

Project Semantic spotter - Insure Assist AI

This application has two important use cases:

- Comparison across multiple policies: Helps insurance agents/ advisors, product teams and customers choosing between plans
- Granular benefit and claim related explanations: Helps in assisting claim handlers and policyholders for more comprehensive way to understand the policy clauses and benefits without going through lengthy documents

Overall workflow:

Document Extraction → Smart Chunking + Metadata → Embedding → VectorDB → User Query → Retrieval (hybrid) → Rerank → LLM Response (Comparison/Q&A) → Frontend_app

Core Use Cases

Policy comparison Questions

1. Compare the availability of a Surrender Value or benefit across the HDFC Life Easy Health (Single Premium), the HDFC Life Group Term Life, and the HDFC Life Sampoorna Jeevan (Regular Pay) policies
2. Compare the Suicide Exclusion clauses for the HDFC Life Group Poorna Suraksha and the HDFC Life Group Term Life policies, specifically differentiating the treatment of employer-employee versus non-employer-employee groups.
3. Contrast the availability of a Policy Loan facility and the stated interest rate structure for the loan under the individual savings policies, HDFC Life Sampoorna Jeevan Plan versus HDFC Life Sanchay Plus
4. Differentiate the Critical Illness Benefit (CIB) trigger and claim effect between the HDFC Life Easy Health policy and the HDFC Life Group Poorna Suraksha policy (Accelerated Critical Illness Option)
5. Compare the definition of "Sum Assured on Death" used in the HDFC Life Sanchay Plus Policy (Limited/Regular Pay) for an entry age of 30, with the definition of Sum Assured on Death used in the HDFC Life Smart Pension Plan.
6. Contrast the Free-Look Period for a policy purchased through Distance Marketing for an Individual Policyholder (e.g., HDFC Life Easy Health or HDFC Life Sanchay Plus) versus a Group Master Policyholder (e.g., HDFC Life Group Poorna Suraksha).

Single Policy Questions

1. For a non-employer-employee group under the HDFC Life Group Term Life Policy, what specific benefit amount is payable if an Insured Member dies by suicide within 12 months of joining?
2. List two requirements that the Policyholder must fulfill for a transfer or assignment of the HDFC Life Sampoorna Jeevan Policy to be operative against the insurer, according to the simplified Section 38 provisions
3. For the HDFC Life Group Poorna Suraksha Master Policy, what defines a "Hospital" based on minimum criteria for a town with a population of 15,00,000?
4. Describe how the Simple Reversionary Income Bonus (Bonus Option 2) is paid out under the HDFC Life Sampoorna Jeevan Plan if the Premium Payment Term is 15 years.
5. For the HDFC Life Easy Health Surgical Benefit Option, what specific percentage of the Sum Insured is payable for undergoing "Coronary Angioplasty with stent implantation (two or more coronary arteries must be stented)"?
6. In the HDFC Life Smart Pension Plan, what is the maximum Policy Discontinuance Charge (in Rupees) for a non-single premium policy with an Annualized Premium of Rs. 60,000, discontinued during Policy Year 4

Tech Stack

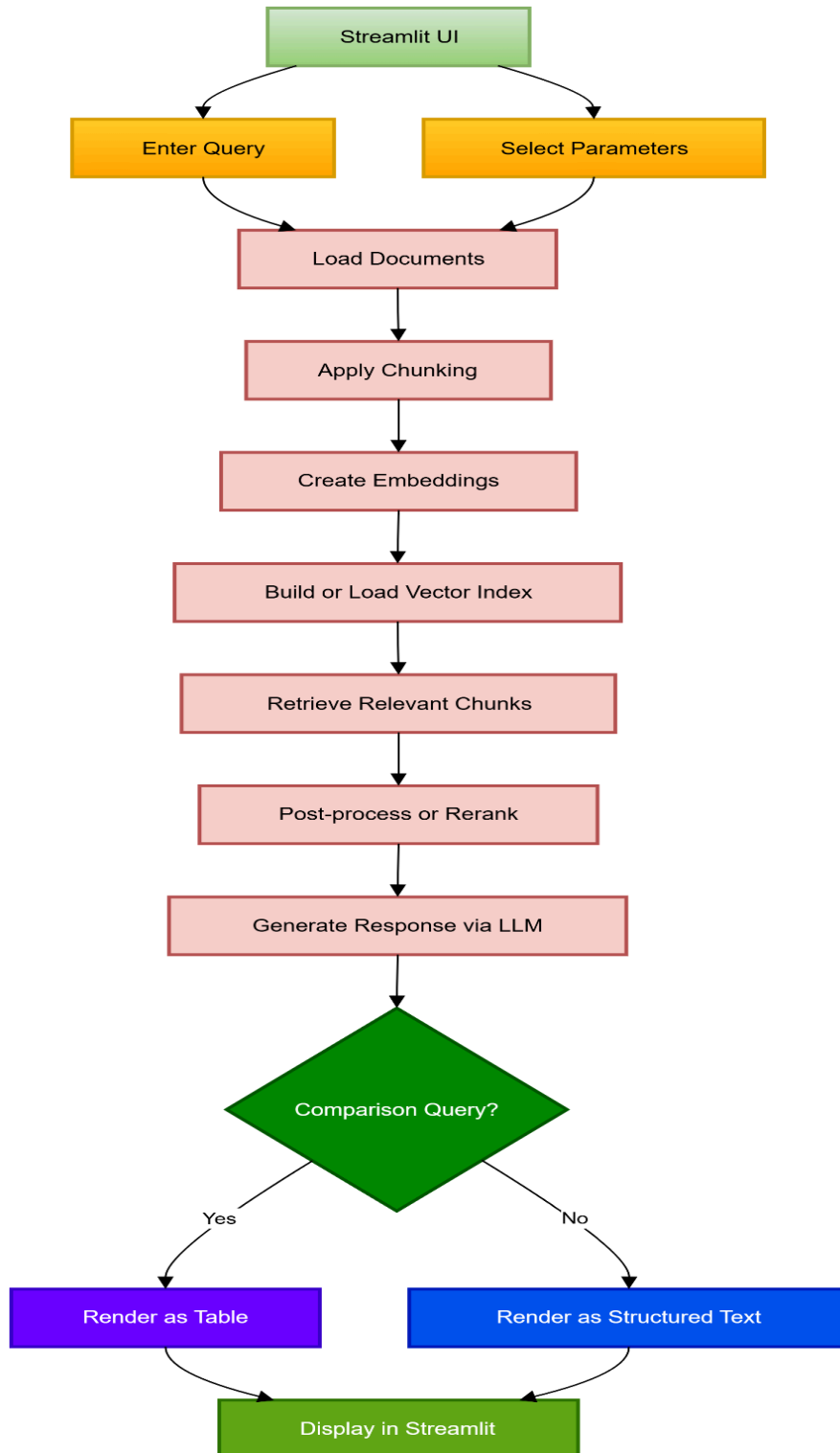
LlamaIndex - LlamaIndex is an open-source framework that excels at the "retrieval" and "augmentation" parts of RAG. It is particularly well-suited for my use case for several key reasons:

- Data-Centric Design
- Advanced data connectors and parsers
- Sophisticated Indexing and retrieval strategies - Mainly focusing on metadata filtering as simple semantic search is not enough for the policy information provided for this problem statement. LlamaIndex makes it straightforward to add metadata (e.g., policy_name, policy_type: 'group', UIN: '101N137V02') to each chunk of text. This allows the system to first filter for only the relevant documents (all group policies) before performing the semantic search, leading to far more accurate results.

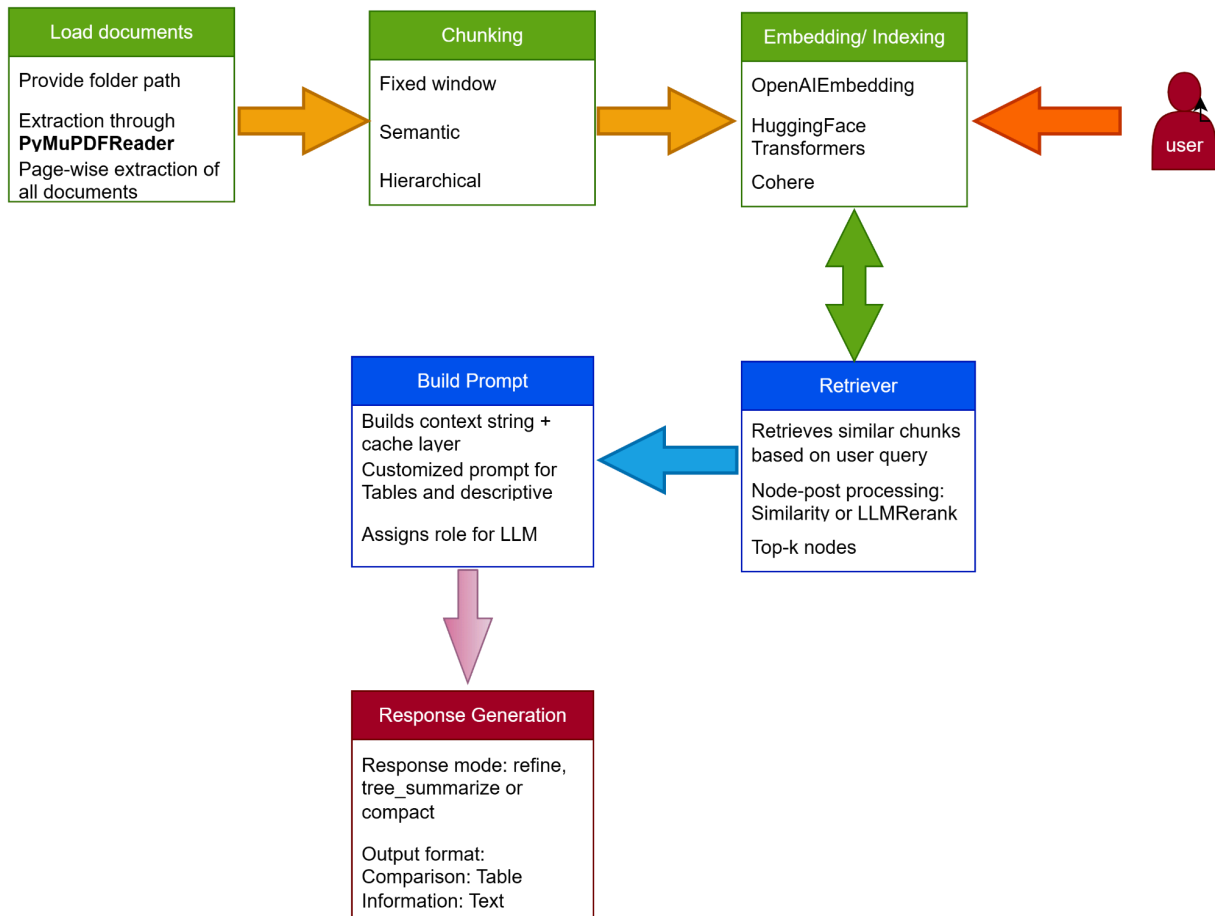
In short, Llama-Index is purposefully built for the RAG system. It is well-suited for ingesting and indexing complex documents like policies.

Let's dive into RAG Architecture of the “Insure Assist AI”

End-to-End flowchart of the Insure Assist AI search Engine



Layer-wise flowchart



Results

Evaluation is done for all the chunking methods across different techniques on both comparison questions and individual policy questions:

Individual Policy level questions:

Question	Embedding	Chunking strategies and response						Challenges
		Fixed window		Hierarchical		Semantic		
		Top nodes (Pages)	Response	Top nodes	Response	Top nodes	Semantic	
What is the claim procedure for HDFC Sanchay Plus Life Policy?	OpenAI	1,14,3,18	<p>The claim procedure for the HDFC Sanchay Plus Life Policy includes submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, the claimant may need to provide the original or copy of the First Information Report, Police Panchnama report, and Postmortem report, all attested by the respective authorities. It is important to note that the claim must be intimated to the insurance company within 90 days from the date of death, although delays may be condoned under certain circumstances.</p>	1, 13,14,3,18	<p>The claim procedure for the HDFC Sanchay Plus Life Policy involves the following steps:</p> <p>1. "Maturity Benefit": This will be paid if:</p> <ul style="list-style-type: none">- The policy has matured and the life assured is alive on the maturity date.- No claim has been made on the policy, except for any survival benefit.- The policy has not been discontinued, surrendered, cancelled, or terminated.- All relevant documents, including the original policy document, have been provided. <p>2. "Death Benefit": This will be paid if:</p> <ul style="list-style-type: none">- The death of the life assured occurs before the maturity date.- The standard policy provisions regarding exclusions and incorrect information are not applicable.- The policy has not been discontinued, surrendered, cancelled, or terminated.- All relevant documents in support of the claim have been provided. <p>**Basic Documentation Required**:</p> <ul style="list-style-type: none">- For death due to natural causes:<ul style="list-style-type: none">- Completed claim form (including NEFT details and bank account proof).- Original policy document.- Original or copy of the death certificate issued by the relevant authority.- Claimant's identity and residence proof.- For death due to unnatural causes:<ul style="list-style-type: none">- Completed claim form (including NEFT details and bank account proof).- Original policy document.- Original or copy of the death certificate.- Original or copy of the First Information Report (FIR) and police panchnama report.- Original or copy of the postmortem report.	1,14,3,18	<p>The claim procedure for HDFC Sanchay Plus Life Policy involves submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, the claimant needs to provide the original or copy of the First Information Report, Police Panchnama report, and Postmortem report, all attested by the respective authorities. The claim must be intimated to the insurer within 90 days from the date of death, although delays may be condoned under certain circumstances.</p>	OpenAI with hierarchical chunking gives the best answer because of it's page-wise sectional chunking, the context stayed intact to extract the info from the relevant page 13 and 14th of the policy document. Hence the chunking makes a lot of difference than the embedding methodology
	HuggingFace	1,14,3,18,2	<p>The claim procedure for the HDFC Sanchay Plus Life Policy includes submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, the claimant needs to provide the original or copy of the First Information Report and Police Panchnama report attested by the Police authorities, as well as the original or copy of the Postmortem report attested by the Hospital authority. It is important to note that depending on the circumstances of the death, further documents may be requested. The claim must be intimated to the insurance company within 90 days from the date of death, although delays may be condoned under certain genuine circumstances.</p>	1, 13,14,3,18	<p>The claim procedure for the HDFC Sanchay Plus Life Policy involves different steps based on the type of benefit being claimed. For the Maturity Benefit, the policyholder must ensure that the policy has matured, the Life Assured is alive on the Maturity Date, no other claims have been made on the policy, and all relevant documents have been provided to the company. On the other hand, for the Death Benefit, the benefit will be paid if the death of the Life Assured has occurred before the Maturity Date, standard policy provisions are not violated, the policy has not been discontinued or surrendered, and all relevant documents supporting the claim have been submitted to the company. The required basic documentation varies depending on whether the death is due to Natural Causes or Un-Natural Causes.</p>	1,14,3,18	<p>The claim procedure for HDFC Sanchay Plus Life Policy involves submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, the claimant needs to provide the original or copy of the First Information Report, Police Panchnama report, and Postmortem report, all attested by the respective authorities. The claim must be intimated to the insurer within 90 days from the date of death, although delays may be condoned under certain circumstances.</p>	
	Cohere	1,14,3,18,2	<p>The claim procedure for HDFC Sanchay Plus Life Policy involves submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, documents such as the First Information Report, Police Panchnama report, and Postmortem report may be required, all attested by the respective authorities. The claim must be intimated within 90 days from the date of death, although delays may be condoned under genuine circumstances.</p>	1, 13,14,3,18	<p>The claim procedure for HDFC Sanchay Plus Life Policy involves submitting relevant documents such as a completed claim form, original policy document, death certificate issued by the appropriate authority, claimant's identity and residence proof. Additional documentation may be required depending on the circumstances of the death. The claim must be intimated within 90 days from the date of death, but delays may be condoned if genuine reasons beyond the claimant's control are provided.</p>	1,14,3,18	<p>The claim procedure for HDFC Sanchay Plus Life Policy involves submitting the original or copy of the Death Certificate issued by the relevant authority, along with the claimant's identity and residence proof. Additionally, the claimant needs to provide the original or copy of the First Information Report, Police Panchnama report, and Postmortem report, all attested by the respective authorities. The claim must be intimated to the insurer within 90 days from the date of death, although delays may be condoned under certain circumstances.</p>	

Note: Please zoom in the image to see the results: To summarise, out of three different categories of chunking, **Hierarchical chunking+OpenAI embedding with GPT-4o-mini** has proven to be the best way of extracting relevant info with a constructive generation of response.

Policy Comparison questions - Sample generated response:

Question: Compare the Suicide Exclusion clauses for the HDFC Life Group Poorna Suraksha and the HDFC Life Group Term Life policies, specifically differentiating the treatment of employer-employee versus non-employer-employee groups.		
Policy Name	Feature/Benefit	Details
HDFC Life Group Poorna Suraksha	Suicide Exclusion for Employer-Employee Groups	In employer-employee groups, the full sum assured is payable to the nominee in the event of death due to suicide, regardless of the time since joining the scheme. This provides comprehensive coverage for employees.
HDFC Life Group Poorna Suraksha	Suicide Exclusion for Non-Employer-Employee Groups	For non-employer-employee groups, if a member dies by suicide within 12 months of joining or reviving the policy, the nominee is entitled to at least 80% of the total premiums paid or the surrender value, whichever is higher.
HDFC Life Group Term Life	Suicide Exclusion for Employer-Employee Groups	Similar to the Poorna Suraksha policy, in employer-employee groups, the full sum assured is payable upon death by suicide, ensuring that employees have robust protection under this clause.
HDFC Life Group Term Life	Suicide Exclusion for Non-Employer-Employee Groups	In non-employer-employee groups, the same 12-month limitation applies, where the nominee receives 80% of premiums paid or the surrender value if the member dies by suicide within that period, providing limited coverage.
Both Policies	Joint Life Provision	For both policies, in the case of joint life, if one member dies by suicide, the respective benefits will be payable to the surviving member, ensuring that the coverage continues for the remaining insured individual.

Policy Name	Feature/Benefit	Details
HDFC Life Easy Health	Critical Illness Benefit (CIB) trigger	The CIB trigger in HDFC Life Easy Health policy is typically based on the diagnosis of specific critical illnesses as defined in the policy document.
HDFC Life Easy Health	Claim Effect	Upon successful claim of the Critical Illness Benefit, the policyholder will receive a lump sum amount to cover medical expenses and other financial needs during the illness period.
HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Critical Illness Benefit (CIB) trigger	The CIB trigger in HDFC Life Group Poorna Suraksha policy (Accelerated Critical Illness Option) is activated upon the diagnosis of specific critical illnesses listed in Annexure IV of the Master Policy.
HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Claim Effect	After the payment of the Accelerated Critical Illness Benefit, the coverage for the Scheme Member ceases, and all benefits under the policy expire. The covered critical illnesses include Myocardial Infarction, Cancer of Specified Severity, Stroke resulting in permanent symptoms, and other critical conditions as specified in the policy document.

Compare the Critical Illness Benefit (CIB) trigger and claim effect between the HDFC Life Easy Health policy and the HDFC Life Group Poorna Suraksha policy (Accelerated Critical Illness Option)			
Policy Name	Feature / Benefit		Details
HDFC Life Easy Health	Critical Illness Benefit (CIB) trigger		Diagnosis of any covered Critical Illness during the Policy Term
HDFC Life Easy Health	Critical Illness Benefit (CIB) claim effect		Benefit payable is the Sum Assured and policy will terminate
HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Critical Illness Benefit (CIB) trigger		Diagnosis of any covered Critical Illness during the Policy Term
HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Critical Illness Benefit (CIB) claim effect		Benefit payable is the Sum Assured and policy will terminate

Application Outputs:

These are examples of successful functioning of the streamlit app

Deploy ⓘ

Configuration

Folder path for policy documents

Policy+Documents

Chunking Type

fixed_window

Chunk Size

600

Chunk Overlap

50

Embedding Type

OpenAIEmbedding

API Key File Path

C:/Users/SHAMBHAVISEN/D

Response Mode

tree_summarize

Insurance Policy Semantic Assistant

Ask your question:

Enter your policy-related question here

What is the claim procedure for HDFC Sanchay Plus Life Policy?

Run Semantic Assistant

Response generated successfully!

The claim procedure for the HDFC Sanchay Plus Life Policy involves submitting documents such as the original or copy of the Death Certificate, claimant's identity and residence proof, First Information Report, Police Panchama report, and Postmortem report. Depending on the circumstances of the death, additional documents may be required. The claim must be intimated within 90 days from the date of death, but delays may be condoned if genuine reasons beyond the claimant's control are provided.

Insurance Policy Semantic Assistant

Ask your question:

Enter your policy-related question here

Describe how the Simple Reversionary Income Bonus (Bonus Option 2) is paid out under the HDFC Life Sampoorna Jeevan Plan if the Premium Payment Term is 15 years.

Run Semantic Assistant

Response generated successfully!

The Simple Reversionary Income Bonus under Bonus Option 2 is paid out in the following manner if the Premium Payment Term is 15 years: The bonus declared at the end of the first Policy Year will be paid at the end of the 16th Policy Year, the bonus declared at the end of the 2nd Policy Year will be paid at the end of the 17th Policy Year, and so on. This pattern continues until the end of the Premium Payment Term, with the bonus being paid out each year after a specific number of years following its declaration.

Configuration

Folder path for policy documents

Policy+Documents

Chunking Type

Hierarchical

Chunk sizes (comma-separated)

512,256,128

Chunk overlaps (comma-separated)

50,25,10

Embedding Type

HuggingFace

API Key File Path

C:/Users/SHAMBHAV/VISEN/D

Response Mode

refine

Insurance Policy Semantic Assistant

Ask your question:

Enter your policy-related question here

Compare the Critical Illness Benefit (CIB) trigger and claim effect between the HDFC Life Easy Health policy and the HDFC Life Group Poorna Suraksha policy (Accelerated Critical Illness Option)

Run Semantic Assistant

Response generated successfully!

Comparison Table

	Policy Name	Feature / Benefit	Details
0	HDFC Life Easy Health	Critical illness Benefit (CIB) trigger	Triggered upon diagnosis of a covered critical illness
1	HDFC Life Easy Health	Critical illness Benefit (CIB) claim effect	Pays out the benefit amount upon diagnosis of a covered critical illness
2	HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Critical illness Benefit (CIB) trigger	Triggered upon diagnosis of a covered critical illness or undergoing a covered surgery
3	HDFC Life Group Poorna Suraksha (Accelerated Critical Illness Option)	Critical illness Benefit (CIB) claim effect	Pays out the Accelerated Critical illness Benefit upon diagnosis of a covered critical illness or undergo

localhost:8501

Finish update

Deploy

Configuration

Folder path for policy documents

Policy+Documents

Chunking Type

Hierarchical

Chunk sizes (comma-separated)

512,256,128

Chunk overlaps (comma-separated)

50,25,10

Embedding Type

Cohere

API Key File Path

C:/Users/SHAMBHAV/VISEN/D

Response Mode

refine

Insurance Policy Semantic Assistant

Ask your question:

Enter your policy-related question here

Contrast the Free-Look Period for a policy purchased through Distance Marketing for an Individual Policyholder (e.g., HDFC Life Easy Health or HDFC Life Sanchay Plus) versus a Group Master Policyholder (e.g., HDFC Life Group Poorna Suraksha).

Run Semantic Assistant

Response generated successfully!

Comparison Table

	Policy Name	Feature / Benefit	Details
0	Individual Policyholder (e.g., HDFC Life Easy Health or HDFC Life Sanchay Plus)	Free-Look Period	Allows the policyholder to review the terms and conditions of the policy and return it if not agreeable with
1	Group Master Policyholder (e.g., HDFC Life Group Poorna Suraksha)	Free-Look Period	Allows the Master Policyholder/Member to review the terms and conditions of the policy and return it if nc

localhost:8501

Finish update

Deploy

Conclusion & Challenges:

- To summarise, out of three different categories of chunking. **Hierarchical chunking+OpenAI embedding with GPT-4o-mini has proven to be the best way of extracting relevant info with a constructive generation of response (Robustness is achieved)**
- The major challenges were to pick the right chunking strategy and LLM model for response generation. GPT-3,5-turbo is the cheapest in terms of cost and credits efficiency. However, it was neither a good semantic extractor nor a response generator. GPT-4o-mini was doing a better job with the massive exhaustion of the credit limit. So, I generated the responses for some questions to stay on budget.
- llama index keeps updating it's version by discommissioning certain modules which hard to follow
- Semantic chunking is consuming so much time with poor performance than the fixed window. Hence it's not necessary to check them on different embedding criteria. I have limited them to just openAI embedding as the embedding strategy does make a significant difference on RAG outputs.