

AVALIAÇÃO – Rodada 1

Parecerista

Ciência aberta*

Em conformidade com as práticas de comunicação da Ciência Aberta a pessoa avaliadora concorda com a publicação do parecer, no qual aprova ou desaprova a publicação do manuscrito, juntamente de suas avaliações.

- ☐ Sim: concordo em abrir o parecer com a minha identificação
☒ Sim: concordo em abrir o parecer sem a minha identificação
☐ Não concordo em abrir o parecer

Trabalho avaliado:

ABRANTES, Paula Cotrim de; GUEDES, Vania Lisboa da Silveira. Construção e análise de protótipo de chatbot generativo com Retrieval-Augmented Generation (RAG) a partir de um vocabulário controlado para aplicação no domínio da Propriedade Intelectual (PI). **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 31, [Fluxo contínuo], p. 1-28, 2026. DOI <http://dx.doi.org/10.1590/1981-5344/61756>

Completo em: 2026-02-25 12:14 PM
Recomendação: Correções obrigatórias

Título do manuscrito:

Construção e análise de protótipo de chatbot generativo com Retrieval-Augmented Generation (RAG) a partir de um vocabulário controlado para aplicação no domínio da Propriedade Intelectual (PI)

O título é adequado, ou seja, representa o menor resumo do seu conteúdo?*¹

- ☒ Sim ☐ Não

Novo título sugerido:

--

¹ * Indica parecer obrigatório.

O resumo é adequado, contendo objetivo, método, resultados e conclusão do trabalho?*

☒ Sim ☐ Não

Melhorias a serem realizadas no resumo:

Como ajustes pontuais, recomenda-se maior padronização terminológica (por exemplo, introduzir a sigla "Propriedade Intelectual (PI)" na primeira ocorrência no resumo).

O trabalho é original?*

☒ Sim ☐ Não

O tema é atual?*

☒ Sim ☐ Não

Considerações a respeito da atualidade do tema:

O tema é atual e relevante. O trabalho aborda a integração entre IA generativa, Retrieval-Augmented Generation (RAG) e vocabulários controlados, tópicos que estão no centro das discussões contemporâneas tanto na Ciência da Informação quanto na Ciência da Computação. Além disso, a articulação entre Organização do Conhecimento e sistemas inteligentes representa uma linha de pesquisa emergente e alinhada às transformações tecnológicas atuais.

Contribuição do artigo a aplicação de conhecimentos para a área de conhecimento:*

☐ Totalmente Satisfatório
☒ Satisfatório
☐ Insatisfatório
☐ Totalmente Insatisfatório

Relevância e consistência teórica do artigo para o desenvolvimento da área de conhecimento:*

☐ Totalmente Satisfatório
☒ Satisfatório
☐ Insatisfatório
☐ Totalmente Insatisfatório

Considerações a respeito da relevância e consistência teórica:

O trabalho apresenta relevância ao propor a integração entre vocabulários controlados e sistemas de IA generativa, evidenciando potencial contribuição para a melhoria da recuperação da informação em domínios especializados. A consistência teórica é satisfatória, especialmente na articulação entre Organização do Conhecimento, Recuperação da

Informação e Inteligência Artificial. O texto demonstra coerência ao sustentar que vocabulários controlados podem reduzir ambiguidade semântica e apoiar sistemas automatizados de resposta. Entretanto, em alguns momentos a discussão técnica sobre IA e RAG permanece descritiva, sem aprofundamento analítico. Por exemplo, a relação entre recuperação semântica, embeddings e geração de resposta poderia ser explorada com maior rigor conceitual. Para versões futuras do texto esse diálogo deve ser melhorado.

Qualidade de redação e organização do texto (clareza, concisão, objetividade, estrutura formal):*

- ☐ Totalmente Satisfatório
☒ Satisfatório
☐ Insatisfatório
☐ Totalmente Insatisfatório

Considerações a respeito de redação e organização do texto:

Há alguns pontos de melhoria:
Pequenas inconsistências terminológicas (ex.: "linguagem natural (LN)" > "Linguagem Natural (LN)"; "processamento de linguagem natural (PLN)" > "Processamento de Linguagem Natural (PLN)").
Possível erro tipográfico ("GiHib", provavelmente "GitHub").
Recomenda-se inserir o link do repositório quando mencionado, para melhorar transparência e reprodutibilidade.
Em alguns trechos técnicos, o texto assume tom descritivo, podendo ser mais analítico (se posicionando melhor).

São ajustes pontuais que não comprometem a compreensão do trabalho.

Qualidade do referencial teórico: bem desenvolvido, articulado e de relevância*

- ☐ Totalmente Satisfatório
☒ Satisfatório
☐ Insatisfatório
☐ Totalmente Insatisfatório

Considerações a respeito do referencial teórico:

A seção sobre Inteligência Artificial contextualiza corretamente conceitos como embeddings, transformers, RAG e chunking, ainda que, em alguns momentos, de forma mais introdutória do que analítica. O referencial cumpre seu papel de sustentar a proposta do estudo, embora pudesse aprofundar a discussão sobre limitações técnicas de RAG e mecanismos de recuperação semântica.

De modo geral, o referencial é consistente e pertinente.

Metodologia utilizada: bem apresentada, de qualidade e de alto nível de sofisticação.*

- ☐ Totalmente Satisfatório
- ☐ Satisfatório
- ☒ Insatisfatório
- ☐ Totalmente Insatisfatório

Considerações a respeito da metodologia:

A metodologia está descrita de forma clara e permite compreender o processo de construção do vocabulário controlado, sua integração ao chatbot e a execução do experimento. Entretanto, observam-se fragilidades no trabalho no sentido de sua validação metodológica, o que limita a robustez científica dos resultados apresentados.

Primeiramente, a avaliação experimental baseia-se em apenas cinco perguntas, o que constitui um universo reduzido para sustentar conclusões sobre desempenho semântico, precisão ou confiabilidade do sistema. A ausência de uma amostra mais ampla impede generalizações e reduz a força empírica do estudo.

Além disso, não há comparação com baseline. O trabalho não avalia, por exemplo, o desempenho do sistema sem RAG, sem vocabulário controlado, ou com diferentes configurações de recuperação. Sem essa comparação, torna-se difícil isolar o ganho efetivo proporcionado pela abordagem proposta.

Outro ponto relevante é que o experimento foi conduzido utilizando apenas uma LLM (gpt-4o-mini). Dessa forma, não é possível avaliar o impacto da escolha do modelo de linguagem no desempenho do sistema, nem verificar a robustez da arquitetura proposta frente a diferentes engines. Essa limitação reduz a generalização dos resultados.

Também se observa a ausência de métricas quantitativas formais. O estudo não apresenta medidas como precisão, acurácia, taxa de recuperação correta, taxa de erro semântico ou avaliação estruturada por especialistas. Consequentemente, a conclusão de "sucesso parcial" baseia-se em evidências qualitativas limitadas. Aqui os autores precisam se posicionar melhor sobre as métricas aplicadas (ou o não uso delas).

Adicionalmente, o trabalho não detalha completamente (ou pelo menos não ficou claro) alguns parâmetros relevantes do sistema, como modelo de embeddings utilizado, tamanho e sobreposição de chunks, parâmetros do FAISS, configuração de recuperação (top-k) e arquitetura detalhada do pipeline RAG. A ausência dessas informações limita a reprodutibilidade técnica do experimento.

Do ponto de vista conceitual, há também uma simplificação na interpretação dos erros observados. O texto associa imprecisões principalmente ao chunking e à temperatura da LLM, porém não diferencia claramente falhas de recuperação (retrieval), ranqueamento, representação semântica ou geração da resposta. Essa distinção seria importante para compreender a real origem das limitações observadas.

Por fim, o escopo experimental é relativamente pequeno (vocabulário com 106 termos), o que também restringe a generalização dos achados.

Essas limitações não invalidam a proposta do trabalho, mas indicam que seus resultados devem ser interpretados como exploratórios. Recomenda-se que ampliem a base experimental, utilizem métricas formais, comparem diferentes configurações e aprofundem a análise técnica dos erros observados. Ou pelo menos posicionem melhor o trabalho nesses aspectos.

Análise e discussão dos resultados: consistência, articulação teórica e metodológica e interpretação sem especulações ou afirmações não sustentadas pelos dados*

- ☐ Totalmente Satisfatório
☐ Satisfatório
☒ Insatisfatório
☐ Totalmente Insatisfatório

Considerações a respeito da análise e discussão dos resultados:

A análise dos resultados é clara e apresenta avaliação honesta do desempenho do sistema, reconhecendo sucessos e limitações. Entretanto, a discussão poderia ser aprofundada em termos técnicos (sobretudo aspectos metodológicos citados anteriormente), especialmente:

Diferenciar erros de recuperação (retrieval) e erros de geração da LLM.
Analisar mais detalhadamente o impacto do chunking e da escolha do top-k.
Discutir limitações do uso de apenas uma LLM.
Ampliar a análise além de exemplos pontuais.
Trazer análises estatísticas de significância dos resultados.

Conclusões: fundamento e coerência, relacionadas com o(s) objetivo(s) do artigo e mostrando o avanço do conhecimento científico:*

- ☐ Totalmente Satisfatório
☒ Satisfatório
☐ Insatisfatório
☐ Totalmente Insatisfatório

Considerações a respeito das conclusões:

As conclusões estão alinhadas ao objetivo do estudo e, em geral, mantêm coerência com os resultados apresentados, especialmente ao reconhecer que o desempenho do protótipo foi parcialmente satisfatório e que ainda há necessidade de ajustes finos e novas investigações. Entretanto, assim como discutido na metodologia e nos resultados, as conclusões acabam ficando mais fortes do que a evidência empírica permite sustentar, principalmente devido às limitações do desenho experimental. Nesse sentido, recomenda-se que as conclusões enfatizem com maior clareza que os achados são exploratórios e não generalizáveis, evitando interpretações amplas sobre eficácia do RAG ou do vocabulário controlado a partir de uma amostra tão reduzida.