VIII ENCONTRO INTERNACIONAL DE GESTÃO, DESENVOLVIMENTO E INOVAÇÃO
NOVOS CONTEXTOS PARA O AGRONEGÓCIO
ISSN: 2594-8083
VIII EIGEDIN
25 a 27 de Maio de 2026
Totalmente Online e Gratuito

(X) Graduação ( ) Pós-Graduação
( ) Artigo completo ( ) Relato de prática (x) Resumo expandido

DESENVOLVIMENTO DE UMA APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL PARA INTERPRETAÇÃO DE DOCUMENTAÇÃO TÉCNICA E GERAÇÃO AUTOMATIZADA DE RESPOSTAS

Eduardo Koketsu Sescato
Unisenai
edusescato@gmail.com

Guilhermo Silveira
Unisenai
guilhermosilveira@gmail.com

Henrique Alves Jonas
Unisenai

Maria Eduarda Becher Santos
Unisenai
mbechersantos@gmail.com

Matheus Palú Brisola
Unisenai
palu.matheus@gmail.com

Gustavo Dambiski Gomes de Carvalho
Unisenai
E-mail: gustavo.dambiski@gmail.com

RESUMO

Este trabalho apresenta o desenvolvimento de uma aplicação baseada em inteligência artificial voltada à interpretação de documentação técnica e geração automatizada de respostas. Em ambientes industriais, a consulta manual a documentos técnicos é frequentemente um processo lento, dependente de conhecimento especializado e sujeito a erros. A proposta utiliza modelos de linguagem combinados com a técnica Retrieval-Augmented Generation (RAG), permitindo integrar conhecimento pré-treinado com bases externas. A metodologia envolve levantamento bibliográfico, definição da arquitetura, implementação de protótipos e testes iniciais. Os resultados indicam melhoria significativa na precisão e relevância das respostas. Conclui-se que a solução possui potencial para otimizar processos de consulta técnica, aumentando eficiência e confiabilidade.

Palavras-chave: Inteligência Artificial; RAG; Documentação Técnica; Automação; Modelos de Linguagem.

ISSN: 2594-8083

1 INTRODUÇÃO

A crescente digitalização de processos industriais tem gerado um aumento significativo no volume de documentação técnica disponível. Manuais, normas, relatórios e especificações são essenciais para a operação e manutenção de sistemas, porém sua consulta ainda é majoritariamente realizada de forma manual, tornando o processo lento e ineficiente.

Nesse contexto, o uso de inteligência artificial surge como uma alternativa promissora para automatizar a recuperação de informações. Modelos de linguagem (LLMs) têm demonstrado capacidade de interpretar linguagem natural e gerar respostas contextualizadas, porém apresentam limitações relacionadas à confiabilidade das informações, como alucinações (KARATOPRAK, 2025). Para mitigar esse problema, destaca-se a técnica Retrieval-Augmented Generation (RAG), que permite ao modelo acessar bases externas de conhecimento, aumentando a precisão das respostas (GHOSH; MITTAL, 2025). Dessa forma, este trabalho propõe o desenvolvimento de uma aplicação baseada em RAG para interpretação de documentação técnica.

2 DISCUSSÃO E ANÁLISE DOS DADOS

A técnica RAG combina recuperação de informação com geração de texto, permitindo que modelos de linguagem utilizem dados externos para fundamentar suas respostas. Estudos recentes demonstram que essa abordagem reduz significativamente erros e melhora a confiabilidade em aplicações críticas (KARATOPRAK, 2025).

Além disso, a literatura evidencia a importância do uso de bancos de dados vetoriais (NGUYEN, 2026) e técnicas de fragmentação de documentos (chunking), que permitem organizar informações de forma eficiente (KUUSISTO, 2025), além do Chain-of-Thought que de acordo com WANG (2025), também aumenta a precisão de dados interpretados. A indexação semântica possibilita que consultas sejam realizadas com base no significado, e não apenas em palavras-chave (KARATOPRAK, 2025).

Outro aspecto relevante é o uso de workflows automatizados e integração com serviços em nuvem, que tornam as soluções mais escaláveis e flexíveis (LIN; KAO, 2026), visto que Scripcenco (2026) aponta que a integração com o ambiente de trabalho é imprescindível para o bom funcionamento do modelo. Ferramentas como n8n permitem orquestrar diferentes etapas do processo, desde a entrada da pergunta até a geração da resposta (LIN; KAO, 2026).

ISSN: 2594-8083

O desenvolvimento do projeto segue uma abordagem aplicada e experimental. Inicialmente, foi realizado um levantamento bibliográfico para identificar as principais técnicas e ferramentas utilizadas em sistemas baseados em RAG. Em seguida, foi desenvolvido um protótipo da aplicação. A arquitetura da solução é composta por:

• Modelos de linguagem (LLMs) acessados via API - ferramenta Groq;

• Banco de dados vetorial para armazenamento de embeddings - ferramenta Cohere/Supabase;

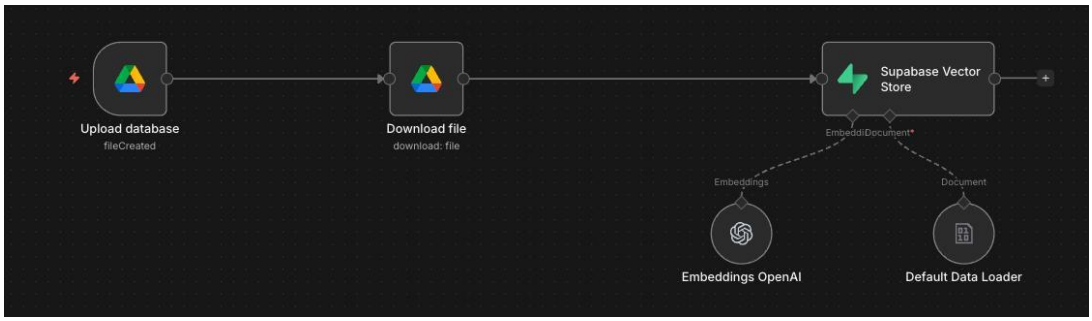
• Ferramenta n8n para orquestração de fluxos e construção de Agentes de IA;

• APIs para comunicação entre serviços via nodes específicos do n8n;

• Ambiente auto-hospedado do n8n para execução e ambiente nuvem para o banco de dados vetorial.

Os documentos técnicos passam por etapas de pré-processamento, incluindo limpeza, segmentação e vetorização. Em seguida, são indexados em um banco vetorial, permitindo consultas semânticas. A figura 1 ilustra no orquestrador n8n o processo de ingestão de informações de um documento pdf na base de dados vetorial utilizando o modelo de embedding do chatgpt.

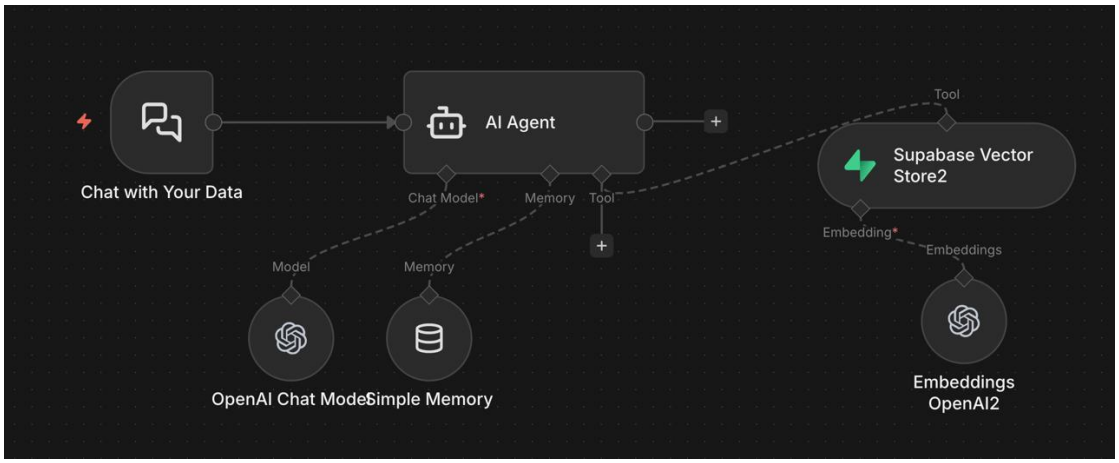
Figura 1 - Ingestão de Informações no Banco Vetorial



Fonte: os autores.

ISSN: 2594-8083

Figura 2 - Recuperação de Informações (rag) com agente de IA



Fonte: os autores.

A Figura 2 ilustra no orquestrador n8n a etapa de recuperação de informações com o agente de IA utilizando como ferramenta (tool) o banco de dados vetorial, isto é, as informações previamente ingeridas sobre o documento pdf. Nesse quesito, foram realizados testes iniciais com diferentes configurações, variando parâmetros como tamanho de chunk (trecho de texto), número de documentos recuperados e estratégias de engenharia de prompt. A avaliação considerou:

- Precisão das respostas;
- Relevância semântica;
- Tempo de resposta;
- Consistência dos resultados.

3 CONCLUSÕES

Os resultados obtidos indicam que a utilização de RAG melhora significativamente a qualidade das respostas em comparação com modelos de linguagem isolados.

Foi desenvolvido um protótipo funcional utilizando n8n, capaz de realizar consultas automatizadas em documentos técnicos. A utilização de bancos vetoriais e fragmentação de documentos demonstrou impacto positivo na recuperação de informações.

A análise dos artigos selecionados também indica que a combinação de RAG com técnicas como Chain-of-Thought pode melhorar ainda mais o desempenho em tarefas complexas.

Observou-se ainda que a escolha adequada de parâmetros, como tamanho dos

ISSN: 2594-8083

fragmentos e número de documentos recuperados, influencia diretamente na qualidade das respostas.

A partir dos resultados obtidos, conclui-se que a aplicação de modelos de linguagem combinados com a técnica RAG tem potencial para sistemas de consulta em documentação técnica. A solução proposta demonstrou potencial para reduzir o tempo de busca por informações, aumentar a precisão das respostas e melhorar a eficiência operacional.

Como trabalhos futuros, propõe-se a ampliação da base de dados, integração com sistemas reais e aprimoramento dos métodos de avaliação com testes experimentais detalhados.

AGRADECIMENTOS

O presente trabalho foi desenvolvido no âmbito do programa da Iniciação Científica do UniSENAI/PR.

REFERÊNCIAS

GHOSH, Soham; MITTAL, Gaurav. Advancing engineering research through context-aware and knowledge graph–based retrieval-augmented generation. *Frontiers in Artificial Intelligence*, v. 8, p. 1697169, 2025.

KARATOPRAK, Burhan. Performance of Retrieval-Augmented Generation (RAG) Systems on Turkish Academic Texts.

LIN, Tzu-Hsuan; KAO, Chih-Hsuan. FROAV: A Framework for RAG Observation and Agent Verification-Lowering the Barrier to LLM Agent Research. *arXiv preprint arXiv:2601.07504*, 2026.

NGUYEN, Duc Phu. Enhancing RAG agent performance through structured knowledge bases from NLP-parsed data warehouses. 2026.

WANG, Zhanliang et al. Integrating Chain-of-Thought and Retrieval Augmented Generation Enhances Rare Disease Diagnosis From Clinical Notes. *Medicine Bulletin*, 2025.