



# Livro Resumo 2025

PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA

**XXXI Seminário Anual PIBIC**  
**XXII Seminário Anual PROBIC**  
**XV Seminário Anual PIBITI**



**26 e 27 de novembro de 2025**



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



# Sistema de Recuperação de Informação com uso de Large Language Models (LLM)

Marcelo Henrique da Cunha Ponciano, Mário Olímpio de Menezes  
IPEN - Instituto de Pesquisas Energéticas e Nucleares

## INTRODUÇÃO

Com o avanço dos Large Language Models (LLMs), surgem novas possibilidades para superar tais limitações. Os quais, segundo Zhai, quando treinados em grandes volumes de dados, os LLMs compreendem profundamente a linguagem e são capazes de aprimorar a indexação, reformular consultas ambíguas e até fornecer respostas diretas a partir de múltiplas fontes.<sup>1</sup>

Aplicados ao Reator Nuclear de Pesquisas IEA-R1, os LLMs podem agilizar o acesso a informações críticas, aumentar a segurança operacional e facilitar o treinamento de novos operadores. Assim, este projeto busca desenvolver um Sistema de Recuperação de Informação especializado para a documentação do reator, utilizando LLMs para superar as restrições dos métodos tradicionais de RI.

## OBJETIVO

O objetivo central deste projeto é validar e implantar um **portal de consulta da documentação crítica do Reator IEA-R1**, utilizando LLMs e interface gráfica. Os objetivos específicos são:

- **Analisar a documentação crítica** do reator e compreender seu uso por operadores, técnicos e gestores;
- **Avaliar diferentes modelos de LLM**, identificando suas potencialidades e aplicações para este tipo de ferramenta;
- **Implementar um protótipo especializado** com o modelo **Gemma**, que possibilita treinamento

local sem riscos de vazamento de informações;

- **Desenvolver a interface de busca**, permitindo consultas em linguagem natural e acesso direto a trechos relevantes da documentação;
- **Avaliar a usabilidade e o impacto** da ferramenta na rotina dos trabalhadores.

## METODOLOGIA

Este trabalho dá continuidade ao projeto de IT, que já avaliou modelos de LLMs e obteve resultados iniciais promissores com documentos de normas de proteção radiológica. A partir dessa base, as etapas propostas são:

1. **Análise e preparo da documentação crítica** do Reator IEA-R1, compreendendo seu conteúdo técnico, padrões de uso e pontos críticos de segurança.
2. **Teste de desempenho dos LLMs e técnicas de Recuperação de Informação**, aprofundando o entendimento das diferentes abordagens e suas particularidades.
3. **Implementação de um protótipo de SRI com LLMs**, utilizando treinamento local com o Ollama para evitar riscos de vazamento de informações sensíveis.
4. **Desenvolvimento de uma interface de busca amigável**, permitindo consultas em linguagem

natural, de forma escrita ou vocal.

5. **Avaliação e refinamento do sistema**, com métricas de precisão, revocação e F1-score, além de testes de usabilidade e coleta de feedback dos usuários para ajustes no modelo, na indexação e na interface, etapa crucial segundo Renato Souza<sup>2</sup>

## RESULTADOS

Até o presente estágio, o protótipo demonstrou ser capaz de:

- Processar documentos técnicos e convertê-los em formato manipulável.
- Gerar embeddings e indexar documentos de forma eficiente.
- Responder a consultas em linguagem natural com base nos documentos do reator.
- Apresentar as fontes utilizadas, garantindo rastreabilidade da resposta.

```
--- Sistema RAG pronto. Digite sua pergunta para começar. ---
--- Digite 'sair' a qualquer momento para fechar o programa. ---

> Digite sua pergunta: qual o limite de exposição anual para não servidores?

--- Processando a pergunta do usuário ---
Pergunta do usuário: "qual o limite de exposição anual para não servidores?"

--- Geração da Query Expansion ---
Original Query for Expansion: 'qual o limite de exposição anual para não servid
Queries used for retrieval: ['qual o limite de exposição anual para não servid
ho sem risco?', '- Quais são os procedimentos para proteger os funcionários de
C:\Users\Maralo\Documents\docling-rag-system-main\docling-rag-system-main\RAG
be removed in 1.0. Use :meth:`~invoke` instead.
docs = vectorstore.as_retriever(search_kwargs={"k": K_RETRIEVE}).get_relevant
Total de documentos únicos retornados: 12

--- Enviando para o LLM para gerar a resposta ---

=====
RESPOSTA
=====

O limite de exposição anual para não servidores é de **1 mSv**.
```

Figura 1. 1

## CONCLUSÕES

O projeto resultou no desenvolvimento de um protótipo de Sistema de Recuperação de Informação baseado em LLMs, aplicado ao contexto do Reator Nuclear IEA-R1. O sistema mostrou-se capaz de processar documentos técnicos, indexá-los em formato vetorial e responder a consultas em linguagem natural, fornecendo respostas contextualizadas e rastreáveis.

Embora em fase inicial, os resultados parciais indicam que o protótipo possui potencial significativo para auxiliar operadores e técnicos, aumentando a confiabilidade nos protocolos e o acesso rápido a informações críticas.

## REFERÊNCIA BIBLIOGRÁFICA

[1] Zhai, C. Large Language Models and Future of Information Retrieval: Opportunities and Challenges. SIGIR, 2024.

[2] Souza, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. Perspectivas em Ciência da Informação, v. 11, n. 2, p. 161–173, 2006.

**APOIO FINANCEIRO AO PROJETO**  
CNPq - CNEN