Introduction

This project focuses on developing a generative search system capable of answering questions from a life insurance policy document.

Traditional keyword-based search often fails to capture the semantic meaning of queries, leading to incomplete or irrelevant results. To overcome this, the system leverages embeddings, vector search, re-ranking, and large language models (LLMs) to provide more accurate and context-aware answers.

By combining these components, the system ensures that information from lengthy and complex policy documents can be retrieved and presented in a user-friendly, natural language format.

This not only improves accessibility but also enables quick decision-making for users seeking clarity from dense legal or insurance texts.

Approach / Workflow

The workflow of the project is structured into three main layers:

1. **Embedding Layer**
    - Process, clean, and chunk the policy document.
    - Experiment with fixed-size chunking strategies.
    - Generate embeddings using models such as OpenAI embeddings or SentenceTransformers.
2. **Search Layer**
    - Formulate at least 3 queries based on the policy document.
    - Embed the queries and perform semantic search using ChromaDB.
    - Implement a cache mechanism for efficiency.
    - Apply a re-ranking step using cross-encoder models to improve retrieval accuracy.
3. **Generation Layer**
    - Design a comprehensive prompt to guide the LLM.
    - Pass the retrieved results to the LLM for final answer generation.
    - Optionally, include few-shot examples to enhance response quality.

Detailed Methodology

1. **Reading & Processing PDF File**
    - The policy document is read using **pdfplumber**, which provides robust parsing capabilities.
    - Unlike basic text extractors, pdfplumber can handle text, tables, and images, offering more accurate preprocessing.
    - It also supports visual debugging, making it easier to validate the extracted content.

2. **Document Chunking**
    - Since the policy document is lengthy, the text is split into smaller chunks before embedding.
    - We begin with a **fixed-size chunking strategy**, but further experimentation with overlap or semantic chunking can improve retrieval quality.

3. **Generating Embeddings**
    - Each chunk is converted into vector representations using the **SentenceTransformer "all-MiniLM-L6-v2" model**.
    - These embeddings capture the semantic meaning of the text, enabling more relevant retrieval.

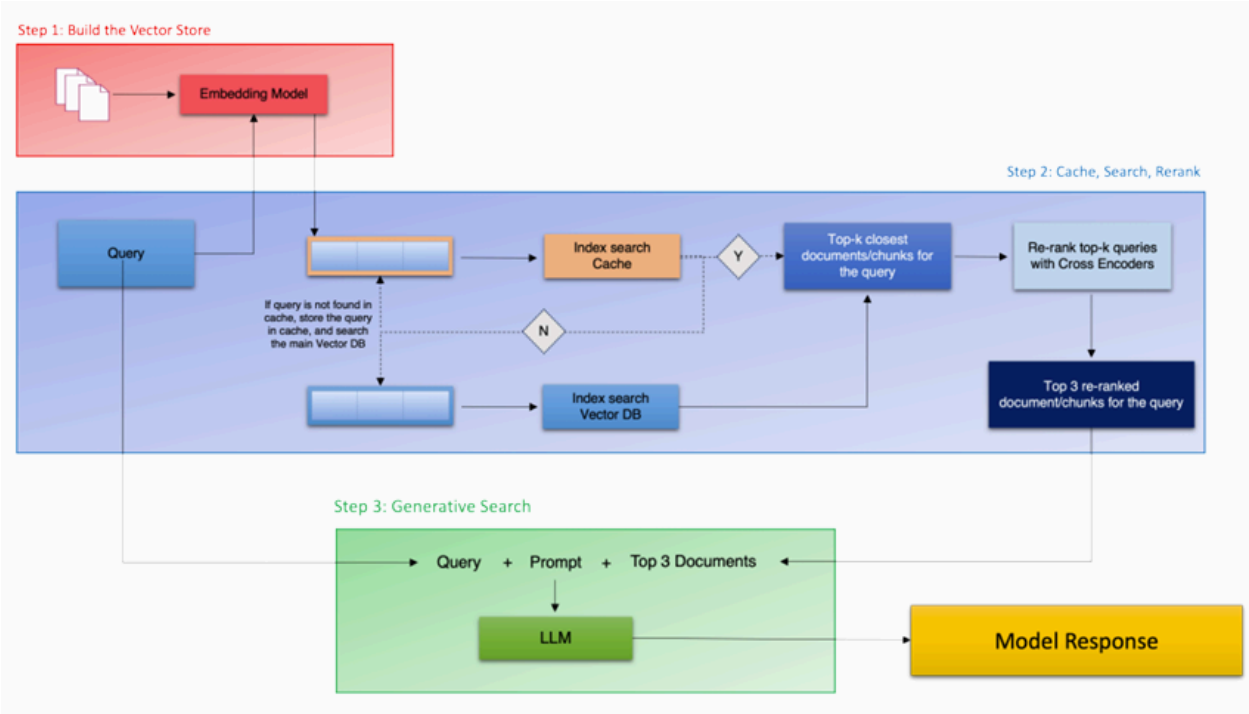
4. **Storing Embeddings in ChromaDB**
    - The generated embeddings are stored in **ChromaDB**, a vector database optimized for similarity search.
    - This enables efficient retrieval of relevant chunks based on user queries.

5. **Semantic Search with Cache**
    - User queries are embedded and searched against the database.
    - A **cache layer** is added to speed up repeated queries and reduce computational cost.

6. **Re-Ranking with a Cross Encoder**
    - Initial retrieval results are refined using a **cross-encoder model**, which evaluates query–response pairs.
    - This improves the relevance and ranking of the final results.

7. **Retrieval-Augmented Generation (RAG)**

- The top-ranked chunks are combined with the user query and passed into GPT-3.5 (or any chosen LLM).
- A carefully designed prompt ensures the model generates accurate, well-structured, and context-aware answers.
- Optionally, few-shot examples can be added to further guide the generation process.



Technical Setup

Language Model: Perplexity AI (*sonar-pro*) accessed via OpenAI client wrapper.

Programming Language: Python.

Interface: Jupyter Notebook (real-time user input handled via input()).

Results

Query 1 : what if i fail to pay premium?

Search Layer O/P

Search layer - Query 1

In [272…
```
from pprint import pprint
from IPython.display import display, HTML
```

In [273…
```
doc = get_first_document("what if i fail to pay premium?", results_df)["Documents"]
display(HTML(f"<pre style='white-space: pre-wrap'>{doc}</pre>"))
```

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. f. Termination (Premium Waiver) This policy has been updated effective January 1, 2014 PART IV - BENEFITS GC 6013 Section A - Member Life Insurance,

Generation Layer O/P

Generative layer - Query 1

In [281…
```
query = 'what if i fail to pay premium?'
df = search(query)
df = apply_cross_encoder(query, df)
df = get_topn(3, df)
response = generate_response(query, df)
print("\n".join(response))
```

When you fail to pay your insurance premium, several consequences can occur depending on the type of insurance and specific policy terms.

Grace Periods and Initial Consequences

Most insurance policies provide a **grace period** after a missed payment, typically ranging from 30 to 90 days depending on your coverage type[1][3][4]. During this time, your policy may remain active, but you'll likely face **late payment fees** and potential impacts to your insurance score, which can affect future premium costs[1].

For health insurance with advanced premium tax credits, you receive a **90-day grace period**, but this only applies if you've paid at least one month's premium in the current plan year[2]. During the first 30 days, insurers must continue paying claims, but after that, they can hold off on paying healthcare claims received during the grace period[2].

Life insurance policies typically offer a **30-day grace period**, though some contracts may extend this to 60 days[3][4]. For permanent life insurance policies with cash value, the accumulated cash value may automatically cover missed premiums to keep the policy active[5].

Policy Cancellation and Termination

If payment isn't received by the end of the grace period, insurance companies have the right to **cancel your policy**[1][3]. Once canceled, you may face several challenges:

- **Reinstatement requirements**: Companies may require not only the missed payment but also the next payment, and sometimes the full remaining premium[1]
- **Evidence of insurability**: For life insurance, you may need to provide proof of insurability for policy reinstatement[4]
- **Higher future premiums**: Finding new coverage after cancellation often results in higher premiums or difficulty obtaining insurance[1]

Long-term Impact

Late payments can negatively affect your **insurance score**, which represents the probability of filing a claim and influences your premium costs with future insurers[1]. This can make obtaining affordable coverage more challenging even after resolving the immediate payment issue.

Query 2 : what is the proof needed for showcasing ADL disability?

Search Layer O/P

Search layer - Query 2

In [274…

```
doc = get_first_document("what is the proof needed for showcasing ADL disability?", results_df)["Documents"]
display(HTML(f"<pre style='white-space: pre-wrap'>{doc}</pre>"))
```

terms of the Prior Policy, to have their premiums waived due to ADL Disability or Total Disability.
NOTE: When insurance under this Group Policy replaces coverage under a Prior Policy and the Active W
ork requirement is waived, any benefits payable will be the lesser of the Scheduled Benefit of this
Group Policy or the amount that would have been paid by the Prior Policy had it remained in force.
b. Effective Date for Initial Insurance When Proof of Good Health is Required This policy has been

Generation Layer O/P

Generative layer - Query 2

In [282…

```
query = 'what is the proof needed for showcasing ADL disability?'
df = search(query)
df = apply_cross_encoder(query, df)
df = get_topn(3, df)
response = generate_response(query, df)
print("\n".join(response))
```

To prove an **ADL (Activities of Daily Living) disability** for insurance purposes, you generally need to provide comprehensi
ve **medical documentation** and undergo an **assessment by a licensed healthcare professional**. The typical proof required
includes:

- **Medical records**: Hospital notes, physical therapy reports, and formal diagnoses documenting your condition and limitati
ons.
- **Chronically Ill Certification**: A statement from a physician certifying that you are chronically ill and require substan
tial assistance with at least two ADLs.
- **Claim forms**: Insurer-specific forms (such as PAQ or insured forms) detailing your daily care needs. It is recommended t
o provide detailed explanations and attach additional pages if necessary.
- **Benefit Eligibility Assessment (BEA)**: The insurance company will usually send a nurse or occupational therapist to eval
uate your ability to perform each ADL. This assessment may be conducted in person or virtually and will focus on:
  - The frequency and type of difficulty you have with each ADL.
  - The level of assistance required (hands-on or standby).
  - Any safety risks, such as falls or confusion.
- **Cognitive assessment** (if applicable): If cognitive impairment is suspected, standardized tests like the Mini-Mental Sta
te Exam (MMSE) or Montreal Cognitive Assessment (MoCA) may be used to determine if substantial supervision is needed.

Key points:
- The insurer's evaluator is not independent and may minimize your needs, so thorough and precise documentation is critical.
- Standby assistance (not just hands-on help) should be documented if it is necessary for your safety.
- Vague or incomplete submissions can lead to delays or denials.

If you are searching for these requirements in your policy, look for sections on "Benefit Eligibility Assessment," "ADL requi
rements," or "Claim Documentation" in your insurance documents[3].

Citations:
- ADL assessment and documentation requirements: Sandstone Law Group, Page 49, Chunk 4[3].

Query 3 : does this cover death due to not wearing seat belt?

Search Layer O/P

Search layer - Query 3

In [275…

```
doc = get_first_document("does this cover death due to not wearing seat belt?", results_df)["Documents"]
display(HTML(f"<pre style='white-space: pre-wrap'>{doc}</pre>"))
```

Payment of benefits will be subject to the Beneficiary and Facility of Payment provisions of this PA
RT IV, Section A. Article 6 - Member Life Insurance - Coverage During Disability A Member may be eli
gible to continue his or her Member Life and Member Accidental Death and Dismemberment Insurance and
Dependent Life Insurance coverage during the Member's ADL Disability or Total Disability. a. Coverag
e Qualification To be qualified for Coverage During Disability, a Member must: (1) become ADL

Generation Layer O/P

Generative layer - Query 3

In [283…

```
query = 'does this cover death due to not wearing seat belt?'
df = search(query)
df = apply_cross_encoder(query, df)
df = get_topn(3, df)
response = generate_response(query, df)
print("\n".join(response))
```

Based on the insurance policy documents provided, I cannot find specific information about coverage for death due to not wear
ing a seatbelt. The documents contain general information about life insurance benefits and termination conditions, but they
do not explicitly address seatbelt-related exclusions or coverage limitations.

What the Documents Show

The available policy excerpts discuss:
- Payment of benefits subject to benefit provisions
- Member Life Insurance coverage during disability periods
- Dependent's Life Insurance termination conditions

However, none of these sections specifically mention seatbelt use as a factor affecting death benefit coverage.

Important Considerations

While I cannot provide a definitive answer based on the available document excerpts, insurance policies typically contain det
ailed exclusions and conditions that may affect coverage. To get a complete answer about seatbelt-related coverage, you shoul
d:

1. **Review the full policy exclusions section** - This would typically outline circumstances where coverage might be limited
or denied
2. **Check the definitions section** - This may clarify what constitutes covered vs. non-covered deaths
3. **Look for accident-related provisions** - These sections often detail how accidental deaths are handled

Next Steps

I recommend reviewing the complete policy document, particularly focusing on:
- Exclusions and limitations sections
- Accidental death benefit provisions
- General conditions of coverage

You may also want to contact your insurance provider directly for clarification on this specific scenario.

Citations:
- Page 49, Chunk 4
- Page 42, Chunk 4
- Page 44, Chunk 4

Challenges Faced and Possible Improvements

Challenges Faced

1. Selecting the right chunking strategy was critical, as poor chunking reduced retrieval accuracy.
2. Ensuring relevant results during semantic search sometimes required tuning and experimentation with different models.
3. Re-ranking introduced additional computation, slightly impacting system latency.

Possible Improvements

1. Experiment with advanced chunking techniques (semantic or overlap-based) for better context preservation.
2. Use larger or domain-specific embedding models to improve semantic representation.
3. Optimize the caching mechanism to handle more queries efficiently.
4. Fine-tune prompts or use few-shot / chain-of-thought prompting for better generation quality.
5. Integrate a user-friendly interface (e.g., web app) for broader accessibility beyond Jupyter Notebook.