**DOCUMENTATION for PDF Based Chatbot**

**A. How the Chatbot Processes User Queries:**

The chatbot processes queries by following these steps:

1. **User Input**: The user enters a question into the chatbot interface.
2. **Fuzzy Matching**: The chatbot uses the find\_answer\_in\_docs function, which leverages the fuzz.partial\_ratio from the fuzzywuzzy library to match the user’s query with questions extracted from the knowledge base.
3. **Score Evaluation**: The chatbot determines the most relevant Q&A pair by evaluating the similarity score. If the score exceeds a defined threshold (e.g., 70), the chatbot responds with the closest match.
4. **Source Attribution**: The chatbot provides the answer along with the source document name to maintain transparency.
5. **Suggestions**: If the chatbot cannot find an exact match, it suggests possible related questions to guide the user.

**B. How to Update and Integrate New Knowledge Base Documents:**

1. **Document Integration Process**:
   * If you have new PDF files, use the extract\_qa\_from\_pdf function to extract Q&A pairs and add them to the knowledge\_base dictionary.
   * For example, to add a new document:

**pdf\_text\_3 = extract\_qa\_from\_pdf('new\_document.pdf')**

**knowledge\_base['Document 3'] = pdf\_text\_3**

1. **Updating Knowledge Base**:
   * You can dynamically update the knowledge base without modifying the chatbot's core logic. Simply extract data from new or updated PDFs and add them to the existing knowledge\_base dictionary.
   * The system doesn’t require major reconfiguration when new documents are added since the chatbot automatically works with the updated knowledge\_base.

**C. Ensure Handling of Document Updates Without Reconfiguration**:

* **Scalability**: The chatbot’s design allows you to scale by loading multiple documents and extracting their content seamlessly. Since the logic for extracting Q&A pairs is generalized, you can easily plug in new documents.
* **Modular Design**: The extract\_qa\_from\_pdf function and the knowledge base structure are independent of specific files. As long as new PDFs follow a similar Q&A format, they can be added without changing the chatbot’s logic.