# Mr.HelpMate AI

There are large documents that speak about insurance policies. Its difficult for a person to go through it and find answers to his questions.

So lets build a robust generative search system capable of effectively and accurately answering questions from a policy document.

Lets use a single long life insurance policy document for this project stored at - [/content/drive/MyDrive/HelpMateAssignment/Principal-Sample-Life-Insurance-Policy.pdf](https://colab.research.google.com/drive/1DLu7SQJaeBsbeSAwCauq4IoKafC6lR_4)

Approach

1. The embedding layer - Here we read and effectively process the PDF documents. Divide the documents into chunks. We have used 2 ways of chunking - one is fixed size chunks and other is paragraph chunks. After chunking lets generate embeddings for those chunks using a pre-trained SentenceTransformer Model - "all-MiniLM-L6-v2"
2. The Searching layer - Here we build a semantic search. Take user query, embed it and then find the cosine similarity with existing chunks and list down top 3 chunks with maximum cosine similarity. Those are the chunk which matches the most with the user query. Also, we store these embedding and chunks to ChormaBD for fast retrieval of query results. ChromaDB provides faster and easier ways of searching.
3. The generative layer - He we create proper prompts that help in generating faster and accurate query results.