

# Federal Constitution using LLM and RAG

E. S. Ramos ; C. C. Aquino ; R. P. Leite ; L. S. Lima ; J. F. F.  
Neto

*<sup>1</sup>Department of Computing / Laboratory D02, Federal University of Sergipe, 49100-000, São  
Cristóvão–Sergipe, Brazil'*

*{edcarlosufs , riquelmewin , cainacastro29 , falcao , waynewyn } @ academico.ufs.br*

---

The use of LLMs (Large Language Models) has become increasingly common across various fields of knowledge, while potentially hallucinated information generated by these tools has also begun to spread. This raises concerns about the reliability of the responses produced, especially in sensitive contexts such as education, law, and healthcare. Nevertheless, LLMs remain valuable allies in task automation, content generation, and research support, as long as they are used with critical thinking and proper validation of the information provided. The current challenge is to find a balance between leveraging the potential of these tools and mitigating their risks, ensuring the ethical and responsible use of generative artificial intelligence.

This work aims to align the generalized knowledge of an LLM with a specialized knowledge base in the field of constitutional rights, in order to disseminate reliable information about the foundation of citizenship in Brazil.

---

## 1. Introduction

The development of Large Language Models (LLMs) has become a widely explored research field within Artificial Intelligence (AI), especially after the success of models developed by OpenAI, such as ChatGPT—an artificial intelligence system designed to answer questions on a wide range of topics across various fields of knowledge. This breakthrough has driven the emergence of many other LLMs, sparking market interest and leading to the adaptation of these tools for specific needs in corporate environments, process automation, and customer service. Today, these technologies are already widely integrated into various applications and services available to the public [1].

In addition to assisting in tasks such as text generation, translation, automatic summarization, and conversational support, these tools are also increasingly used as sources of information by millions of users. However, a significant challenge persists: the tendency of these models to produce “hallucinated” responses—that is, false, inaccurate, or entirely fabricated information [2]. This limitation can be especially critical in contexts where the accuracy of information is essential, such as legal, medical, or educational consultations. In Brazil, this issue is compounded by the population’s low level of familiarity with the country’s own Federal Constitution [3]. Many citizens turn to AI-based tools for guidance regarding their rights and obligations, but due to the possibility of receiving incorrect information, they become vulnerable to misinformation and may even contribute to its spread.



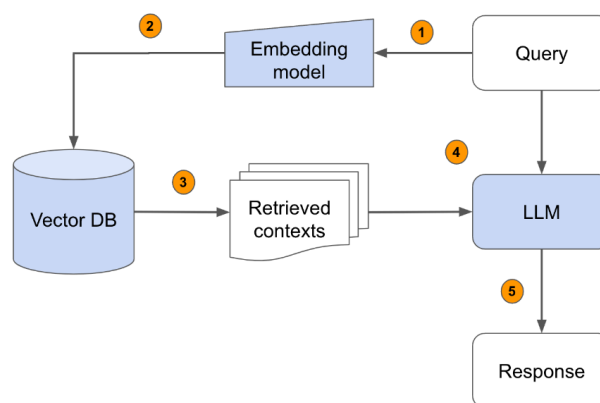
A promising approach to mitigate the hallucination problem is the use of a technique known as Retrieval-Augmented Generation (RAG). This method combines the language generation capabilities of LLMs with the retrieval of verified information from external and reliable sources, such as PDF files, Word documents, CSV spreadsheets, or structured databases. Instead of relying solely on the model's pre-trained knowledge, the system dynamically queries these sources during response generation, offering greater accuracy, traceability, and contextual relevance.

Therefore, the goal of this project is to enhance the knowledge of a pre-trained LLM with a specialized knowledge base—specifically, the text of the Brazilian Federal Constitution—using the RAG technique. The ultimate objective is to create an intelligent chatbot, hosted on a WhatsApp user profile, capable of answering questions related to the articles and laws within the Constitution. This solution has the potential to democratize access to legal knowledge, combat misinformation, and promote informed citizenship, while also exploring the advanced capabilities of artificial intelligence in a socially relevant context.

## 2. Methodology

### 2.1 Application architecture

- a. **Query Frontend** (end user interacts via WhatsApp).
- b. **RAG Module**: semantic retrieval engine + generation:
  1. Transformation of the Constitution PDF file into Markdown using Docling.
  2. Splitting by headings.
  3. Smart re-splitting to divide overly large chunks and concatenate small ones.
  4. Embedding of the chunks using “text-embedding-3-small”.
  5. Storage in a vector database.
- c. **Vector Database**: responsible for storing and retrieving embeddings.
  1. Stores all embeddings.
  2. Returns embeddings using similarity search.
  3. Adds the search result to the prompt before sending it to the LLM.



- a. **Answer the user.**

---

## 2.2 Pseudocode for the Main Algorithms

- Reading the Document.

```

FUNCTION process_pdf_to_markdown(pdf_path):
  raw_text = Docling.extract_text(pdf_path)
  markdown = add_markdown_headers(raw_text)
  RETURN markdown

```

- Splitting into smaller segments.

```

FUNCTION split_by_headers(markdown):
  splitter = MarkdownHeaderTextSplitter(
    headers_to_split_on=[("##", "Section")]
  )
  chunks = splitter.split_text(markdown)
  RETURN chunks
FUNCTION rebalance_chunks(chunks):
  final_chunks = []
  skip_next = False
  FOR EACH chunk IN chunks:
    IF skip_next:
      skip_next = False
      CONTINUE
    IF chunk.length > 2000:
      split_chunks = RecursiveTextSplitter(
        chunk_size=1100,
        separators=["\n\n", "\n", "."],
        keep_separator=True
      ).split_text(chunk.content)
      FOR split_chunk IN split_chunks:
        final_chunks.append({
          "content": split_chunk,
          "metadata": chunk.metadata,
          "length": len(split_chunk)
        })
    ELIF chunk.length < 500:
      IF NOT is_last_chunk(chunk, chunks):
        merged_chunk = merge_with_next(chunk, chunks)
        final_chunks.append(merged_chunk)
    ELSE:
      final_chunks.append(chunk)
  ELSE:
    final_chunks.append(chunk)
  RETURN final_chunks
FUNCTION merge_with_next(current_chunk, all_chunks):
  next_index = all_chunks.index(current_chunk) + 1
  next_chunk = all_chunks[next_index]
  RETURN {
    "content": f"{current_chunk.content}\n\n{next_chunk.content}",
    "metadata": combine_metadata(current_chunk.metadata,
next_chunk.metadata),
    "length": current_chunk.length + next_chunk.length
  }
  skip_next = True

```

```
}
```

- Embedding generation using a transformer model..

```
FUNCTION generate_embeddings(final_chunks):
    embedder = OpenAIEmbeddings(model = "text-embedding-3-small")
    texts = [chunk.content FOR chunk IN final_chunks]
    vectors = embedder.embed_documents(texts)
    metadata = [
        {
            "header": chunk.metadata.Section,
            "length": chunk.length
        }
        FOR chunk IN final_chunks
    ]
    RETURN vectors, metadata
```

- Storing embeddings in the vector database along with metadata.

```
FUNCTION store_in_vectordb(vectors, metadata, texts):

    vectordb = ChromaDB(

        persist_path = "path/to/storage",
        collection_name = "constitution"
    )
    vectordb.add(
        embeddings = vectors,
        documents = texts,
        metadata = metadata
    )
    RETURN vectordb
```

---

## 2.3. Used dataset

- **Data Source:**
  - **Brazilian Federal Constitution [4]**
- **Corpus Structure:**
  - Split into chunks larger than 500 and smaller than 2000 characters for compatibility.

---

## 2.4. Frameworks, Libraries and Tools

- **LLM and Embeddings:**

- LLM : gpt-4o-mini
  - Embedding : text-embedding-3-small
  - **Vector Database:**
    - ChromaDB
  - **Language and Environment:**
    - Python ( 3.10+).
    - Most used Frameworks and Libraries: LangChain, Transformers , Docling and Flask.
    - All requirements are in the requirements.txt in repository.
  - **Communication interface:**
    - API from Whatsapp, Openai.
- 

## 2.5. Hardware

- CPU: AMD Ryzen 5 5500U with Radeon Graphics × 6
- RAM: 20 GB DDR4
  - PLACA DE VÍDEO: Advanced Micro Devices, Inc. [AMD/ATI] Lucienne

### 3. Conducted tests

**1 Q:** “Can the President of the Republic order the intervention of the Armed Forces in a state’s public security operations without a request from the state government, in the event of a ‘prolonged crisis’ not declared as a state of defense?”

**R (GPT-4 Mini):** “In Brazil, the intervention of the Armed Forces in public security matters without a request from the state government is a complex issue and depends on the legal context. According to the Federal Constitution, for the Armed Forces to intervene in a public security situation, there is usually a need for a request from the state government or that the action be grounded in situations of serious disturbance of public order.

However, the Constitution also provides that the President of the Republic may, in cases of ‘extreme crisis’ or emergency situations, order interventions. This is generally applied in cases where there is a threat to public order that is not adequately resolved by local authorities.

Therefore, the use of the Armed Forces in public security operations without a state request may occur, but only in exceptional situations and based on specific legal norms, considering the complexity of the legal and constitutional aspects involved. The absence of a formal declaration of a state of defense or siege would complicate the situation and could generate legal debates about the legality of the action. In summary, although the President may have this possibility in extreme cases, the issue is legally delicate and must follow the established legal precepts.”

**R (RAG + GPT-4 Mini):** “According to the provided documents, the President of the Republic cannot order the intervention of the Armed Forces in public security operations of a state without a request from the state government, in the event of a ‘prolonged crisis’ not declared as a state of defense. The Constitution provides, in Article 136, that the President can declare a state of defense to preserve public order, but this requires a formal declaration and communication with the National Congress. For situations classified as crises but not qualifying as a state of defense or siege, there is no legal provision for intervention without the state's request.

Reference: Constitution, Article 136 (State of Defense).”

**2 Q:** “Can a public servant refuse to issue marriage certificates to same-sex couples, invoking religious conscientious objection, without violating the Constitution?”

**R (GPT-4 Mini):** “A public servant’s refusal to issue marriage certificates to same-sex couples, invoking religious conscientious objection, may conflict with constitutional principles, particularly the right to equality and non-discrimination guaranteed by the Brazilian Constitution.

The Federal Supreme Court (STF) has already ruled in favor of recognizing stable unions and marriage between same-sex couples, considering that sexual orientation should not be a reason for the denial of rights. Refusing to provide an essential public service, such as issuing marriage certificates, may constitute discrimination and violate the servant’s duty to act impartially and with respect for fundamental rights.

Although conscientious objection is a right recognized in certain situations, its application in public services—especially in cases involving fundamental civil rights—is complex and must be carefully considered. Generally, the protection of the rights of historically marginalized groups, such as the LGBTQ+ population, tends to prevail over personal objections of conscience that could lead to discrimination.

Therefore, the refusal may be deemed unconstitutional and contrary to the public servant’s duty to ensure equal access to all citizens. In this context, conscientious objection cannot be used as a valid justification for discrimination.”

**R (RAG + GPT-4 Mini):** “A public servant cannot refuse to issue marriage certificates to same-sex couples by invoking religious conscientious objection, as such a refusal would violate constitutional principles. The Federal Constitution, according to Article 226, ensures that civil marriage is guaranteed and celebrated free of charge, with no discrimination based on participants’ sexual orientation. Furthermore, Article 19, II, reinforces that

the public administration must act impartially and with respect for the rights of all citizens.”

**3 Q:** “Can civilians who commit crimes against military personnel during a Law and Order Guarantee (GLO) operation be tried by Military Justice, even without a link to the Armed Forces?”

**R (GPT-4 Mini):** Yes, civilians who commit crimes against military personnel during a Law and Order Guarantee (GLO) operation can be tried by Military Justice. This is because, during a GLO, the Armed Forces are mobilized to maintain public order, and crimes involving military personnel performing their duties can be considered under the jurisdiction of Military Justice, as established by Brazilian law. Jurisprudence and legal interpretation support that, in the context of the Armed Forces' activities, Military Justice may have jurisdiction over acts committed by civilians against military personnel.

**R (RAG + GPT-4 Mini):** Based on the provided documents, the answer is no. Civilians who commit crimes against military personnel during a Law and Order Guarantee (GLO) operation cannot be tried by Military Justice, as its jurisdiction is limited to the processing and judgment of military crimes defined by law, with no provision for trying civilians in these cases. This information was extracted from paragraphs 4 and 5, which address the jurisdiction of Military Justice concerning military personnel and judicial actions against military disciplinary acts, highlighting the exclusion of jury jurisdiction when the victim is a civilian.

#### 4. Results

- Reduction of Hallucinations
  - RAG eliminated incorrect answers in sensitive issues (e.g., Military Justice).
  - Without RAG, the LLM made mistakes in 1 out of 3 cases (33% severe inaccuracy).
- Solid Legal Foundation
  - All answers with RAG cite specific articles of the Constitution.
  - Without RAG, answers are generic and lack references.
- Clarity and Authority
  - RAG standardizes responses with:
    - Direct citations of the law.



- Logical structure (e.g., "According to Article X...").
- Without RAG, the llm tends to ramble and include assumptions.

LLM Metrics	With Rag	Without Rag
Articles citation	100%	0%
Legal accuracy	100%	66%

## 5. Conclusion:

Although it provided a correct answer based on the 1988 Constitution, on November 9, 2023, the Supreme Federal Court (STF) validated the trial of civilians by the Military Justice of the Union in cases of crimes committed against military personnel during Law and Order Guarantee (GLO) operations.[5]

Therefore, even when using an expert document on the subject, due to highly specific situations, the LLM with RAG produced an answer that was consistent with the information available to both the LLM and the retrieval system. However, because the recent legal update was not present in the retrieval base, the response was outdated and hallucinatory.

This limitation highlights one of the main challenges in using LLMs with RAG: the reliance on the timeliness and comprehensiveness of the dataset used for information retrieval.

## 6. References

1. PORTAL DA INDÚSTRIA. Como as empresas têm usado a inteligência artificial e qual o impacto no mercado de trabalho? Notícias, Educação, [S. l.], [2025]. Disponível em: <https://noticias.portaldaindustria.com.br/noticias/educacao/como-as-empresas-tem-usado-a-inteligencia-artificial-e-qual-o-impacto-no-mercado-de-trabalho/>. Accessed on: Apr. 8, 2025.
2. TECNOBLOG. Pesquisa mostra que chatbots estão errados em 60% do tempo. Notícias, [S. l.], 2025. Disponível em: <https://tecnoblog.net/noticias/pesquisa-mostra-que-chatbots-estao-errados-em-60-do-tempo/>. Accessed on: Apr. 7, 2025.
3. SENADO FEDERAL. Pesquisa DataSenado mostra que poucos conhecem realmente a Constituição do país. Notícias, Brasília, 25 out. 2013. Disponível em: <https://www12.senado.leg.br/noticias/materias/2013/10/25/pesquisa-datasenado-mostra-que-poucos-conhecem-realmente-a-constituicao-do-pais>. Accessed on: Apr. 7, 2025.
4. BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Supremo Tribunal Federal, 2024. Disponível em: [https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88\\_Livro\\_EC91\\_2016.pdf](https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf). Accessed on: Apr. 7, 2025.