

Step 1: Resume Processing

- I wrote a script (process_resumes.py) to extract text from PDF resumes using pdfminer.six.
- The text is split into smaller chunks (eg 1000 characters each) to make it easier to process. “Ely 2oltlk 3leha”
- The chunks are saved in a file called chunks.pkl.

Step 2: Embedding Generation

- I used a pre-trained model (all-MiniLM-L6-v2) from the sentence-transformers library to convert the text chunks into vector embeddings.
- The embeddings are saved in a file called embeddings.pkl.

Step 3: FAISS Index Creation

- I used the faiss library to create a vector index from the embeddings.
- The index is saved in a file called resumes_index.faiss.

Step 4: Chatbot Interface

- I built a chatbot interface using gradio.
- The chatbot allows users to query the resumes using natural language (eg "Find candidates with experience in Python").
- The chatbot retrieves the most relevant resume chunks based on the query.

How It Works

1. **Input:** PDF resumes are placed in the resumes directory.
2. **Processing:**
 - The text is extracted and split into chunks.
 - The chunks are converted into embeddings.
 - The embeddings are stored in a FAISS index.
3. **Querying:**
 - Users can ask questions in natural language.

- The chatbot retrieves the most relevant resume chunks and displays them.
-

Tools and Technologies Used

- **Python:** For scripting and automation.
 - **pdfminer.six:** For extracting text from PDFs.
 - **sentence-transformers:** For generating embeddings.
 - **FAISS:** For storing embeddings and performing similarity search.
 - **Gradio:** For building the chatbot interface.
-

Challenges Faced

- **Dependency Issues:** Some libraries (eg email-validator) caused errors, but I resolved them by upgrading the packages.
 - **File Path Issues:** I had to ensure that all files (eg resumes_index.faiss) were in the correct directory.
-

Next Steps

- **Improve the Chatbot:**
 - Add filters (eg by skills, experience level).
 - Use a larger embedding model for better accuracy.
- **Deploy the System:**
 - M3rfsh ezay lesa bs 7asa me7tagen n3ml 7aga a7sn mn keda
- **Add More Resumes:**
 - Test the system with a larger dataset of resumes. “Lw 3rfna”