Saheta is capable of reading PDF documents and engaging in interactive communication with users. It can also accept user inputs such as name, phone number, email, and appointment date. Phone numbers and email addresses are validated to ensure proper formatting. If the input does not meet the required format, the data will not be saved.

All saved user information is stored in JSON format, making it easily accessible across other platforms.

For the language model, I used *Llama 3.2-1b* as it is a lightweight and efficient option compared to larger models like *Llama 70b*, which could not run on my laptop due to its large size and computational requirements. During this task, I explored models from Hugging Face but encountered issues with vectorization. I also experimented with ChatGroq, but it required a significant amount of time to convert documents into vector embeddings. Ultimately, I decided to use FAISS for its efficiency and reliability.

The embedding process was powered by the Ollama Embeddings model (`model="llama3.2:1b"`), which enabled effective document vectorization. The vector embeddings were stored using FAISS, and retrieval was handled through a retriever with top-k query capabilities, making the system efficient for real-time interaction.

For the user interface, I utilized Streamlit to provide a clean and interactive experience. Validation for email and phone numbers was implemented using regular expressions, and date parsing was handled using a date parser.

This project allowed me to explore cutting-edge technologies and make informed decisions to ensure efficiency and accuracy in delivering the desired functionality.