HelpMateAI – Generative Search using RAG on Insurance Data

Objective

The objective of this project is to build an efficient Generative Search system for insurance domain. A robust system capable of effectively and accurately answering questions from a policy document. We will be using a single long life insurance policy document (Principal-Sample-Life-Insurance-Policy.pdf) for this project.

System Design

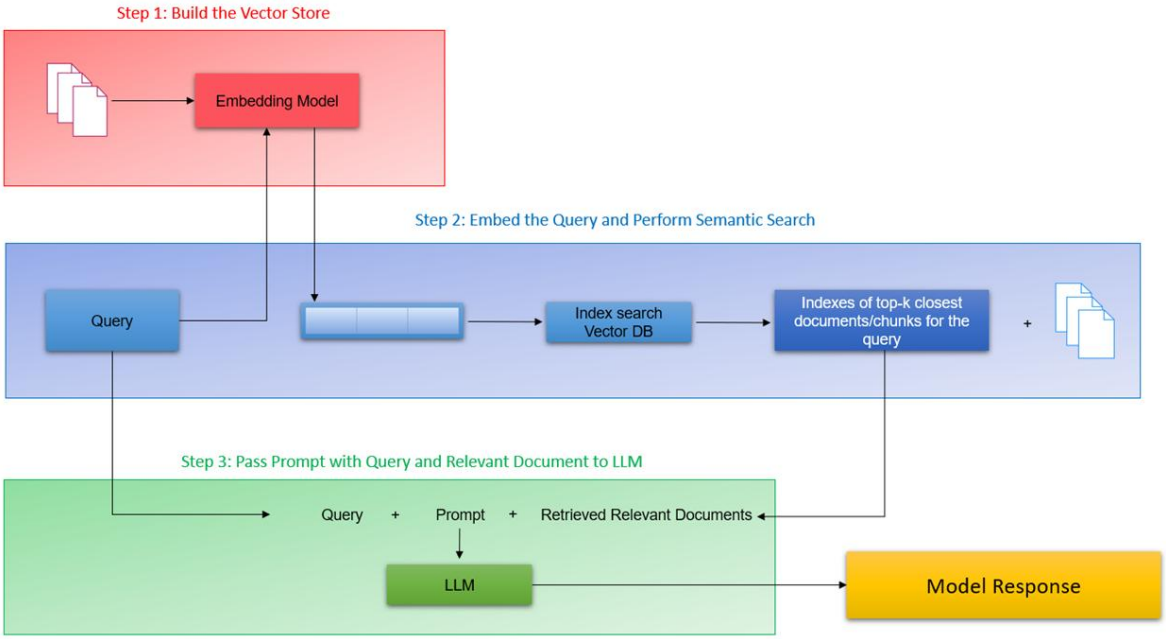
The Generative AI search system has the following key layers.

a) **Embedding Layer:** The PDF document is effectively processed, cleaned, and chunked for the embeddings. The choice of the chunking strategy will have a large impact on the final quality of the retrieved results. Two chunking strategies – Page level chunking and Fixed size chunking are tried to verify the results.
Another important aspect in the embedding layer is the choice of the embedding model. We tried using OpenAI's *text-embedding-ada-002* model to generate embeddings. And *sentence-transformers/all-MiniLM-L6-v2* model from the SentenceTransformers library on HuggingFace.

b) **Search Layer:** The search layer is built using ChromaDB. We will embed the chunks and store them in a ChromaDB collection. We will then perform a semantic search of a query in the collections embeddings to get several top semantically similar results. A Cache layer is also implemented in Semantic Search, so that results can be fetched easily if a query indeed matches to a query in cache.

   Re-Ranking is also implemented using Cross Encoder. Re-ranking the results obtained from semantic search can sometime significantly improve the relevance of the retrieved results. We tried two cross encoder models in CrossEncoder library from sentence_transformers – *ms-marco-MiniLM-L6-v2* and *ms-marco-MiniLM-L12-v2*.

c) **Generation Layer:** In this layer, we pass the final top search results to OpenAI's GPT 3.5 LLM along with the user query and a well-engineered prompt. To generate a direct answer to the query along with citations, rather than returning whole pages/chunks.



Retrieval Augmented Generation - System Design

Two experiments were conducted.

Approach 1

Chunking Strategy - Page Level Chunking

Embeddings - Using OpenAI's text-embedding-ada-002 model

Cross Encoder - Using cross-encoder ms-marco-MiniLM-L6-v2

Approach 2

Chunking Strategy - Fixed-Size Chunking

Embeddings - Default Embedding of Chroma DB (sentence-transformers/all-MiniLM-L6-v2 model)

Cross Encoder - Using cross-encoder ms-marco-MiniLM-L12-v2

Major functions implemented in the System:

- extract_text_from_pdf(): Function to extract text from a PDF file.
- search_vectordb(): Performs a semantic search of a query in the collections embeddings to get several top semantically similar results. Includes implementing Cache in Semantic Search.

- get_rerank_scores(): Utility function to fetch rerank scores using cross encoder.
- generate_response(): Generates a response using GPT-3.5's ChatCompletion based on the user query and retrieved information (the top 3 results after reranking)

Output

The following are the screenshots of results from the system. 3 screenshots against 3 self-designed queries that clearly showcase the top 3 results/chunks retrieved from the search layer. 3 screenshots of the same 3 queries with the final output generated by the LLM in the generation layer.

One observation is that the citations need improvement. The policy name returned by the LLM is not accurate, not picked the exact name from the metadata. Instead of relying on LLM response, we probably might have to use techniques like function calling to provide the citations exactly.

```
[54] # Query 1

     query = "what are the insurance benefits for hearing loss"
```

```
results_df = search_vectordb(query)
results_df = get_rerank_scores(cross_encoder, results_df, query)
top_3_semantic = results_df.sort_values(by='Distances')
top_3_semantic[:3]
```

1 to 3 of 3 entries Filter

index	Metadatas	Documents	Distances	IDs	Reranked_scores
0	{'Chunk No.': 1, 'Page_No.': 'Page 57', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	% of Scheduled Covered Loss Benefit Loss of Speech and/or Hearing Speech and Hearing 100% Speech or Hearing 50% Hearing in One Ear 25% Loss must be determined by a Physician to be permanent, complete, and irreversible. Total payment for all losses that result from the same accident will not exceed the Scheduled Benefit (or approved amount, if applicable). Payment for Loss will be to the Member. For this benefit, the term "Loss" means a total and irrevocable Loss of speech or hearing, which has continued for 12 consecutive months. Article 7 - Repatriation Benefit If a benefit is paid under this Section B for loss of the Member's life and death occurred at least 100 miles away from the Member's permanent place of residence, all customary and reasonable expenses incurred for preparation of the body and its transportation to the place of burial or cremation will be paid up to a maximum benefit payment of $2,000. Article 8 - Educational Benefit If a benefit is paid under this Section B for	0.6959014398996258	117	3.7280330657958984
1	{'Chunk No.': 1, 'Page_No.': 'Page 8', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	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	0.9734622240066528	8	-0.3218878209590912
2	{'Chunk No.': 2, 'Page_No.': 'Page 53', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	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. the injury must be the direct and sole cause of a loss listed in this Section B, Article 3; and d. the loss must occur within 365 days of the injury; and e. the limitations listed in this Section B, Article 9, must not apply; and This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6015 Section B - Member Accidental Death and Dismemberment Insurance, Page 1	1.1242238619044378	109	-5.6277384757995605

```
[57] top_3_RAG = top_3_rerank[["Documents", "Metadatas"]][:3]
     print("Query: ",query)
     response = generate_response(query, top_3_RAG)
     print("Response:")
     print("\n".join(response))
```

```
Query:  what are the insurance benefits for hearing loss
Response:
Insurance benefits for hearing loss typically vary depending on the specific insurance policy. Some insurance policies may cover partial or full costs of hearing aids, cochlear implants

Citations:
1. Policy Name: Member Life Insurance Schedule of Benefits
   Page Number: Page 8

Please refer to the above-cited policy document for detailed information on insurance benefits related to hearing loss.
```

```
[58] # Query 2
     query = 'In case of accident, what are the policies related to seat belt?'
```

```
results_df = search_vectordb(query)
results_df = get_rerank_scores(cross_encoder, results_df, query)
top_3_semantic = results_df.sort_values(by='Distances')
top_3_semantic[:3]
```

1 to 3 of 3 entries Filter

index	Metadatas	Documents	Distances	IDs	Reranked_scores
0	{'Chunk No.': 2, 'Page_No.': 'Page 55', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	was driving an Automobile equipped with a properly functioning driver-side air bag or riding as a passenger in an Automobile equipped with a properly functioning passenger-side air bag, although the Member's Seat Belt may not have been fastened at the time of the accident. The properly functioning and/or deployment of the air bag must be certified in the official report of the accident or by the investigating officer. For the purpose of this benefit, "Automobile" means a four-wheel passenger vehicle, station wagon, pick-up truck, or van-type vehicle, but excludes recreational-type vehicles such as a "dune-buggy" or an "all-terrain" vehicle. The term "Seat Belt" means a factory-installed device that forms an occupant restraint and injury avoidance system. Article 5 - Loss of Use or Paralysis Benefit This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6015 Section B - Member Accidental Death and Dismemberment Insurance, Page 3	0.5778053840339527	114	3.3792362213134766
1	{'Chunk No.': 1, 'Page_No.': 'Page 55', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	Exposure Exposure to the elements will be presumed to be an injury if: a. such exposure is due to an accidental bodily injury; and b. within 365 days after the injury, the Member incurs a loss that is the result of the exposure; and c. this Group Policy would have covered the injury resulting from the accident. Article 4 - Seat Belt/Airbag Benefit If the Member loses his or her life as a result of an accidental injury sustained while driving or riding in an Automobile, an additional benefit of $10,000 will be paid to the beneficiary named for Member Life Insurance, provided all Benefit Qualifications as described in Article 2 are met and: a. the Automobile is equipped with factory-installed Seat Belts; and b. the Seat Belt was in actual use by the Member and properly fastened at the time of the accident; and c. the position of the Seat Belt is certified in the official report of the accident or by the investigating officer. This additional benefit payment will also apply if the Member	0.6094187853070837	113	3.7365190982818604
2	{'Chunk No.': 1, 'Page_No.': 'Page 8', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	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	1.1743313074111938	8	-5.322818756103516

```
[61] top_3_RAG = top_3_rerank[["Documents", "Metadatas"]][:3]
     print("Query: ",query)
     response = generate_response(query, top_3_RAG)
     print("Response:")
     print("\n".join(response))
```

```
Query:  In case of accident, what are the policies related to seat belt?
Response:
In case of an accident, insurance policies related to seat belts typically revolve around the requirement for all occupants of a vehicle to wear seat belts at all times for safety rea

Please note that while the provided documents do not explicitly mention seat belt policies, it is important to refer to the specific policy documents related to auto insurance or pers

---

Citations:
1. Policy Name: Policy Name from Page 55
2. Policy Name: Policy Name from Page 8
```

```
[62] # Query 3
     query = 'What are the guidelines for adding or removing dependants from the insurance policy?'
```

```
results_df = search_vectordb(query)
results_df = get_rerank_scores(cross_encoder, results_df, query)
top_3_semantic = results_df.sort_values(by='Distances')
top_3_semantic[:3]
```

1 to 3 of 3 entries Filter

index	Metadatas	Documents	Distances	IDs	Reranked_scores
0	{'Chunk No.': 3, 'Page_No.': 'Page 43', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	insurance, then issued by The Principal to persons in the risk class to which the Dependent belongs on the individual policy's effective date. This policy has been updated effective January 1, 2014 PART III - INDIVIDUAL REQUIREMENTS AND RIGHTS GC 6011 Section F - Individual Purchase Rights, Page 2	0.7573674917221069	84	-2.5613622665405273
1	{'Chunk No.': 1, 'Page_No.': 'Page 17', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	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	0.817592978477478	27	-4.138967514038086
2	{'Chunk No.': 2, 'Page_No.': 'Page 18', 'Policy_Name': 'Principal-Sample-Life-Insurance-Policy'}	Group Policy. Article 7 - Certificates The Principal will give the Policyholder Certificates for delivery to insured Members. The delivery of such Certificates will be in either paper or electronic format. The Certificates will be evidence of insurance and will describe the basic features of the coverage. They will not be considered a part of this Group Policy. Article 8 - Assignments No assignments of Member Life Insurance will be allowed under this Group Policy. Article 9 - Dependent Rights A Dependent will have no rights under this Group Policy except as set forth in PART III, Section F, Article 2. Article 10 - Policy Interpretation This policy has been updated effective January 1, 2014 PART II - POLICY ADMINISTRATION GC 6003 Section A - Contract, Page 3	0.8255649209022522	30	-5.298238754272461

```
[65] top_3_RAG = top_3_rerank[["Documents", "Metadatas"]][:3]
     print("Query: ",query)
     response = generate_response(query, top_3_RAG)
     print("Response:")
     print("\n".join(response))
```

```
Query:  What are the guidelines for adding or removing dependants from the insurance policy?
Response:
To add or remove dependants from the insurance policy, you need to follow the guidelines outlined in the policy document:
1. Premiums will be based on the dependent's age and coverage.
2. Family members can contribute to a part of the premium for their dependent's coverage.
3. Any changes to the dependants covered under the policy need to be communicated to the insurance provider for proper documentation and premium adjustments.

For detailed guidelines on adding or removing dependants from the insurance policy, please refer to the policy document sections provided below:

Policy Document: Page 44
Policy Document: Page 22
Policy Document: Page 43
```