

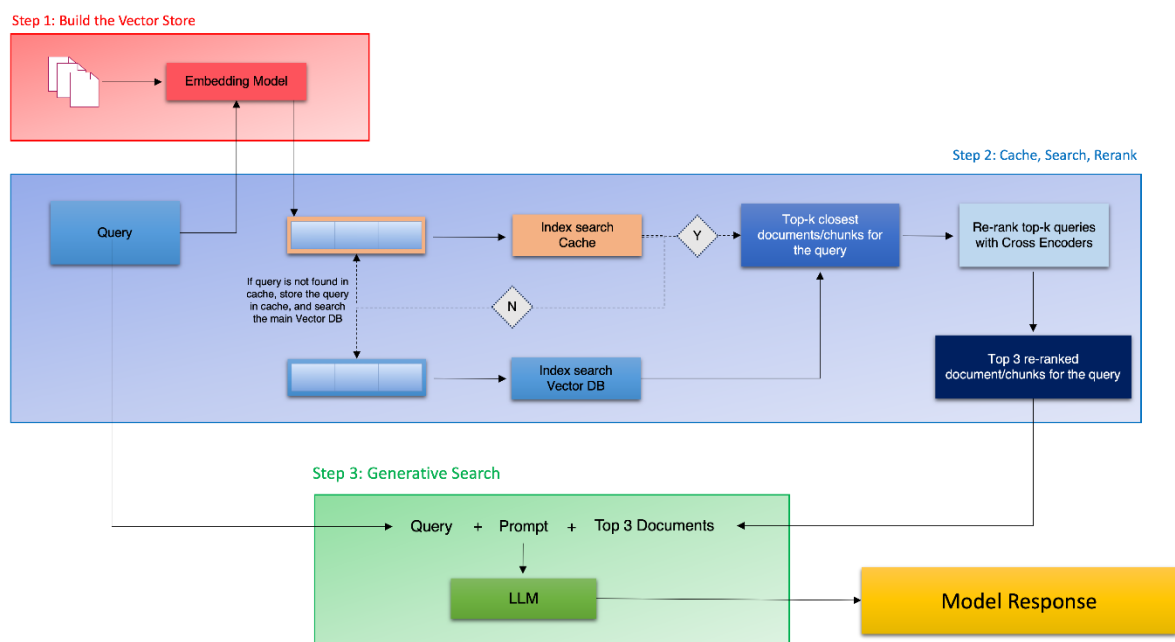
# Building HELPMATE AI using RAG

## Dataset:

A pdf consisting of the Principal Life Insurance policy that outlines terms, conditions, and benefits for member life, accidental death, and dismemberment, and dependent coverage, including accelerated benefits for terminal illnesses.

AIM: To create a model which will process this pdf document and create a vector store, and if a user queries any information, it must check the cache/main db retrieve the answer to the llm, which will generate the answer in a human readable format.

## Flow diagram:



## STEP 1: Building the vector store

- Extracted the text from the pdf pages using pdf plumber library and stored it into a data frame.

```
1 extracted_text_df
✓ 0.0s
```

	Page No.	Page_Text	Text_Length	Metadata
0	Page 1	DOROTHEA GLAUSE S655 RHODE ISLAND JOHN DOE 01/...	30	{'Page_No.': 'Page 1'}
2	Page 3	POLICY RIDER GROUP INSURANCE POLICY NO: S655 C...	230	{'Page_No.': 'Page 3'}
4	Page 5	PRINCIPAL LIFE INSURANCE COMPANY (called The P...	110	{'Page_No.': 'Page 5'}
5	Page 6	TABLE OF CONTENTS PART I - DEFINITIONS PART II...	153	{'Page_No.': 'Page 6'}
6	Page 7	Section A – Eligibility Member Life Insurance ...	176	{'Page_No.': 'Page 7'}
7	Page 8	Section A - Member Life Insurance Schedule of ...	171	{'Page_No.': 'Page 8'}
8	Page 9	P ART I - DEFINITIONS When used in this Group ...	387	{'Page_No.': 'Page 9'}
9	Page 10	T he legally recognized union of two eligible ...	251	{'Page_No.': 'Page 10'}
10	Page 11	(2) has been placed with the Member or spouse ...	299	{'Page_No.': 'Page 11'}
11	Page 12	An institution that is licensed as a Hospital ...	352	{'Page_No.': 'Page 12'}
12	Page 13	a . A licensed Doctor of Medicine (M.D.) or Os...	260	{'Page_No.': 'Page 13'}
13	Page 14	c . end stage renal failure; or d. acquired im...	316	{'Page_No.': 'Page 14'}
14	Page 15	A record which is on or transmitted by paper o...	36	{'Page_No.': 'Page 15'}
15	Page 16	PART II - POLICY ADMINISTRATION Section A - Co...	325	{'Page_No.': 'Page 16'}
16	Page 17	a. be actively engaged in business for profit ...	280	{'Page_No.': 'Page 17'}

- Embedding were created using the model- text-embedding-ada-002 using OpenAIEmbeddingFunction
- Storing these embeddings in the chroma db collections.

## STEP 2: Cache, search, rerank

- Cache collection was created and conditions were added to check if the information is present in the cache collection, then the search provides the top 10 results from there, else from the main collection that has all the embeddings

- After the top 10 results is obtained

```
Results are not found in the cache db, it is found in the main db
```

	Metadatas	Distances	IDs	\
0	{'Page_No.': 'Page 18'}	0.463001	15	
1	{'Page_No.': 'Page 16'}	0.479960	13	
2	{'Page_No.': 'Page 6'}	0.509958	3	
3	{'Page_No.': 'Page 37'}	0.510362	34	
4	{'Page_No.': 'Page 14'}	0.513758	11	
5	{'Page_No.': 'Page 15'}	0.517100	12	
6	{'Page_No.': 'Page 17'}	0.520049	14	
7	{'Page_No.': 'Page 19'}	0.524604	16	
8	{'Page_No.': 'Page 13'}	0.526859	10	
9	{'Page_No.': 'Page 25'}	0.539437	22	

```
Documents
```

0	c . a copy of the form which contains the stat...
1	PART II - POLICY ADMINISTRATION Section A - Co...
2	TABLE OF CONTENTS PART I - DEFINITIONS PART II...
3	b. a business assignment; or c. full-time stud...
4	c . end stage renal failure; or d. acquired im...
5	A record which is on or transmitted by paper o...
6	a. be actively engaged in business for profit ...
7	T he Principal has complete discretion to cons...
8	a . A licensed Doctor of Medicine (M.D.) or Os...
9	Section D - Policy Renewal Article 1 - Renewal...

- We can re-rank them by sorting them wrt to the cross encoder model's ranking

```
1 top_3_rerank = results_df.sort_values(by='Reranked_scores', ascending=False)
2 top_3_rerank[:3]
```

✓ 0.0s

	Metadatas	Distances	IDs	Documents	Reranked_scores
0	{'Page_No.': 'Page 18'}	0.463001	15	c . a copy of the form which contains the stat...	-6.536820
1	{'Page_No.': 'Page 16'}	0.479960	13	PART II - POLICY ADMINISTRATION Section A - Co...	-8.311792
6	{'Page_No.': 'Page 17'}	0.520049	14	a. be actively engaged in business for profit ...	-9.575331

### STEP 3: Generative search

- Written a prompt using the openai.chat.completions functions, the top 3 results would go into the input to the llm, and it provides an output of the query in a more human like format.

## OUTPUT:

### Query 1:

```
Customer: Give me steps in brief on how can i fill a claim.
Found in cache
Top 10 results are as follows:

Documents      Metadatas
0 Section D - Claim Procedures Article 1 - Notic... {'Page_No.': 'Page 61'}
1 A claimant may request an appeal of a claim de... {'Page_No.': 'Page 62'}
2 c . a copy of the form which contains the stat... {'Page_No.': 'Page 18'}
3 a. be actively engaged in business for profit ... {'Page_No.': 'Page 17'}
4 f . claim requirements listed in PART IV, Sect... {'Page_No.': 'Page 54'}
5 Section A - Member Life Insurance Schedule of ... {'Page_No.': 'Page 8'}
6 TABLE OF CONTENTS PART I - DEFINITIONS PART II... {'Page_No.': 'Page 6'}
7 T he Principal has complete discretion to cons... {'Page_No.': 'Page 19'}
8 Coverage During Disability will cease on the e... {'Page_No.': 'Page 51'}
9 b. a business assignment; or c. full-time stud... {'Page_No.': 'Page 37'}

Bot response:
To fill a claim, you can follow these steps:
1. Obtain a claim form containing the necessary details.
2. Complete the claim form accurately with all required information.
3. Attach any supporting documents as specified (like medical reports, receipts, etc.).
4. Submit the completed claim form and supporting documents to the insurance provider.

Based on the provided documents from the corpus:

The documents containing relevant information on filling a claim are from:
- Document Title: "PART II - POLICY ADMINISTRATION"
- Article Name: "Section A - Claims"
- Page Number: Page 16

Therefore, for detailed steps and information on filing a claim, refer to "PART II - POLICY ADMINISTRATION" under "Section A - Claims" on Page 16 of the insurance policy document.
```

### Query 2:

```
Query: what are the policy termination conditions?
Results are not found in the cache db, it is found in the main db
Top 10 results are as follows:

Documents      Metadatas
0 T he Principal may terminate the Policyholder'... {'Page_No.': 'Page 24'}
1 Section C - Policy Termination Article 1 - Fai... {'Page_No.': 'Page 23'}
2 Section C - Individual Terminations Article 1 ... {'Page_No.': 'Page 35'}
3 A Member's insurance under this Group Policy f... {'Page_No.': 'Page 36'}
4 Any individual policy issued will then be in f... {'Page_No.': 'Page 43'}
5 b. a business assignment; or c. full-time stud... {'Page_No.': 'Page 37'}
6 (1) If termination is as described in b. (1) a... {'Page_No.': 'Page 45'}
7 TABLE OF CONTENTS PART I - DEFINITIONS PART II... {'Page_No.': 'Page 6'}
8 a. be actively engaged in business for profit ... {'Page_No.': 'Page 17'}
9 c . a copy of the form which contains the stat... {'Page_No.': 'Page 18'}

Bot response:
The policy termination conditions are typically outlined in the insurance policy document under a specific section related to termination or cancellation. Please refer to the specifi

Based on the information found in the document excerpts provided:

- Document titled "PART II - POLICY ADMINISTRATION Section A - Company Requirements" on Page 16 may contain information relevant to policy termination conditions.
- Document excerpt starting with "a. be actively engaged in business for profit" on Page 17 may also contain details related to policy termination conditions.

I recommend reviewing these sections for specific information on policy termination conditions. If you have trouble locating the exact details, you can refer to the headings or secti
---
```

**\*\*Citations:\*\***

1. Title: PART II - POLICY ADMINISTRATION Section A - Company Requirements  
- Page Number: Page 16

**\*\*Citations:\*\***

1. Title: PART II - POLICY ADMINISTRATION Section A - Company Requirements  
- Page Number: Page 16
2. Document excerpt starting with "a. be actively engaged in business for profit"  
- Page Number: Page 17

### Query 3:

```
Query: What is the life insurance policy on accidental death?
Results are not found in the cache db, it is found in the main db
Top 10 results are as follows:

Documents      Metadata
0 Exposure Exposure to the elements will be pres... {'Page.No.': 'Page 55'}
1 Section A - Eligibility Member Life Insurance ... {'Page.No.': 'Page 7'}
2 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS ... {'Page.No.': 'Page 26'}
3 Section C - Individual Terminations Article 1 ... {'Page.No.': 'Page 35'}
4 Section A - Member Life Insurance Schedule of ... {'Page.No.': 'Page 8'}
5 f. claim requirements listed in PART IV, Sect... {'Page.No.': 'Page 54'}
6 (1) marriage or establishment of a civil union... {'Page.No.': 'Page 32'}
7 Section 8 - Member Accidental Death and Dismem... {'Page.No.': 'Page 53'}
8 Any individual policy issued will then be in f... {'Page.No.': 'Page 43'}
9 Member's death, the Death Benefits Payable ma... {'Page.No.': 'Page 47'}

Bot response:
The life insurance policy on accidental death provides coverage in case the insured person passes away due to an accident. Specific details regarding the coverage amount, exclusions,
However, you may refer to the following sections in the provided documents for insurance-related information:
1. Document: PART II - POLICY ADMINISTRATION Section A
  - This section might contain details on policy administration, which could include information on different coverage types.
2. Document: a. be actively engaged in business for profit
  - This section may provide insights into the eligibility criteria or terms related to insurance coverage.

Please review the mentioned sections in the documents to explore further details related to insurance policies and coverage specifics.
```

```
Citations:
- Document: PART II - POLICY ADMINISTRATION Section A
  - Page No.: Page 16

- Document: a. be actively engaged in business for profit
  - Page No.: Page 17
```

### My learnings:

- Learnt about pdfplumber, with which we can extract the page text from the pdfs
- Learnt about chroma db and how the collections are created, and updated, about how chromadb creates the embeddings by itself as well.
- Learnt the difference between cross encoders and bi-encoders.
- Also created a pipeline for all the steps in the flow to have a one step execution experience present in the last step of the notebook.