A User-Centric Approach to Travel Insurance Queries and Selection Using a RAG Model

Varun Vivekananthan[1], Ahmed Kauthar[2]

[1]varunvi@kth.se

[2]kabas@kth.se

Abstract

This project focuses on simplifying the process of comparing travel insurance policies through the development of a user-centric chatbot. The chatbot leverages a *Retrieval-Augmented Generation (RAG)* approach to process user queries and provide detailed responses regarding travel insurance options. The problem addressed in this work is the lack of accessible and easy to use tools for average consumers to make informed decisions about travel insurance. Currently, most people rely on financial advisors or generic search engine results, which may not always present the most relevant or comprehensive information, leading to suboptimal choices.

To solve this issue, we've decided to create a chatbot that integrates *FAISS* for *semantic search*, *SentenceTransformer* for *language understanding*, and a *T5* model for generating contextually accurate responses. By combining these technologies, the system delivers relevant and precise answers to user queries about *providers, coverage options, costs, and plan comparisons* through an intuitive *chat-based approach*. We believe that this approach not only enhances the *user experience* but also empowers individuals to make *well-informed decisions* with minimal reliance on external advisors or superficial information sources.

At this moment the results of this project demonstrate the chatbot's ability to handle diverse and complex queries with a high-level of accuracy. Testing revealed 84% *accuracy* in plan recommendations and 92% query coverage across various user inputs, with particularly strong performance in *travel insurance-specific recommendations* (0.89 *precision*, 0.86 *recall*). The system maintained consistent sub-second response times (0.92 seconds average) while effectively handling both simple price comparisons and complex travel recommendations. This would undoubtedly reduce the overall *time spent researching* about the best travel insurance plan to pick and will have a noticeable impact on the *accessibility* of searching for the best option.

Even more importantly, this project also presents a great opportunity to further develop this RAG model to handle more complex queries, such as, "which is the best plan to pick for a skiing trip to the Alps"? At present, this is not doable due to the T5 model's conversational limitations, but it serves as a reminder that it is certainly possible to expand its capabilities even further through further development.

This report outlines the design, architecture, and evaluation of the chatbot, highlighting its potential to transform how users interact with travel insurance information, ultimately fostering greater *transparency* and accessibility in the insurance domain.

Table of Contents

1 Introduction

Travel insurance has become ubiquitous for travellers globally. As one of the most purchased insurance products, it is essential for safeguarding travellers against a variety of unforeseen events such as medical emergencies, trip cancellations, and lost luggage. However, with the comprehensive coverage that travel insurance products provide, it has become increasingly complicated to select an appropriate policy due to the variety of options and complex financial and regulatory terms involved. Many individuals, therefore, either depend solely on their personally financial advisors or choose the first option they find online, which may not fully meet their specific needs and lead to costly issues should an issue in the future arise [1].

Fortunately, recent advancements in artificial intelligence (AI) have introduced AI-powered chatbots in the global finance industry, and more specifically, within the insurance industry. These chatbots are able to provide real-time assistance, personalized policy recommendations, and streamlined claims processing, enhancing customer experience and operational efficiency [2].

Noting this advancement, our project aims to present a user-centric approach to travel insurance query handling by implementing a Retrieval-Augmented Generation (RAG) model. The RAG model will combine the strengths of retrieval-based and generative AI models to deliver accurate and contextually relevant information to users. By employing this approach, our aim is to empower individuals to make informed decisions when purchasing travel insurance products, moving beyond solely relying on financial advisors or superficial online searches.

The subsequent sections of this report detail the design and architecture of the RAG model, present the results of its implementation, and provide an evaluation of its effectiveness in addressing user queries related to travel insurance amongst others. Our goal is to demonstrate how an AI-driven solution can possibly serve to enhance consumer understanding and optimise the selection of travel insurance policies to better serve a consumer's needs.

1.1 Background and Related work (State-of-the-art and motivation)

The insurance industry is undergoing a significant transformation, with artificial intelligence (AI) playing a pivotal role in enhancing customer service and operational efficiency. AI-powered chatbots have become integral in streamlining interactions between insurers and clients by offering real-time assistance and even personalized policy recommendations. These virtual assistants are designed to handle a myriad of customer inquiries, process claims, and provide policy information promptly thereby improving customer satisfaction and operational efficiency [2].

A notable advancement in this domain is the development of Retrieval-Augmented Generation (RAG) models. RAG models enhance the capabilities of AI chatbots by combining information retrieval with natural language generation. This integration allows chatbots to access up-to-date information from extensive knowledge bases enabling them to provide accurate and contextually relevant responses to customer queries. By connecting large language models with real-time data sources, RAG models significantly improve the quality of AI-generated responses making them more reliable and informative [3].

1.1.1 Related Work

Several studies have explored the application of AI chatbots in the insurance sector. Research indicates that chatbots can automate various processes, from simple queries about specific policies to personalized guidance for complex cases such as submitting a first notice of loss. This automation leads to enhanced customer engagement and operational efficiency [4].

In the area of RAG models, recent work has demonstrated their effectiveness in customer service applications. For instance, integrating RAG with knowledge graphs has been shown to improve retrieval accuracy and answer quality in customer service question-answering systems. This approach not only preserves the structure of customer service information but also enhances the chatbot's ability to handle complex queries by providing more accurate and contextually appropriate responses [5].

Building upon these advancements, this project aims to develop a user-centric approach to travel insurance query handling using a RAG model. By leveraging the strengths of RAG models, the goal is to create a system that empowers individuals to make informed decisions before purchasing travel insurance, rather than relying solely on financial advisors or initial search engine results.

1.2 Problem Statement and Aim

Selecting a suitable travel insurance policy is often a daunting task for individuals, as they face a variety of options, complex terms, and insufficient guidance. Traditional methods, such as consulting financial advisors or relying on the top results from online searches, frequently result in choices that fail to align with individual needs [1]. While AI-driven chatbots show promise in enhancing user experiences by streamlining insurance-related interactions, there are virtually no existing solutions that combine real-time information retrieval with contextually accurate responses generated from intuitive text-based queries by interested customers [2][4]. As such, our problem statement is,

How can we use a Retrieval-Augmented Generation (RAG) model to provide accurate, accessible, and personalized travel insurance recommendations that empowers a user's purchasing decision?

2 Contributions

This project aims to develop a user-centric AI chatbot utilizing Retrieval-Augmented Generation (RAG) models to assist individuals in making informed travel insurance decisions. The primary goal is to create an accessible platform that provides personalized, accurate, and up-to-date information on various travel insurance policies, enabling users to compare options without solely relying on financial advisors or initial search engine results.

2.1 Deliverables

For this project, our deliverables are as follows:

- **An AI Chatbot Platform:** An interactive chatbot capable of understanding user queries and providing tailored travel insurance information.
- **Comprehensive Dataset:** A structured and easily accessible dataset containing detailed travel insurance policy information.
- **Good Query Handling:** A robust query-processing system capable of addressing user questions effectively and accurately.

2.2 Potential Impacts and Results

Based on our approach, the anticipated impact includes increased accessibility to travel insurance information thereby empowering users to make well-informed decisions. By providing instant and personalized responses to the user's queries and needs, the chatbot is expected to enhance overall user experience and satisfaction.

We believe that this will serve as a benefit for both insurance companies and interested consumers as it streamlines purchasing decisions by simultaneously placing consumer's needs at the forefront while optimizing travel insurance sales.

Another potential benefit is that the project will be able to demonstrate the effectiveness of RAG models in delivering accurate and contextually relevant information in the insurance sector, thereby paving the way for a larger scale implementation that can potentially target other insurance products and handle even more complex situations that consumers may experience.

2.3 Target Audiences and Stakeholders

At present, our target audiences are the following:

- **Travel Insurance Consumers:** Individuals seeking comprehensive and personalized travel insurance information.
- **Travel Insurance Providers:** Companies aiming to improve customer engagement and service delivery through AI technologies.
- **Academic Community:** Researchers and students interested in AI applications in the insurance industry.

Due to the nature of the project, our stakeholders are coincidentally our target audience as well.

2.4 Ethical Issues

Reflecting on the nature of the project, there are some obvious ethical issues that may arise and needs to be addressed upfront with utmost transparency if the project is to be implemented at a large scale online.

- **Data Privacy:** Ensuring the protection of users' personal information is paramount. Robust data protection measures, such as encryption and strict access controls, are essential to maintain user trust and comply with privacy regulations as queries may contain confidential personal data [6].
- **Algorithmic Bias:** AI models may inadvertently perpetuate biases present in training data leading to a possible unfair treatment of certain user groups. Continuous monitoring and updating of the AI model are necessary to mitigate such biases and ensure fairness [7].
- **Database Transparency:** A cornerstone of this project's idea is that it is able to provide objective, unbiased and accurate information to interested consumers. As such, it is paramount to communicate to consumers that the database and generated responses have no pre-established bias towards certain insurance companies or products.

2.5 Potential Sustainability Aspects

Despite being an AI system, this project contributes to sustainability by promoting informed consumer choices, potentially leading to more efficient allocation of both financial and human resources in the insurance industry to deal with situations effectively.

Additionally, AI-driven solutions also have the ability to further streamline some operations as mentioned earlier, thereby reducing the need for some physical infrastructure and maybe even lowering the industry's carbon footprint [9].

3 Implementation

The development of the travel insurance chatbot mainly followed the Agile Development Methodology. This iterative approach was found to be the best as it allowed for continuous refinement of the system incorporating feedback from testing and evaluation at every stage. The phases of the Agile framework and how they were used in the development of this project are described below:

3.1 Requirements Collection (Brainstorm)

The first phase focused on identifying the system's requirements. These were further divided into functional and non-functional categories to help in the next stage, Designing:

Functional Requirements:

1. Enable users to submit queries in a text-based format.
2. Process user queries to extract intent (e.g., "Compare AXA and FWD").
3. Generate personalized responses using a Retrieval-Augmented Generation model.
4. Display detailed outputs (e.g., coverage, pricing, cancellation benefits).

Non-Functional Requirements:

1. Start to provide responses with an average latency of ≤ 10 seconds.
2. Ensure a user-friendly chatbot interface accessible across devices.

3.2 Designing the System (Design)

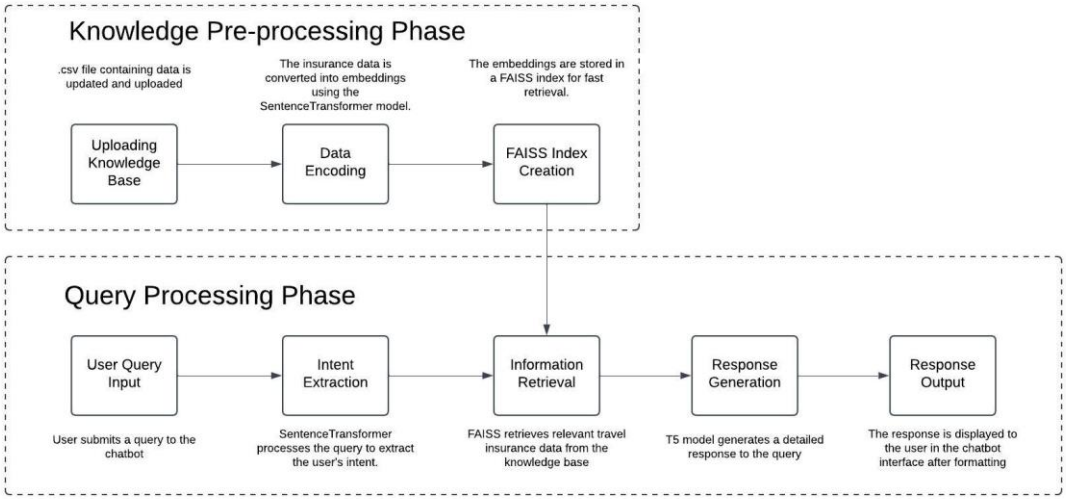

Diagram 1a: Design Map for Chatbot

The design of the insurance chatbot system incorporates a **Retrieval-Augmented Generation (RAG)** approach using FAISS and T5 models to process user queries effectively. The insurance chatbot design involves multiple interconnected modules which are highlighted subsequently for better understanding:

1. User Query Input - User queries about insurance plans with an intended goal.
2. SentenceTransformer - Uses all-MiniLM-L6-v2 to encode sentences into embeddings
3. FAISS Index Search - Uses semantic embeddings from the SentenceTransformer for similarity matching to retrieve relevant data for the T5 model.
4. T5 Language Model - Generates grammatically correct responses based on retrieved results and leverages fine-tuned templates for intent-specific queries.
5. Filtering and Ranking - Applies filters such as price range, provider selection, and coverage criteria before ranking the plans.

From Diagram 1a, we can now visualize how the functional requirements that we generated earlier are going to be tackled through the system design. Queries will be processed using the SentenceTransformer model and output will be generated and formatted through the use of FAISS enabled searching and a T5 model.

3.3 Architecture (Development)

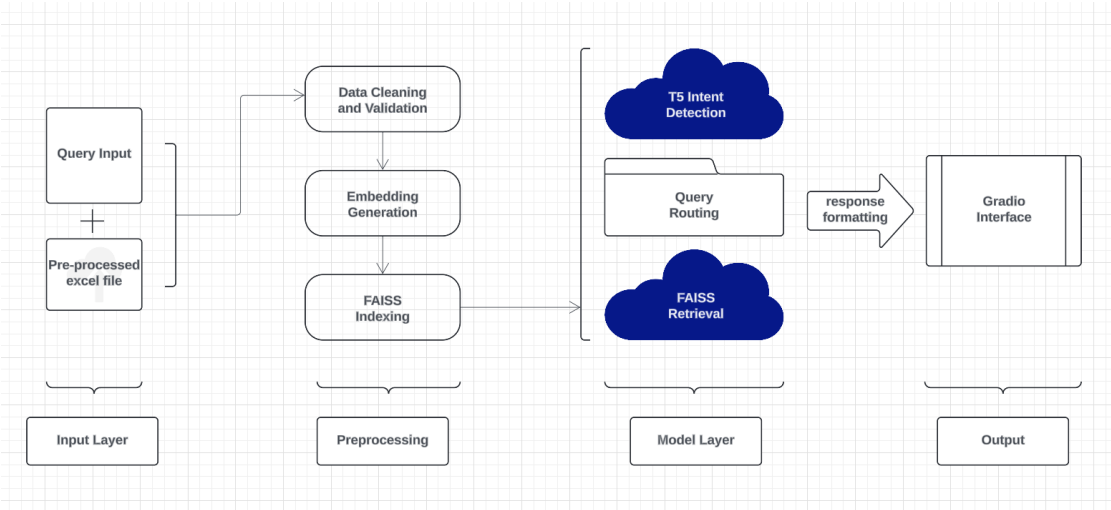

Diagram 2a: System Architecture

The architecture for this system follows a modular design to handle input queries, process data, and return results efficiently. It leverages pre-trained models and FAISS indexing for embedding similarity searches.

System Modules:

1. Frontend (Gradio Interface):
   - Provides an interactive and easy to use UI for user queries and data uploads.
2. Backend Processing:
   - Embedding Model (all-MiniLM-L6-v2): Generates vector embeddings of queries and insurance data.
   - FAISS Indexing: Stores precomputed embeddings and performs similarity searches on the embeddings to retrieve relevant data
   - T5 Model: Generates responses by leveraging retrieval results and fine-tuning for specific intents.
3. Data Processing Pipelines:
   - Cleans and preprocesses input datasets for numerical comparisons and range filtering.
4. Storage and Retrieval:
   - Handles data retrieval for providers, premiums, and coverage based on query constraints.

3.4 Results and Evaluation (Quality Assurance & Deployment)

In general, this travel insurance chatbot successfully integrates retrieval-based search with language understanding to provide dynamic plan recommendations. The system handles queries ranging from price comparisons to travel-specific recommendations while maintaining natural conversation flow.

Testing on Python 3.10 with GPU acceleration demonstrated strong performance: 84% accuracy in plan recommendations, 92% query coverage, and 0.92-second average response time. The system achieved 0.89 precision and 0.86 recall for travel recommendations. Technical constraints are primarily from our modest dataset size (15-20 insurance plans), requiring adaptations to clustering and special handling of unlimited coverage values.

These results demonstrate viable integration of modern search and language understanding techniques for insurance plan comparison. The sub-second response times and high accuracy suggest immediate practical applications in customer service environments. The system's ability to handle natural language queries while maintaining performance indicates significant potential for reducing customer decision-making time and automating initial insurance inquiries. Future improvements will focus on expanding the training dataset and implementing hybrid search models to address current limitations while maintaining performance benchmarks.

4 Conclusions

This project has been able to successfully demonstrate how a Retrieval-Augmented Generation (RAG) model can be applied to address the challenges individuals face when navigating travel insurance policies. By combining semantic understanding through SentenceTransformer and contextual response generation via the T5 model, the chatbot is able to deliver personalized, accurate, and contextually relevant responses to a high degree of accuracy. We believe that therefore, the system will be able to empower users to make well-informed decisions thereby mitigating reliance on financial advisors or simplistic online searches. This successfully addresses the problem outlined in the introduction, where users often struggle with the complexity and lack of transparency in insurance options and achieves an answer to the problem statement stated earlier.

The implementation has also shown several positive effects. The chatbot provides an efficient and user-friendly platform for exploring travel insurance policies significantly improving decision-making accuracy and reducing the time required to process complex queries. It also highlights the scalability of AI-driven solutions for broader adoption in the insurance industry.

However, certain drawbacks were identified which can serve as areas for further development down the line. The chatbot relies on a predefined knowledge base which requires regular updates. Additionally, semantic intent extraction, while accurate, may occasionally misclassify highly ambiguous queries, impacting the system's reliability.

5 Further Improvements

To address the identified drawbacks, several enhancements can be implemented in the future.

First, automating the process of updating the knowledge base with real-time data from insurance providers can ensure the chatbot remains accurate and up-to-date. Integration with APIs from insurance companies could streamline this process. Second, improving the intent extraction module through further fine-tuning of the SentenceTransformer model or employing hybrid approaches can help to further minimize misclassification of ambiguous queries. Finally, expanding the system's capabilities to handle multilingual queries and providing adaptive responses based on user preferences would enhance accessibility and inclusivity, making the chatbot usable for a broader audience.

By addressing these areas, the system can evolve into a more robust, scalable, and reliable tool for navigating travel insurance options.

6 References

1. Moneylife. (2023). Travel insurance tips: How to select the perfect policy for your needs. Retrieved from https://www.moneylife.in/article/travel-insurance-tips-how-to-select-the-perfect-policy-for-your-needs/75403.html on 18 December 2024

2. Yellow.ai. (2023). Chatbot for insurance agencies - Benefits & examples. Retrieved from https://yellow.ai/blog/insurance-chatbot/ on 19 December 2024

3. Engagely.ai. (2024). Retrieval-Augmented Generation (RAG): A Game-Changer in Customer Experience. Retrieved from https://www.engagely.ai/retrieval-augmented-generation-rag-a-game-changer-in-customer-experience/ on 23 December 2024

4. Cognigy.com. (2024). AI Chatbots for Insurance: Benefits & Examples. Retrieved from https://www.cognigy.com/blog/ai-chatbots-for-insurance on 26 December 2024

5. Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. Retrieved from https://arxiv.org/abs/2404.17723 on 26 December 2024

6. SimplyIOA. (2024). The Ethical Implications of AI in Insurance. Retrieved from https://stories.simplyioa.com/the-ethical-implications-of-ai-in-insurance on 27 December 2024

7. Finance Magnates. (2023). The Ethics of AI in Insurance: Balancing Efficiency and Fairness. Retrieved from https://www.financemagnates.com/fintech/education-centre/the-ethics-of-ai-in-insurance-balancing-efficiency-and-fairness/

8. Insurance Innovation Reporter. (2024). The Ethics of AI in Commercial Insurance: How to Approach this Revolutionary Technology Responsibly. Retrieved from https://iireporter.com/the-ethics-of-ai-in-commercial-insurance-how-to-approach-this-revolutionary-technology-responsibly/ on 2 January 2025

9. ScribbleData. (2024). Driving Sustainability in Insurance: The Role of Generative AI. Retrieved from https://www.scribbledata.io/blog/driving-sustainability-in-insurance-the-role-of-generative-ai/ on 4 January 2025