

Live Demo Url -

<https://www.loom.com/share/e41176810e6647ba9a4004a80b7ece6e?sid=9c8c38ce-9a84-4a0c-9905-37a81a3fc00a>

AMLGOLABS Project Report

With Integrated Demo Responses and Example Interactions

1. Document Structure and Chunking Logic

The AMLGOLABS system processes user-uploaded documents in PDF, DOCX, or TXT format placed in the `data/` directory. Text extraction is handled by PyPDF2 (PDF) or python-docx (DOCX), followed by cleaning to remove extraneous spaces and unwanted characters.

Chunking is performed using the NLTK sentence tokenizer, splitting the cleaned text into coherent segments of 100–300 words. Each chunk ends at a sentence boundary to ensure context and readability. The resulting chunks are stored as JSON objects—each with a unique ID and text—in `chunks/chunks.jsonl`. This structure supports efficient retrieval and accurate context delivery during question answering.

2. Embedding Model and Vector Database

The system uses the `all-MiniLM-L6-v2` model from SentenceTransformers to generate dense vector embeddings for each chunk. These embeddings capture the semantic meaning of the text, enabling effective similarity search.

All chunk embeddings are indexed using FAISS, a high-performance vector search library. The FAISS index and the chunk-to-ID mapping are stored in the `vectordb/` directory as `faiss.index` and `chunk_mapping.json`. This setup allows for rapid retrieval of relevant chunks in response to user queries, even with large document sets.

3. Prompt Format and Generation Logic

When a user submits a query, it is embedded using the same embedding model. The system retrieves the top-k (default: 3) most relevant chunks from the FAISS index.

A prompt is then constructed for the Gemini LLM in this format:

You are an AI assistant. Use ONLY the following sources to answer the user's question.

[Source 1: ...]

[Source 2: ...]

...

User question: [query]

Answer (cite sources as Source 1, Source 2, etc. if used):

The Gemini 2.0 Flash model (via Google Generative AI API) generates the answer, which is returned to the user along with the cited sources. This ensures transparency and allows users to verify the factual basis of each answer.

4. Example Queries and Responses (from Demo)

Below are real interactions with the bot, demonstrating both successful retrieval and the system's ability to cite sources:

Q1: Tell me about eBay from the Document

Bot:

eBay offers access to and use of its Services based on the terms outlined in the User Agreement (Source 1). The User Agreement, Mobile Application Terms of Use, and all policies and additional terms posted on eBay's sites, applications, tools, and services are collectively referred to as "Services" (Source 1). eBay is not a party to contracts for sale between third-party sellers and buyers, nor is eBay a traditional auctioneer (Source 3). eBay may use AI-powered tools to provide and improve their services, but the availability and accuracy of these tools are not guaranteed (Source 3). eBay may help facilitate the resolution of disputes between buyers and sellers through various programs (Source 3). Users are responsible for the content they provide and must ensure it is accurate, appropriate, legal, and does not infringe any Intellectual Property Rights (Source 2).

Q2: Tell me about Using eBay

Bot:

Using eBay requires you to agree to comply with the User Agreement, eBay's policies, terms, and all applicable laws, rules, and regulations (Source 3). You agree not to breach any laws, regulations, third-party rights, or eBay's systems, services, or policies (Source 3). You cannot use eBay if you are unable to form legally binding contracts, are suspended from using eBay, or are a person with whom transactions are prohibited under economic or trade sanctions (Source 3). You must pay for items you purchased unless you have a valid reason (Source 3). You must deliver items you sold unless you have a valid reason (Source 3). You are not allowed to manipulate the price of any item or interfere with any other user's listings (Source 3). You cannot take any action that may undermine the feedback or ratings systems (Source 3). Transferring your eBay account to another party without consent is prohibited (Source 3). Sharing your login credentials with third parties is prohibited (Source 3). eBay is a marketplace that allows users to offer, sell, and buy goods and services (Source 1).

Q3: Tell me about Policy Enforcement From Document

Bot:

When a buyer or seller issue arises, the platform may consider the user's performance history and the specific circumstances when applying its policies. It may choose to be more lenient with policy enforcement in an effort to do the right thing for both buyers and sellers (Source 1). However, this does not limit their right to refuse, modify, or terminate all or part of their Services to anyone, or to terminate the agreement with anyone, for any reason at their discretion (Source 1).

5. Additional Example Scenarios (Success and Failure Cases)

Success:

- The bot accurately summarizes eBay's User Agreement and usage policies, referencing the correct chunks and sources.
- The bot explains policy enforcement, including eBay's discretion and flexibility, with proper citation.

Partial/Failure:

- If a user asks about topics not present in the document (e.g., eBay's cryptocurrency policy), the bot may either state that the information is not available or, if not properly constrained, hallucinate an answer.
- Questions unrelated to the document (e.g., "What is the weather like in San Jose?") may result in hallucinated responses unless the prompt and retrieval logic are robustly designed.

6. Notes on Hallucinations, Model Limitations, and Slow Responses

- The Gemini LLM can hallucinate or fabricate information if the prompt or retrieved context is insufficient or ambiguous.
- If the document does not contain the answer, the model may still attempt to answer; users should always verify with the cited sources.
- The system's accuracy is limited by the quality and coverage of the input document. Missing topics cannot be answered reliably.
- Slow responses can occur due to network latency with the Gemini API or when processing large documents, which can affect retrieval speed.
- Chunking logic ensures context is coherent, but very large or very small chunks may impact retrieval accuracy.

7. System Dependencies and Usage**Dependencies:**

- Python libraries: streamlit, faiss-cpu, sentence-transformers, google-generativeai, nltk, PyPDF2, pandas, python-dotenv, requests
- Data files: Place your document in the data/ directory as PDF, DOCX, or TXT.
- Chunks and vector DB are stored in chunks/ and vectordb/ respectively.

Usage:

- Run the Streamlit app (app.py) to interact with the chatbot.

- Enter your Gemini API key when prompted.
- Ask questions about the uploaded document; the system will display answers with cited sources for verification.

8. Recommendations and Additional Insights

- **Best Practices:** Upload comprehensive, well-structured documents to maximize answer accuracy.
- **Transparency:** Always review cited sources to confirm the validity of the model's responses.
- **Extensibility:** The modular design allows for easy updates to embedding models or LLMs as technology evolves.
- **User Guidance:** Encourage users to ask specific, document-relevant questions for best results.