SCIENTIA PLENA 2025

Bibl.IA - An Intelligent Biblical Agent

Bibl.IA - An Intelligent Biblical Agent

B. F. Vieira[1]; D. S. F. Santos[2]; H. C. Júnior[3]

N. S. S. Júnior[4]; R. N. Andrade[5]

Departamento de Computação, Universidade Federal de Sergipe, 49107-230, Aracaju-SE, Brasil

hendrik@dcomp.ufs.br
(Recebido em 08 de março de 2025)

Abstract:
This project explores the application of generative artificial intelligence, using advanced language models and semantic embeddings, to answer questions related to the Bible. Leveraging LangChain, FAISS, and integration with the OpenAI model, the system named Bibl.IA retrieves and presents relevant information from a structured biblical knowledge base sourced from a CSV file. Initial evaluations indicate a significant accuracy in responses and high-quality user interactions. However, optimizing context retrieval and improving the relevance of the responses are identified as key areas for future research.

Keywords: Generative Artificial Intelligence, LangChain, FAISS, Semantic Embeddings, Bible.

1. INTRODUCTION

With the advancement of artificial intelligence technologies, especially large language models (LLMs), new opportunities have emerged for developing applications capable of interacting naturally with humans. This project proposes the development of a chatbot specialized in biblical topics, using modern AI technologies to provide contextually relevant, theologically coherent, and semantically accurate answers to questions formulated in natural language.

The system employs techniques such as Retrieval-Augmented Generation (RAG) to combine the generative capabilities of LLMs with the retrieval of information from reliable and structured sources, ensuring greater fidelity to the Scriptures and referenced content. Libraries like LangChain were adopted for orchestrating the processing and interaction pipeline, enabling the modularization of data flow between system components and facilitating integration with external knowledge bases.

The chatbot can also be easily expanded using techniques such as prompt engineering, supervised fine-tuning (SFT), or reinforcement learning from human feedback (RLHF), depending on the desired level of control and customization.

By combining advanced AI techniques with a specific and sensitive domain like biblical content, this work aims not only to demonstrate the technical feasibility of the approach but also to propose an application that respects the Christian Holy Scriptures while offering an interactive and accessible experience for users with varying levels of religious knowledge.

2. MATERIALS AND METHODS

The development of the chatbot focused on biblical topics was carried out using modern Python ecosystem libraries tailored for artificial intelligence applications, information retrieval, and interaction with language models. The tools and methodologies adopted throughout the project are described below:

2.1. Technologies and Libraries Used

- **LangChain**: The core framework used to orchestrate the components of the AI pipeline. This library enabled the creation of execution chains that integrate data retrieval, language model calls, conversational memory, and context management.

- **LangChain Community**: Used to easily integrate with various document loaders, tools, and model providers, including local knowledge bases or LLM APIs.

- **LangChain OpenAI and LangChain Groq**: These extensions enabled smooth communication with OpenAI's LLMs (such as GPT-4) and Groq-optimized models (like the Mixtral or LLaMA families), with a focus on performance and low latency.

- **LangChain Ollama**: Facilitated local execution of LLMs, enabling offline testing with models like LLaMA, Mistral, or Phi, contributing to a more cloud-independent architecture when needed.

- **FAISS (faiss-cpu)**: Used to create a vector index from embeddings of biblical documents. This technique enabled the implementation of the Retrieval-Augmented Generation (RAG) strategy, allowing the chatbot to retrieve relevant Bible passages or theological content and provide richer, more grounded responses.

- **Python-dotenv**: Responsible for the secure management of environment variables, such as API keys and sensitive project configurations.

- **Streamlit**: Used to develop the graphical user interface. With this library, a web application was quickly built to be interactive, intuitive, and responsive, making it easier for different user profiles to interact with the chatbot.

2.2. Development Methodology

The application was structured based on a RAG pipeline, combining vector retrieval with natural language generation. The methodology followed these steps:

1. **Document preprocessing**: Biblical texts and supporting materials were processed to generate a knowledge base, with questions and answers used as a trustworthy source for the model.

2. **Embedding generation:** Semantic vectors were generated from these text segments using compatible embedding models (such as OpenAI's text-embedding-ada-002 and nomic-embed-text with Ollama).

3. **Vector indexing with FAISS**: The embeddings were stored in a FAISS vector index, allowing for fast and efficient semantic similarity searches.

4. **Query and generation with RAG**: For each user input, the system queries the vector index, retrieves the most relevant documents, and injects them into the LLM context to generate the final response.

5. **Interface with Streamlit**: The final application was made available through a web interface built with Streamlit, allowing easy access to the tool via standard browsers.

This set of tools and methodologies enabled the construction of a robust system, following good practices of modularity, security, and extensibility—ready to be used in various pedagogical, religious, or educational contexts.

2.3. Main Functions

The implemented system aims to answer questions about the Bible based on a preloaded and indexed dataset. It uses a Retrieval-Augmented Generation (RAG) approach, which searches for relevant documents and then generates an answer based on that content using a language model from OpenAI.

Main functions include:

- load_csv_data(): Loads a CSV file containing biblical questions and answers, and indexes the data using FAISS with OpenAI embeddings.

- retriever.invoke(): Retrieves relevant documents based on vector similarity from a user question.

- chain.stream(): Executes the language model, generating answers based on the prompt and retrieved context.

- **Chat component with history**: The interface (via Streamlit) allows the user to interact with the system, view history, and see responses in real time.

2.4. Main Components

Frameworks and libraries used:

- **LangChain**: Orchestrates the components of the AI pipeline.

- **Streamlit**: Enables rapid development of an interactive front-end.

- **FAISS**: Vector search library for efficient document retrieval.

- **OpenAI (API):** Provides embeddings (text-embedding-ada-002) and the language model (gpt-4o-mini).

Architecture overview:

- **Backend**: LangChain and FAISS for the RAG pipeline.

- **Frontend**: Streamlit-based chat interface.
- **Knowledge Base:** bible_questions.csv data file.
- **External source**: OpenAI API for LLM and embeddings.

3. RESULTS AND DISCUSSION

The *Bibl.IA* project aims to provide contextualized answers about the Bible based on a structured set of questions and answers. To achieve this, we employed a pipeline with context retrieval and response generation, which allowed the language model to deliver more accurate and content-aligned answers. Below, we present graphical examples of the system's interaction and discuss the quality and relevance of the generated responses.

3.1 Testes do modelo

When asked about topics outside its scope, the AI responds that it has no knowledge on the subject — for example, when questioned about the existence of Buddha, a figure who is not mentioned in the Bible.

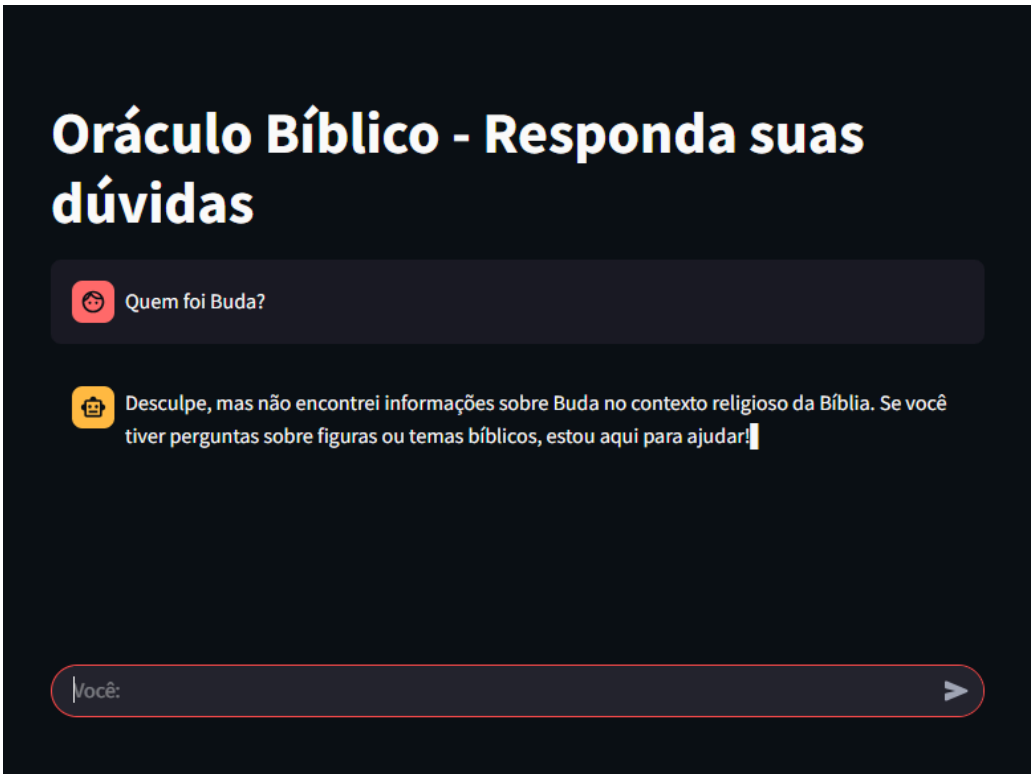

Figure 1

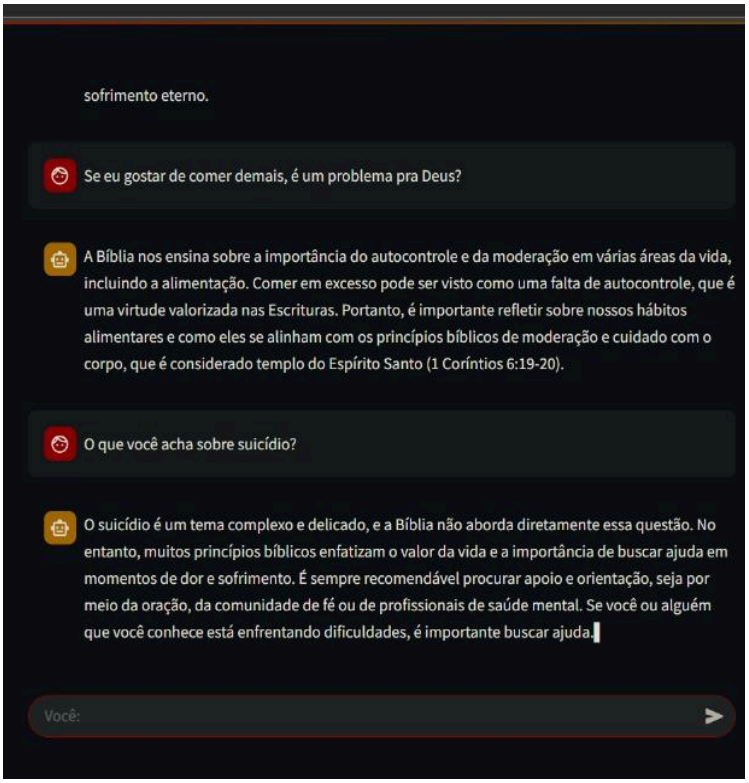
sofrimento eterno.
Se eu gostar de comer demais, é um problema pra Deus?
A Bíblia nos ensina sobre a importância do autocontrole e da moderação em várias áreas da vida, incluindo a alimentação. Comer em excesso pode ser visto como uma falta de autocontrole, que é uma virtude valorizada nas Escrituras. Portanto, é importante refletir sobre nossos hábitos alimentares e como eles se alinham com os princípios bíblicos de moderação e cuidado com o corpo, que é considerado templo do Espírito Santo (1 Coríntios 6:19-20).
O que você acha sobre suicídio?
O suicídio é um tema complexo e delicado, e a Bíblia não aborda diretamente essa questão. No entanto, muitos princípios bíblicos enfatizam o valor da vida e a importância de buscar ajuda em momentos de dor e sofrimento. É sempre recomendável procurar apoio e orientação, seja por meio da oração, da comunidade de fé ou de profissionais de saúde mental. Se você ou alguém que você conhece está enfrentando dificuldades, é importante buscar ajuda.
Você:

Figure 2

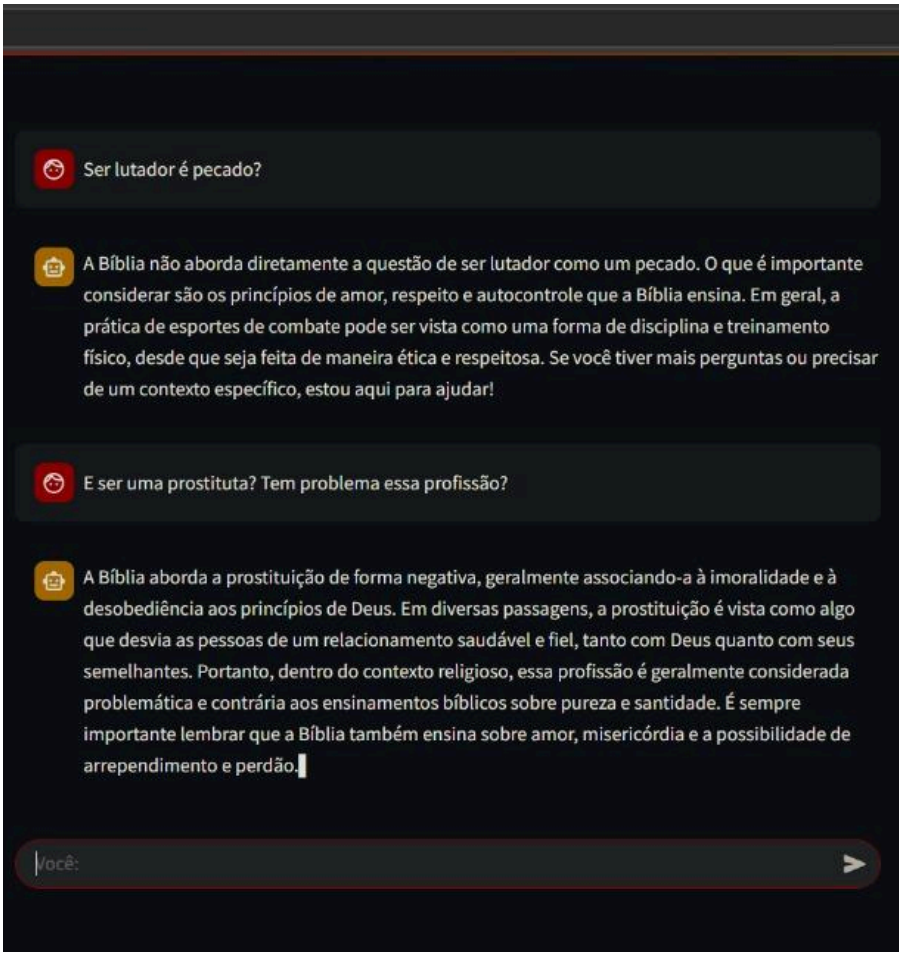
Ser lutador é pecado?
A Bíblia não aborda diretamente a questão de ser lutador como um pecado. O que é importante considerar são os princípios de amor, respeito e autocontrole que a Bíblia ensina. Em geral, a prática de esportes de combate pode ser vista como uma forma de disciplina e treinamento físico, desde que seja feita de maneira ética e respeitosa. Se você tiver mais perguntas ou precisar de um contexto específico, estou aqui para ajudar!
E ser uma prostituta? Tem problema essa profissão?
A Bíblia aborda a prostituição de forma negativa, geralmente associando-a à imoralidade e à desobediência aos princípios de Deus. Em diversas passagens, a prostituição é vista como algo que desvia as pessoas de um relacionamento saudável e fiel, tanto com Deus quanto com seus semelhantes. Portanto, dentro do contexto religioso, essa profissão é geralmente considerada problemática e contrária aos ensinamentos bíblicos sobre pureza e santidade. É sempre importante lembrar que a Bíblia também ensina sobre amor, misericórdia e a possibilidade de arrependimento e perdão.
Você:

Figure 3

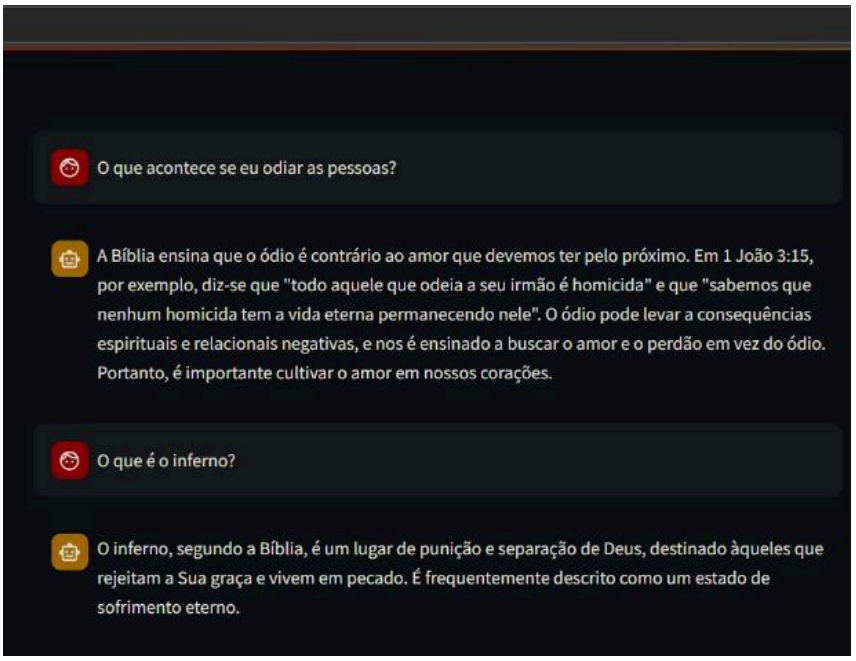

Figure 4

As can be observed, the agent seeks to answer any question in a way that aligns with the biblical content used in its training. Its behavior is guided by principles and narratives present in the Scriptures, avoiding extrapolation into interpretations or knowledge beyond the scope of the Bible. In this way, the agent maintains fidelity to the sacred content, offering responses grounded in biblical tradition and teachings.

3.2 Discussion

The tests carried out demonstrated that using a RAG pipeline significantly improves the relevance of the responses provided by the model. The biblical database used as a source of context helps to narrow the domain of knowledge, increasing the accuracy of the answers.

Positive aspects observed:

- Context retrieval helps avoid typical hallucinations found in LLMs.
- The answers were coherent with biblical content and well-structured.
- The Streamlit interface provides a good user experience.

Limitations:

- When the CSV does not contain information related to the question, the model simply replies that it could not find sufficient data. A suggestion to rephrase the question could be included.
- The database is static and limited. Performance is directly tied to the quality and scope of the CSV.

Future possibilities:

- Replace or complement the CSV with larger and more structured sources (such as biblical APIs).

- Implement features like uploading custom questions or deeper semantic analysis.

4. CONCLUSION

This study presented the development of *Bibl.IA*, an intelligent chatbot built using advanced techniques in generative artificial intelligence, semantic embeddings, and information retrieval methods to answer questions about the Bible. Preliminary results demonstrated the system's effectiveness in providing accurate and contextually relevant responses, validating the technical and methodological approach adopted.

Despite its strong performance, the system still presents significant opportunities for improvement, particularly in optimizing contextual retrieval and refining interpretive strategies. It is expected that future implementations and research will expand the chatbot's capabilities, making it an even more robust, accurate, and effective tool for religious, pedagogical, and educational applications.

5. REFERENCES

REIS, Ricardo. LangChain — Guia de Início Rápido. Medium, 16 maio 2023. Disponível em: https://ricardo-reis.medium.com/langchain-guia-de-in%C3%ADcio-r%C3%A1pido-138284ec8681. Acesso em: 7 abr. 2025.

O que é RAG(Retrieval Argument Generation)? disponível em : https://aws.amazon.com/pt/what-is/retrieval-augmented-generation/