**Project Goals**

The project aims to develop a robust Generative Search System for answering queries from complex insurance policy documents. The traditional keyword search is inadequate due to the legal jargon and verbosity in policies. This project leverages LlamaIndex to implement semantic search that understands context and delivers precise answers.

Data Sources

The data sources include PDFs of insurance policies and claim documents. These documents were loaded using SimpleDirectoryReader from LlamaIndex, which handles directory-based PDF reading and document parsing. For this exercise we have used the document provided by UpGrad

Design Choices

1. **Document Loading:** SimpleDirectoryReader was used to load multiple PDFs, offering simple document management.
2. **Chunking:** The default chunking and splitting mechanism was employed using a sentence splitter to ensure that each document was broken down logically.
3. **Indexing:** The documents were transformed into nodes and indexed using VectorStoreIndex for efficient retrieval and semantic search.
4. **Query Engine:** We configured the query engine to return the top 3 most relevant documents (similarity\_top\_k=3) for each user query, ensuring focused and accurate responses.
5. **Evaluation:** Human feedback was utilized to evaluate the system, which resulted in positive outcomes and satisfactory answers to user queries.

Challenges

The primary challenge was handling document structure and ensuring the system did not hallucinate answers. Leveraging LlamaIndex's chunking and semantic search capabilities overcame this issue.

Code:

Please refer: https://github.com/Himachallad/InsuranceDocumentHelperAI