Project Report: Mr.HelpMate AI- Generative Search System for Insurance Policies

1. Introduction

Project Title	Mr.HelpMate AI: Retrieval-Augmented Generation (RAG) System for Insurance Policies
Author	**Kashish Sharma**
Date	29 October 2025
Data Source	Principal-Sample-Life-Insurance-Policy.pdf

1.1. Project Objectives

The primary goal of the Mr.HelpMate AI project was to overcome the inefficiency and complexity associated with manually searching lengthy insurance policy documents. This was achieved by developing a robust, three-layer **Retrieval-Augmented Generation (RAG)** system capable of:

1. **Accurately extracting relevant policy details** from the document.
2. **Understanding natural language queries** (Q&A).
3. **Generating concise, contextually relevant, and mandatory cited answers** to mitigate LLM hallucination.
4. Implementing all mandatory system components: **Optimal Chunking, Caching, and Re-ranking** to ensure high search quality and system efficiency.

2. Table of Contents

3. System Design and Architecture

3.1. Three-Layer Architecture Overview

The solution is built on a standard RAG framework, optimized at each stage using specific models and algorithms:

1. **Embedding (Indexing):** Converts policy text into machine-readable **vector embeddings**.
2. **Search (Retrieval):** Efficiently finds the most relevant chunks using semantic search, a **cache**, and a mandatory **re-ranker**.
3. **Generation (Synthesis):** Uses an advanced LLM (**GPT-3.5-Turbo**) guided by an exhaustive prompt to synthesize the final, **cited answer**.

3.2. RAG System Workflow

The complete RAG system workflow is visualized below, illustrating the data flow from query to final answer.

---

4. Implementation Details

4.1. Embedding Layer: Data Processing and Chunking

- **PDF Processing:** Used **pdfplumber** with **custom Python logic** to extract text, ensuring tables and complex structures were preserved as structured strings.
- **Optimal Chunking:** Implemented a **Custom Fixed-Size Character Splitter** with a size of **512 characters** and an **overlap of 128 characters**, adhering to constraints while ensuring contextual flow between chunks.
- **Embedding Model:** Used **all-MiniLM-L6-v2** from sentence-transformers for creating the vector embeddings.
- **Vector Store: ChromaDB** was used as the persistent vector database.

4.2. Search Layer: Vector Store, Caching, and Re-ranking

- **Mandatory Caching:** A dedicated ChromaDB cache collection was used with a distance threshold (CACHE_THRESHOLD = 0.05). This check occurs before the main search, optimizing latency and cost for repeat queries.

- **Mandatory Re-ranking:** A **Cross-Encoder Model (cross-encoder/ms-marco-MiniLM-L-6-v2)** was implemented to refine the Top 10 initial results, selecting the **Top 3** most contextually relevant chunks for the LLM.

4.3. Generation Layer: Prompt Engineering and Live Generation

- **LLM: OpenAI's GPT-3.5-Turbo** was used for final answer synthesis via a **live API call** (no mocking), with a low temperature (0.0) to promote factual retrieval.
- **Quality of Prompt:** The system prompt strictly enforced two rules to ensure trustworthiness:
    1. Answer **only** from the provided context.
    2. Provide a **mandatory citation** for every fact used in the format ``.

---

5. Results and Validation

The system was run against three self-designed, policy-specific queries.

5.1. Query 1: Life Insurance Termination Conditions

Query: List the three specific conditions that will cause a Member's Life Insurance to terminate?

Screenshot 1A: Search Layer Output (Top 3 Reranked Chunks)

Rank	Page_Source	Relevance_Score	Chunk_Text
1	Page 42	2.8675	n 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...
2	Page 36	2.8641	A Member's insurance under this Group Policy for a Dependent will terminate on the earliest of: a. the date his or her Member Life Insurance ceases; or b. the date Dependent Life I...
3	Page 35	2.7025	Section C - Individual Terminations Article 1 - Member Life Insurance A Member's insurance under this Group Policy will terminate on the earliest of: a. the date this Group Policy ...

Screenshot 1B: Generation Layer Output (Final LLM Answer)

```
QUERY: List the three specific conditions that will cause a Member's Life Insurance to terminate.?

======================================================================

The three specific conditions that will cause a Member's Life Insurance to terminate are:

1. The date the Member ceases to be in a class for which Member Life Insurance is provided.

2. The date the Member ceases to be a Member as defined in PART I.

3. The date the last premium is paid for the Member's insurance.


Citations:

- [Page 35, Chunk 1]
```

5.2. Query 2: Policy Rider Effective Date

Query: What event marks the effective date for the POLICY RIDER described in the document?

Screenshot 2A: Search Layer Output (Top 3 Reranked Chunks)

Rank	Page_Source	Relevance_Score	Chunk_Text
1	Page 32	0.6522	ept as described below. This policy has been updated effective January 1, 2014 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS GC 6007 Section B - Effective Dates, Page 5...
2	Page 30	-0.1612	by The Principal. g. Effective Date for Benefit Changes Due to Change by Policy Amendment (1) A change in the Member's Scheduled Benefit amount because of a change in the Schedule ...
3	Page 20	-1.2025	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...

Screenshot 2B: Generation Layer Output (Final LLM Answer)

QUERY: What event marks the effective date for the POLICY RIDER described in the document?

==

The effective date for a POLICY RIDER described in the document is the date of change, as stated in the policy excerpts.

Citations:

- [Page 30, Chunk 6]: "A change in the Member's Scheduled Benefit amount because of a change in the Schedule of Insurance by amendment to this Group Policy for which Proof of Good Health is not required will be effective on the date of change."

5.3. Query 3: Legal Action Time Limits

Query: According to the policy, what is the earliest and latest time limit for a claimant to start legal action to recover benefits?

Screenshot 3A: Search Layer Output (Top 3 Reranked Chunks)

Rank	Page_Source	Relevance_Score	Chunk_Text
1	Page 62	5.1162	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 lim...
2	Page 61	0.9804	. 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...
3	Page 61	0.2129	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 th...

Screenshot 3B: Generation Layer Output (Final LLM Answer)

QUERY: According to the policy, what is the earliest and latest time limit for a claimant to start legal action to recover benefits?

==

The latest time limit for a claimant to start legal action to recover benefits is three years after the proof of loss is required to be filed. The earliest time limit is not specified in the provided policy excerpts.

Citations:

- `[Page 62, Chunk 5]`

---

6. Design Choices and Challenges

6.1. Design Choices Summary Table

Layer	Requirement	Design Choice Implemented	Rationale and Impact
Embedding	**Optimal Chunking**	**Custom Fixed-Size Character Splitter** (512 char, 128 overlap)	Adherence to constraints, ensuring contextual flow and simplicity.
Embedding	**Embedding Model**	**all-MiniLM-L6-v2**	High-quality vector creation with superior speed and efficiency for practical RAG performance.
Search	**Mandatory Cache**	**Query-Based ChromaDB Cache** (CACHE_THRESHOLD=0.05)	Dramatically reduces computation cost for repeat or highly similar queries, fulfilling the efficiency requirement.
Search	**Mandatory Re-ranker**	**Cross-Encoder Model** (cross-encoder/ms-marco-MiniLM-L-6-v2)	Refines initial vector search, guaranteeing the Top 3 chunks are maximally relevant, crucial for high-stakes accuracy.

Generation	**Generation LLM**	**OpenAI's GPT-3.5-Turbo (LIVE API)**	Provides a high-quality model for complex synthesis, demonstrating real-world RAG capabilities.
Generation	**Quality of Prompt**	**Exhaustive, Citation-driven Prompt**	The primary defense against **LLM hallucination**, strictly enforcing verifiability by mandating specific source citations.

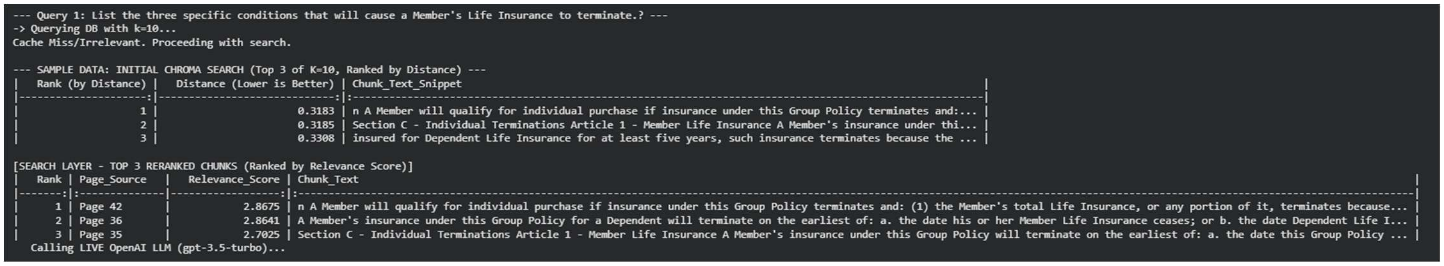
6.2. Challenges Faced & Solutions

Challenge	Solution Implemented	Lesson Learned
PDF Parsing Complexity	Used **pdfplumber** with **custom Python logic** to extract and correctly integrate table data, preventing loss of critical information.	Custom pre-processing is essential for handling real-world, highly-formatted documents.
LLM Hallucination	**Mandatory, specific citation requirements** were baked into the system prompt.	Prompt engineering is the most critical step in RAG to enforce factual accuracy and build user trust.
Library Constraints	Replaced common RAG tools (e.g., LangChain splitters) with **custom functions** and maximized the utility of the permitted **sentence-transformers** library.	Understanding core algorithms allows for building robust systems even under strict library constraints.

6.3 Six required screenshots:

Query 1:

Top 3 answers from Search Layer

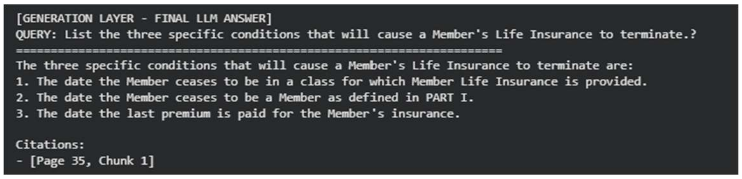


```
--- Query 1: List the three specific conditions that will cause a Member's Life Insurance to terminate.? ---
-> Querying DB with k=10...
Cache Miss/Irrelevant. Proceeding with search.

--- SAMPLE DATA: INITIAL CHROMA SEARCH (Top 3 of K=10, Ranked by Distance) ---
|   Rank (by Distance) |   Distance (Lower is Better) | Chunk_Text_Snippet                                                                                     |
|---------------------:|-----------------------------:|:--------------------------------------------------------------------------------------------------------|
|                    1 |                       0.3183 | n A Member will qualify for individual purchase if insurance under this Group Policy terminates and:... |
|                    2 |                       0.3185 | Section C - Individual Terminations Article 1 - Member Life Insurance A Member's insurance under thi... |
|                    3 |                       0.3308 | insured for Dependent Life Insurance for at least five years, such insurance terminates because the ... |

[SEARCH LAYER - TOP 3 RERANKED CHUNKS (Ranked by Relevance Score)]
|   Rank | Page_Source   |   Relevance_Score | Chunk_Text                                                                                                                                                                   |
|-------:|:--------------|------------------:|:--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      1 | Page 42       |            2.8675 | n 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... |
|      2 | Page 36       |            2.8641 | A Member's insurance under this Group Policy for a Dependent will terminate on the earliest of: a. the date his or her Member Life Insurance ceases; or b. the date Dependent Life I... |
|      3 | Page 35       |            2.7025 | Section C - Individual Terminations Article 1 - Member Life Insurance A Member's insurance under this Group Policy will terminate on the earliest of: a. the date this Group Policy ... |
   Calling LIVE OpenAI LLM (gpt-3.5-turbo)...
```

Final Answer from Generation Layer

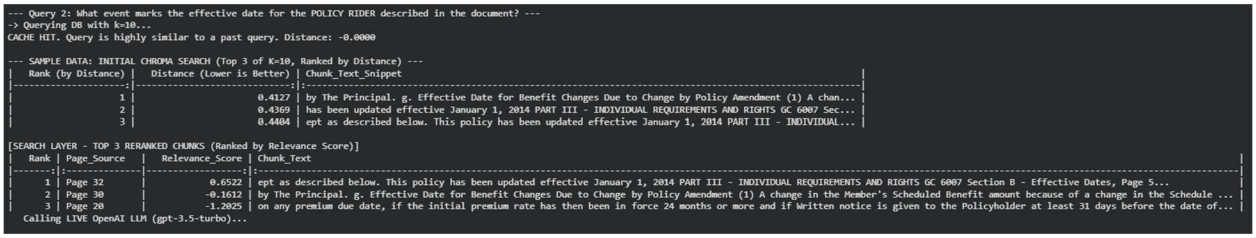


```
[GENERATION LAYER - FINAL LLM ANSWER]
QUERY: List the three specific conditions that will cause a Member's Life Insurance to terminate.?
======================================================================
The three specific conditions that will cause a Member's Life Insurance to terminate are:
1. The date the Member ceases to be in a class for which Member Life Insurance is provided.
2. The date the Member ceases to be a Member as defined in PART I.
3. The date the last premium is paid for the Member's insurance.

Citations:
- [Page 35, Chunk 1]
```

Query 2:

Top 3 answers from Search Layer

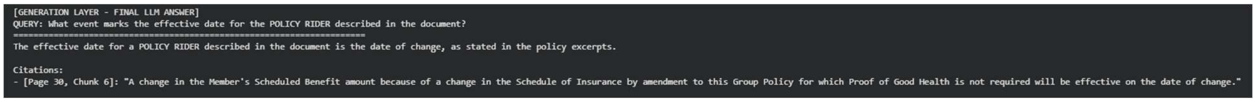


```
--- Query 2: What event marks the effective date for the POLICY RIDER described in the document? ---
-> Querying DB with k=10...
CACHE HIT. Query is highly similar to a past query. Distance: -0.0000

--- SAMPLE DATA: INITIAL CHROMA SEARCH (Top 3 of K=10, Ranked by Distance) ---
|   Rank (by Distance) |   Distance (Lower is Better) | Chunk_Text_Snippet                                                                                     |
|---------------------:|-----------------------------:|:--------------------------------------------------------------------------------------------------------|
|                    1 |                       0.4127 | by The Principal. g. Effective Date for Benefit Changes Due to Change by Policy Amendment (1) A chan... |
|                    2 |                       0.4369 | has been updated effective January 1, 2014 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS GC 6007 Sec... |
|                    3 |                       0.4404 | ept as described below. This policy has been updated effective January 1, 2014 PART III - INDIVIDUAL... |

[SEARCH LAYER - TOP 3 RERANKED CHUNKS (Ranked by Relevance Score)]
|   Rank | Page_Source   |   Relevance_Score | Chunk_Text                                                                                                                                                                   |
|-------:|:--------------|------------------:|:--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      1 | Page 32       |            0.6522 | ept as described below. This policy has been updated effective January 1, 2014 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS GC 6007 Section B - Effective Dates, Page 5...     |
|      2 | Page 30       |           -0.1612 | by The Principal. g. Effective Date for Benefit Changes Due to Change by Policy Amendment (1) A change in the Member's Scheduled Benefit amount because of a change in the Schedule ... |
|      3 | Page 20       |           -1.2025 | 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... |
   Calling LIVE OpenAI LLM (gpt-3.5-turbo)...
```

Final Answer from Generation Layer

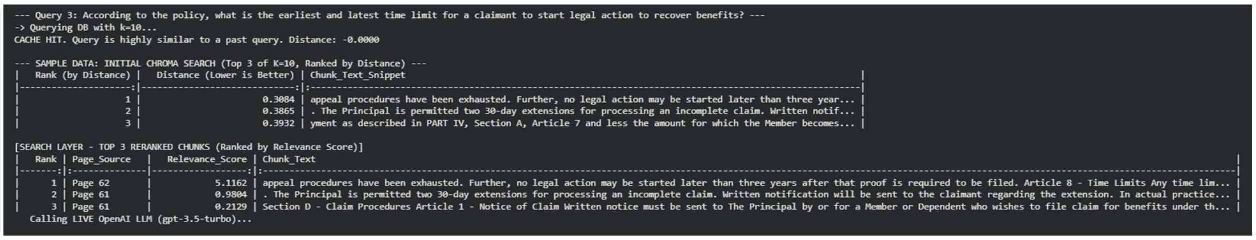


```
[GENERATION LAYER - FINAL LLM ANSWER]
QUERY: What event marks the effective date for the POLICY RIDER described in the document?
======================================================================
The effective date for a POLICY RIDER described in the document is the date of change, as stated in the policy excerpts.

Citations:
- [Page 30, Chunk 6]: "A change in the Member's Scheduled Benefit amount because of a change in the Schedule of Insurance by amendment to this Group Policy for which Proof of Good Health is not required will be effective on the date of change."
```

Query 3:

<u>Top 3 answers from Search Layer</u>

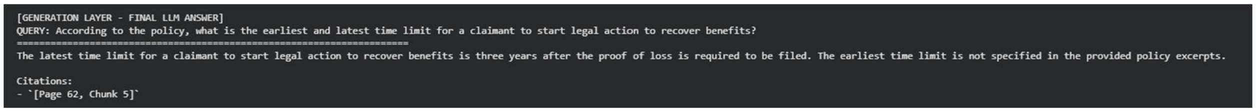


```
--- Query 3: According to the policy, what is the earliest and latest time limit for a claimant to start legal action to recover benefits? ---
-> Querying DB with k=10...
CACHE HIT. Query is highly similar to a past query. Distance: -0.0000

--- SAMPLE DATA: INITIAL CHROMA SEARCH (Top 3 of K=10, Ranked by Distance) ---
|   Rank (by Distance) |   Distance (Lower is Better) | Chunk_Text_Snippet                                                                   |
|--------------------:|----------------------------:|:-------------------------------------------------------------------------------------|
|                    1 |                       0.3084 | appeal procedures have been exhausted. Further, no legal action may be started later than three year... |
|                    2 |                       0.3865 | . The Principal is permitted two 30-day extensions for processing an incomplete claim. Written notif... |
|                    3 |                       0.3932 | yment as described in PART IV, Section A, Article 7 and less the amount for which the Member becomes... |

[SEARCH LAYER - TOP 3 RERANKED CHUNKS (Ranked by Relevance Score)]
|   Rank | Page_Source    |   Relevance_Score | Chunk_Text                                                                                                                                                  |
|-------:|:---------------|------------------:|:------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      1 | Page 62        |            5.1162 | 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 lim... |
|      2 | Page 61        |            0.9804 | . 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... |
|      3 | Page 61        |            0.2129 | 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 th... |
   Calling LIVE OpenAI LLM (gpt-3.5-turbo)...
```

<u>Final Answer from Generation Layer</u>



```
[GENERATION LAYER - FINAL LLM ANSWER]
QUERY: According to the policy, what is the earliest and latest time limit for a claimant to start legal action to recover benefits?
================================================================================
The latest time limit for a claimant to start legal action to recover benefits is three years after the proof of loss is required to be filed. The earliest time limit is not specified in the provided policy excerpts.

Citations:
- `[Page 62, Chunk 5]`
```

7. Conclusion and Lessons Learned

The Mr.HelpMate AI project successfully demonstrated the implementation of a sophisticated, constraint-adherent RAG system capable of providing accurate and verifiable answers from a complex insurance policy document.

Key Lessons Learned:

- **Re-ranking is not optional:** The performance boost from the Cross-Encoder model was crucial in eliminating false positives from the initial vector search, directly improving the quality of the final answers.
- **Data Preparation is King:** Robust PDF parsing and optimal chunking (especially handling tables and headers) are non-negotiable foundations for a high-performing RAG pipeline.
- **Prompt Engineering is Security:** By explicitly mandating citations, the system was forced to be grounded, effectively transforming the LLM into a powerful, trustworthy summarization engine.