

Insurance Documents RAG QA Chatbot

Bridging the Gap Between Users and Policies

Table of Contents

- O1 About the Project
- 02 Key Features
- 03 Tech Stack
- 04 Challenges Addressed
- 05 Future Scope
- 06 Conclusion



About the Project

Intelligent Query Answering

- The Insurance Documents RAG QA Chatbot simplifies complex insurance policies by integrating retrieval and generation techniques, providing accurate answers in real time.
- Employs a Retrieval-Augmented Generation (RAG) pipeline for highly relevant answers.
- Combines LlamaIndex retrieval with OpenAI
 GPT-4o-mini/GPT-4o for user-centric response generation.
- Seamlessly retrieves relevant information and generates concise answers.



Key Features

Innovative Solutions

- Precise Responses: Uses RAG pipeline for relevance.
- Efficient Embedding: Utilizes ChromaDB for storage.
- Al-Driven Answers: Combines LlamaIndex with Al models.
- Dynamic Processing: Splits documents for retrieval.



Tech Stack



Powerful Technologies

- Language: Python.
- Frameworks/Libraries: LlamaIndex, ChromaDB, DiskCache.
- APIs/Models: OpenAI's Embedding Model, GPT-40, LlamaIndex.
- Cutting-edge technologies for robust performance.





Challenges Addressed

Overcoming Hurdles

- Optimized PDF Parsing: Enhanced extraction using Llamaindex.
- Embedding Efficiency: Added cache for redundant embeddings.
- Query Optimization: Integrated cache for efficiency.
- Enhanced Passage Ranking: Introduction of reranker.



Future Scope

- **Expanding Horizons**
- Multi-lingual Support: Extend processing capabilities.
- Generative Models: Integrate Claude AI for responses.
- Innovative advancements for broader functionality.
- Enhancing capabilities for diverse user needs.





Conclusion

Ensuring Precision

- Bridging the Gap: Delivers precise, contextual responses.
- Robust Caching: Efficient embeddings and querying.
- Seamless Experience: For policyholders and professionals.
- Enabling understanding of complex insurance policies.