Simple RAG-Based Insurance QA Project

1. Introduction

This project focuseson building a Retrieval-Augmented Generation (RAG) system for answering questions about Anadolu Insurance (Kasko) policy documents. The main goal is to reduce hallucinations and provide document-grounded answers by combining large language models with vector-based document retrieval.

---

2. What the System Does

- Takes an insurance-related question
- Searches the PDF document
- Tries to answer using the found text

If the information is not found, the system says so.

---

3. How It Was Built

- The insurance PDF was loaded
- The text was split into small parts
- These parts were saved in a vector database
- A language model was used to answer questions

That’s basically it.

---

4. Retrieval Process

When a question is asked: - The system searches for similar text in the document - A few related text parts are returned - The answer is generated using those parts

Sometimes the results are good, sometimes not.

---

5. Models Used

Two models were tested: - Groq (LLaMA-based) - Google Gemini

Both were used in the same way. No special tuning was done.

6. Testing

Some example questions were asked manually. Later, a small benchmark was run.

The benchmark was limited because: - API limits were reached - Running too many questions caused errors

---

7. Problems Faced

During the project, commonstudent-level problems were faced:

- API limits were exceeded
- Some requests failed
- The model sometimes did not follow instructions
- Debugging took time

Most problems were solved by retrying or simplifying the setup.

---

8. Limitations

- Answers depend heavily on the PDF quality
- Some answers are incomplete
- No advanced evaluation was done

---

9. Conclusion

This project helped understand how AI Agents work under the hood. It is not perfect, but it works well enough for learning purposes.

Overall, the project achieved its goal of experimenting with document-based question answering.

Yusuf Yiğit Gültekin