group Policy for at least five years, his or her total Member Life Insurance terminates because this Group Policy terminates or is amended to exclude the Member's insurance class; or (3) the Member's Coverage During Disability as described in PART IV, Section A, ceases because Total Disability ends and he or she does not return to Active Work within 31 days; or (4) the Member's Accelerated Benefits Premium Waiver Period as described in PART IV, Section A, ceases and he or she does not qualify for Coverage During Disability. 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	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 42'}

#1 What are the defaulter cases in Policy?

- Generative Search Result

```
[89]: display(Markdown(final_response))
```

Query:

What are the defaulter cases in this policy?

Gemini Response:

The provided documents do not explicitly define "defaulter cases" in the traditional sense, such as a failure to pay premiums. However, they outline situations where benefits might be affected or denied.

- **Beneficiary Disqualification:** If a beneficiary is found guilty of causing the Member's death, they may be disqualified from receiving benefits. These benefits might then be paid to a contingent beneficiary, the Member's estate, or as directed by the Policyholder. **[Principal-Sample-Life-Insurance-Policy, Page 47]**
- **Beneficiary Death Before Member:** If the designated beneficiary dies before the Member, benefits will be divided equally among the Member's surviving beneficiaries. **[Principal-Sample-Life-Insurance-Policy, Page 47]**

Recommendations:

- **Review the full policy document:** Pay close attention to sections related to "beneficiary" or "payment of benefits" for a comprehensive understanding of potential disqualifications.
- **Contact the insurance provider:** Reach out to the insurer directly for clarification on the policy's definition of "defaulter cases" and any other specific concerns.

Citations:

- Document Name: Principal-Sample-Life-Insurance-Policy
- Page Numbers: 47

#2 What are the accidental death in this policy?

- Semantic Search Result

```
[90]: query = "What are the accidental death in this policy?"
      final_response, search_results = gemini_pipeline_with_original_results(query)
```

```
[91]: search_results
```

	Documents	Metadatas
6	Section A - Member Life Insurance Schedule of Insurance Article 1 Death Benefits Payable Article 2 Beneficiary Article 3 Facility of Payment Article 4 Settlement of Proceeds Article 5 Member Life Insurance - Coverage During Disability Article 6 Accelerated Benefits Article 7 Section B - Member Accidental Death and Dismemberment Insurance Schedule of Insurance Article 1 Benefit Qualification Article 2 Benefits Payable Article 3 Seat Belt Benefit Article 4 Loss of Use or Paralysis Benefit Article 5 Loss of Speech and/or Hearing Benefit Article 6 Repatriation Benefit Article 7 Educational Benefit Article 8 Limitations Article 9 Section C - Dependent Life Insurance Schedule of Insurance Article 1 Death Benefits Payable Article 2 Beneficiary Article 3 Section D - Claim Procedures Notice of Claim Article 1 Claim Forms Article 2 Proof of Loss Article 3 Payment, Denial and Review Article 4 Medical Examinations Article 5 Autopsy Article 6 Legal Action Article 7 Time Limits Article 8 This policy has been updated effective January 1, 2014 GC 6001 TABLE OF CONTENTS, PAGE 3	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 8'}
0	Section B - Member Accidental Death and Dismemberment Insurance Article 1 - Schedule of Insurance Subject to the Effective Date provisions of PART III, Section B, and the qualifying provisions of this Section B, the Scheduled Benefit for an insured Member will be based on his or her class: Class *Scheduled Benefit ALL MEMBERS \$10,000 *The Scheduled Benefit is subject to the Proof of Good Health requirements as shown in PART III, Section B, Article 1. Because of the Proof of Good Health requirements, the amount of insurance approved by The Principal may be different than the Scheduled Benefit. If the approved amount of insurance is different than the Scheduled Benefit, the approved amount will apply. For the age(s) shown below, the amount of a Member's insurance will be the percentage of the Scheduled Benefit (or approved amount, if applicable) as shown below. Age % of Scheduled Benefit (or approved amount, whichever applies) Age 70 but less than age 75 65% Age 75 and over 45% Article 2 - Benefit Qualification To qualify for benefit payment, all of the following must occur: a. the Member must be injured while insured for Member Accidental Death and Dismemberment Insurance under this Group Policy; and b. the injury must be through external, violent, and accidental means; and c	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 53'}
5	d Effective Dates and Premium Waiver Coverage During Disability will be effective for a qualified Member on the earliest of: (1) the date nine months after the date the Member becomes Totally Disabled; or (2) the date one month after the date the Member becomes ADL Disabled; or (3) the date the Member dies. Premium will not be charged for Member Life and Member Accidental Death and Dismemberment Insurance and Dependent Life Insurance while the Member's Coverage During Disability is in force. e Benefits Payable If death occurs while Coverage During Disability is in force, The Principal will pay the Member's beneficiary the Member Life Insurance benefit amount that would have been paid had the Member remained insured under the Schedule of Insurance in force on the date Total Disability began. Such benefit shall be subject to any reduction provided under the Schedule of Insurance. However, NO BENEFIT WILL BE PAID if Written proof of ADL Disability or Total Disability is not sent to The Principal within one year of the date ADL Disability or Total Disability began. Failure to give Written proof within the time specified will not invalidate or reduce any claim if Written proof is given as soon as reasonably possible. Further, if a death benefit is paid under this section of the Group Policy, it will be in place of all other Member Life Insurance benefits provided under this Group Policy	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 50'}

#2 What are the accidental death in this policy?

- Generative Search Result

Query:

What are the accidental death in this policy?

Gemini Response:

This policy offers Accidental Death and Dismemberment (AD&D) insurance to members, with a scheduled benefit of \$10,000. The actual amount of insurance may vary based on the member's age, health, and class. For members aged 70 or older, the benefit is reduced: 65% for those aged 70-74 and 45% for those 75 and older. To qualify for AD&D benefits, the member must be injured while insured under the policy, and the injury must be caused by external, violent, and accidental means.

Citations:

- **Document Name:** Principal-Sample-Life-Insurance-Policy
- **Page Numbers:** 8, 53

#3 What are the benefits of taking policy in young age?

- Semantic Search Result

```
[96]: query = "What are the benifits of taking the policy in young age?"
      final_response, search_results = gemini_pipeline_with_original_results(query)

[97]: search_results
```

[97]:	Documents	Metadatas
1	<p>If the approved amount of insurance is different than the Scheduled Benefit, the approved amount will apply. For the age(s) shown below, the amount of a Member's insurance will be the percentage of the Scheduled Benefit (or approved amount, if applicable) as shown below. Age % of Scheduled Benefit (or approved amount, whichever applies) Age 70 but less than age 75 65% Age 75 and over 45% Article 2 - Death Benefits Payable If a Member dies while insured for Member Life Insurance under this Group Policy, The Principal will pay his or her beneficiary the Scheduled Benefit (or approved amount, if applicable) in force on the date of death, less any Accelerated Benefit payment as described in PART IV, Section A, Article 7. However, if a beneficiary is suspected or charged with the This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6013 Section A - Member Life Insurance, Page 1</p>	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 46'}
7	<p>Section B - Member Accidental Death and Dismemberment Insurance Article 1 - Schedule of Insurance Subject to the Effective Date provisions of PART III, Section B, and the qualifying provisions of this Section B, the Scheduled Benefit for an insured Member will be based on his or her class: Class *Scheduled Benefit ALL MEMBERS \$10,000 *The Scheduled Benefit is subject to the Proof of Good Health requirements as shown in PART III, Section B, Article 1. Because of the Proof of Good Health requirements, the amount of insurance approved by The Principal may be different than the Scheduled Benefit. If the approved amount of insurance is different than the Scheduled Benefit, the approved amount will apply. For the age(s) shown below, the amount of a Member's insurance will be the percentage of the Scheduled Benefit (or approved amount, if applicable) as shown below. Age % of Scheduled Benefit (or approved amount, whichever applies) Age 70 but less than age 75 65% Age 75 and over 45% Article 2 - Benefit Qualification To qualify for benefit payment, all of the following must occur: a. the Member must be injured while insured for Member Accidental Death and Dismemberment Insurance under this Group Policy; and b. the injury must be through external, violent, and accidental means; and c</p>	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 53'}
3	<p>Section C - Dependent Life Insurance Article 1 - Schedule of Insurance Subject to the Effective Date provisions of PART III, Section B, the Scheduled Benefit for an insured Dependent will be based on the status of the Dependent: Class ALL MEMBERS Dependent *Scheduled Benefit Spouse 5,000Children(ageatdeath)0daysoldbutlessthan6months1,000 6 months and older \$2,000 *The Scheduled Benefit is subject to the Proof of Good Health requirements as shown in PART III, Section B, Article 1. Because of the Proof of Good Health requirements, the amount of insurance approved by The Principal may be different than the Scheduled Benefit. If the approved amount of insurance is different than the Scheduled Benefit, the approved amount will apply. In no event will a Dependent's Scheduled Benefit be more than 50% of the Member's Scheduled Benefit amount. If a Member elects a Dependent Life benefit in excess of 50% of the Member's Scheduled Benefit amount, the Dependent will be given the highest amount available, not to exceed 50%. Article 2 - Death Benefits Payable If a Dependent dies while insured for Dependent Life Insurance under this Group Policy, The Principal will pay the Scheduled Benefit (or approved amount, if applicable) in force for that Dependent on the date of death</p>	{'Document Name': 'Principal-Sample-Life-Insurance-Policy', 'Page No.': 'Page 59'}

#3 What are the benefits of taking policy in young age?

- Generative Search Result

Query:

What are the accidental death in this policy?

Gemini Response:

This policy offers Accidental Death and Dismemberment (AD&D) insurance to members, with a scheduled benefit of \$10,000. The actual amount of insurance may vary based on the member's age, health, and class. For members aged 70 or older, the benefit is reduced: 65% for those aged 70-74 and 45% for those 75 and older. To qualify for AD&D benefits, the member must be injured while insured under the policy, and the injury must be caused by external, violent, and accidental means.

Citations:

- **Document Name:** Principal-Sample-Life-Insurance-Policy
- **Page Numbers:** 8, 53