PARECER*

Artigo Avaliado CASTILLO BADILLA, Carlos Andrés. Modelo replicável de engenharia de prompts para recuperação de literatura científica com ChatGPT. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis/SC, Brasil, v. 31, p. 1–30, 2026.

Rodada de Avaliação 01

- [ ] Rejeitar
- [O] Correções obrigatórias (requer grandes ajustes e nova rodada de análise pelo avaliador)
- [ ] Aceitar com pequenos ajustes (não necessita nova análise)
- [ ] Aceitar sem alterações

Originalidade e Plágio: espera-se que o trabalho seja original e não contenha plágio ou outras formas de fraude e má conduta, caso contrário se sugere justificar e rejeitar de imediato. Se o artigo provém de uma publicação em evento, esta versão deve conter melhorias significativas em relação ao original *

Fraco

Contribuição/Relevância para a área *

Bom

Título e Objetivo: o título deve representar adequadamente o artigo e o objetivo devem estar explicitado com clareza no texto *

Bom

Referencial teórico: deve ser suficientemente aprofundado e conter citações a outros estudos de prestígio relacionados e publicados em revistas nacionais (inclusive nesta) e/ou internacionais *

Bom

Metodologia: os métodos utilizados devem ser claros e adequados aos fins perseguidos *

Fraco

Resultados e Conclusões: devem estar em consonância com as evidências do estudo e os objetivos propostos, demonstrando o atingimento dos mesmos *

Muito Fraco

Redação e normas ABNT: o texto está escrito de forma clara, coerente, sem erros e cumpre com as normas ABNT *

Bom

Avaliação Geral: indique seu parecer e as recomendações exigidas em caso de aprovação, em caso de rejeição indique os motivos de forma clara (este parecer será visível para os autores) *

In a moment in which society in general, and academia in particular, is dealing with the overwhelming presence of Generative AI, this paper poses an interesting and relevant research question. Generative AI, generally in the form of text-producing software systems, is known to be able to generate convincing texts of all kinds. However, given the very nature of Generative AI, especially that of broad- or general-scope GenAI (i.e., ChatGPT and the like), the veracity of the claims included in the generated text content cannot be assumed (and thus are described, rather than errors like in "regular software," as "hallucinations"). Admittedly, companies developing these black-box systems are constantly fine-tuning and rail-guarding these products, so the attention of the research community is justified. In this regard, the research question that the author(s) aim for, whether prompt engineering (i.e. careful crafting of input text) contributes to reduce/eliminate hallucinations when text-based Generative AI is used for literature retrieval, is relevant.

However, the paper in its current form presents quite a number of flaws that make it unfit for publication, in the opinion of this reviewer. Here's the list of my most pressing concerns about it:

* The research claims to evaluate ChatGPT's effectiveness in scientific literature retrieval, but it fails to start by giving a proper definition of what "effectiveness" is, in the context of the work. Given that the results of the experiments (i.e., the retrieved results) are not compared with an "expected output" (i.e., set of results), "effectiveness" is not being used with its most common meaning in computer science. Thus, the author(s) must clearly state what definition they are using, or else refer to other, more specific metrics (i.e., lack of invented references).
* The research question is rather open in order to be answered with the designed experiment. The author(s) will not uncover "how" their proposed prompt engineering protocol contributes, but rather whether it actually does or not, in terms of the KPIs (key performance indicators) that they choose to examine (which, referring to the previous point of this list) are loosely, if not at all, described.
* The paper references terms that may not be assumed to be known to the general reader. For instance, "ChatGPT Plus (Investiga a fondo)" seems to be some sort of specific agreement product-as-a-service for the use of ChatGPT. Instead of using this term, the author(s) should simply state which version of ChatGPT they are using (i.e., ChatGPT-5.2, ChatGPT-4.5, ChatGPT-4o, ChatGPT-4o-mini...). Only this way may the experiments (i.e., searches) be reproduced by others.
* Another piece of information that is neither cited, nor explained, is the Zero-Hallucination protocol (ZHP-7).
* Even if the author(s) claim to adhere to FAIR principles, and they do include their engineered prompt in their appendix, there's other data that would be adamant for result verification and validation that, even if they say they'll make it available, is not cited or

referred to. This includes the dataset (i.e., the result of the 30 searches and the specific dates they were run).
* There are some measures that are not defined nor reported in a precise manner in the paper. For instance, why and how is the "reply time" (i.e., assuming this is the time that ChatGPT needs to process the input and produce the output) considered relevant? And, if it is, why is it later on only reported as wide, overlapping time ranges (10-15, 15-20, 15-25)? Later on, the concept of "expected temporal range" (!!).
* Similarly, sample saturation is used as a methodological criterion to claim that 30 searches were enough but is never defined nor justified.
* When reporting their results, the author(s) seem to differentiate hallucinations from invented/unverifiable references from conceptual consistency (i.e., topic relevance?), but without ever clearly describing what these are, the reader is left puzzled and wondering.
* In presenting their qualitative analysis, the author(s) state they select a representative sample of 8 results. Without ever knowing the number of absolute results (with/without duplicates) returned by the 30 searches, it is impossible to judge whether 8 might be a reasonable representative sample. The author(s) do not explain or justify this representativeness, nor do they include the details of those 8 papers.

Regarding the list of references in the paper, there are a couple that are broken at the day of writing this review (ANID, 2024; BRASIL, 2016) which are not cited in the text anyway. Up to an additional six references are not cited either (HAMAN and SKOLNIK, 2024; OCDE, 2019; PANDA, BHATT and SATAPATHY, 2024; SUPPADUNGSUK et al 2023; UNESCO, 2021; WILKINSON, 2016), and one that is repeated (CHILE 2021 and CHILE 2024).

HISTÓRICO

Designado: 2/12/2025 - **Confirmado**: 10/12/2025 - **Concluído**: 12/01/2026

Encontros Bibli

PARECER*

Rodada de Avaliação	02

- [ ] Rejeitar
- [ ] Correções obrigatórias (requer grandes ajustes e nova rodada de análise pelo avaliador)
- [O] Aceitar com pequenos ajustes (não necessita nova análise)
- [ ] Aceitar sem alterações

Originalidade e Plágio: espera-se que o trabalho seja original e não contenha plágio ou outras formas de fraude e má conduta, caso contrário se sugere justificar e rejeitar de imediato. Se o artigo provém de uma publicação em evento, esta versão deve conter melhorias significativas em relação ao original *

Regular

Contribuição/Relevância para a área *

Regular

Título e Objetivo: o título deve representar adequadamente o artigo e o objetivo devem estar explicitado com clareza no texto *

Bom

Referencial teórico: deve ser suficientemente aprofundado e conter citações a outros estudos de prestígio relacionados e publicados em revistas nacionais (inclusive nesta) e/ou internacionais *

Bom

Metodologia: os métodos utilizados devem ser claros e adequados aos fins perseguidos *

Regular

Resultados e Conclusões: devem estar em consonância com as evidências do estudo e os objetivos propostos, demonstrando o atingimento dos mesmos *

Bom

Redação e normas ABNT: o texto está escrito de forma clara, coerente, sem erros e cumpre com as normas ABNT *

Bom

Avaliação Geral: indique seu parecer e as recomendações exigidas em caso de aprovação, em caso de rejeição indique os motivos de forma clara (este parecer será visível para os autores) *

Los autores parecen haber atendido en modo suficientemente satisfactorio todas las recomendaciones realizadas en la primera revisión.

Como nota de edición, debe eliminarse la etiqueta "TÍTULO SIGNIFICATIVO: " que precede al título en el cuerpo del manuscrito.

HISTÓRICO

Designado: 18/02/2026 - **Confirmado**: 19/02/2026 - **Concluído**: 11/03/2026