Revista GCTI
Gestão do Conhecimento e Tecnologia da Informação
RGCTI 10:68-87 2026
ISSN: 2526-8384

A Inteligência Artificial Generativa (IAG) como ferramenta de apoio à pesquisa em Ensino de Ciências: exemplos de uso e estratégias de elaboração de prompts

Generative Artificial Intelligence (GenAI) as a tool to support research in Science Education: examples of use and prompt-design strategies

Gabriel Luiz Nalon Macedo [a], **Luciano Carvalhais Gomes** [b], **Marcos Cesar Danhoni Neves** [c]

[a] *gabrielnalonmacedo@gmail.com*

[b] *lcgomes2@uem.br*

[b] *mcdneves@uem.br*



Data de submissão: 14/10/2026
Data de aceite: 25/11/2026

Resumo

Objetivo: Analisar as possibilidades e os limites da utilização da inteligência artificial generativa (IAG) como ferramenta de apoio em diferentes tarefas da pesquisa em Ensino de Ciências e examinar, nos casos estudados, como o refinamento dos prompts se relaciona com a qualidade das respostas produzidas.

Metodologia: Estudo qualitativo, descritivo e comparativo de cinco casos intencionalmente selecionados: revisão bibliográfica, tradução, localização documental, interpretação de fonte histórica e revisão da escrita acadêmica. Compararam-se as respostas da primeira e da última versão de cada prompt, produzidas no ChatGPT (modelo GPT-5, plano gratuito) entre setembro e novembro de 2025, segundo cinco dimensões: adequação à tarefa, precisão/fidelidade, completude, rastreabilidade/verificabilidade e necessidade de intervenção humana.

Resultados: As versões finais apresentaram diferenças na adequação às instruções, na cobertura dos elementos solicitados e na apresentação de informações passíveis de conferência, sem que isso demonstrasse maior precisão ou validade científica. A comparação evidenciou que conferir referências e passagens documentais, interpretar fontes históricas e revisar a escrita exigem operações distintas de validação.

Implicações práticas/gerenciais: Os resultados oferecem subsídios para pesquisadores, grupos e programas de pós-graduação discutirem a formulação de instruções, a documentação das interações e os procedimentos de verificação adequados a cada tarefa.

Implicações sociais: O estudo contribui para a discussão sobre formação crítica e uso transparente da IAG na pesquisa, preservando a responsabilidade intelectual dos pesquisadores pelas decisões e conclusões.

Palavras-chave: Inteligência Artificial Generativa; Interação pesquisador-IAG; Prompts; ChatGPT; Mediação na pesquisa.

Abstract

Purpose: To analyze the possibilities and limitations of using generative artificial intelligence (GenAI) as a support tool for different research tasks in Science Education and to examine, in the cases studied, how prompt refinement relates to the quality of the responses produced.

Methodology: A qualitative, descriptive, and comparative study of five purposively selected cases: literature review, translation, document passage retrieval, interpretation of a historical source, and academic writing revision. Responses generated by the first and final versions of each prompt using ChatGPT (GPT-5 model, free plan) between September and November 2025 were compared across five dimensions: task alignment, accuracy/fidelity, completeness, traceability/verifiability, and need for human intervention.

Results: Final versions showed differences in adherence to instructions, coverage of requested elements, and provision of information that could be checked, without demonstrating greater accuracy or scientific validity. The comparison showed that checking references and documentary passages, interpreting historical sources, and revising academic writing require different validation procedures.

Practical/Managerial Implications: The findings provide a basis for researchers, research groups, and graduate programs to discuss prompt formulation, documentation of interactions, and task-appropriate verification procedures.

Social Implications: The study contributes to discussions of critical training and transparent use of GenAI in research while preserving researchers' intellectual responsibility for decisions and conclusions.

Keywords: Generative Artificial Intelligence; Researcher–GenAI interaction; Prompts; ChatGPT; Research mediation.

1. Introdução

A incorporação recente de sistemas de inteligência artificial generativa (IAG) baseados em grandes modelos de linguagem (LLMs, do inglês large language models) ampliou as possibilidades de sua utilização em diferentes operações da atividade acadêmica. Estudos têm examinado seu emprego em tarefas como revisão e síntese de literatura, produção e sumarização de textos, tradução, organização de informações e análise de dados qualitativos (Christou, 2023; Rodrigues et al., 2024; Sampaio et al., 2024), bem como na produção de imagens e animações (Neves, 2026). Essas possibilidades envolvem riscos de imprecisões, informações fabricadas, vieses e problemas de transparência, privacidade e integridade científica. Assim, sua utilização requer examinar não apenas a plausibilidade das respostas, mas as condições em que são produzidas, verificadas e incorporadas à pesquisa.

No campo do Ensino de Ciências, investigações recentes já permitem observar algumas dessas possibilidades e tensões em situações concretas. Anjos et al. (2024), ao comparar o desempenho do ChatGPT 4.0 e do Claude 2.0 na análise de entrevistas utilizadas em pesquisas sobre aprendizagem de conceitos de Física, identificaram diferenças importantes entre os modelos e destacaram a necessidade de supervisão humana para reconhecer suas limitações. Em outra vertente, Silva e Paula (2024) discutiram a utilização do ChatGPT como ferramenta de apoio à Análise Textual Discursiva em pesquisas da área de Educação em Ciências e Matemática, indicando possibilidades de utilização em etapas iniciais de tratamento textual sem transferir à ferramenta a responsabilidade pela interpretação e categorização. Esses estudos mostram que a discussão sobre IAG no Ensino de Ciências não se reduz à adoção de uma nova tecnologia, mas envolve decisões metodológicas sobre quais operações podem ser apoiadas pela ferramenta e quais permanecem dependentes do julgamento teórico e interpretativo do pesquisador.

Uma dessas decisões refere-se à formulação das instruções fornecidas à IAG. Em revisão de escopo assistida por inteligência artificial (IA), Sampaio et al. (2024) relatam a necessidade de modular perguntas quando as respostas se mostravam evasivas ou pouco relacionadas aos critérios de análise. No contexto da Análise Textual Discursiva, Silva e Paula (2024) discutem o emprego de prompts para orientar procedimentos de unitarização, enquanto Naeem et al. (2025) enfatizam a explicitação de objetivos, contexto e referenciais metodológicos. Esses trabalhos indicam que o prompt participa da delimitação da tarefa e das condições de avaliação de suas respostas. Neste estudo, ele é compreendido como elemento da interação entre pesquisador e IAG, cuja reformulação pode estar associada a diferenças observáveis nas devolutivas, sem assegurar sua validade científica.

A maior especificidade de um prompt, contudo, não constitui garantia de exatidão ou validade científica. Christou (2023) enfatiza, nesse sentido, a necessidade de familiarização do pesquisador com os dados, verificação cruzada das informações, controle do processo de análise e demonstração de aporte cognitivo humano. Sampaio et al. (2024), igualmente, ressaltam que o apoio da IA não elimina a interpretação qualitativa realizada por pesquisadores. A relação entre formulação dos prompts e qualidade das respostas precisa, portanto, ser examinada sem assumir que maior detalhamento do comando produz automaticamente resultados cientificamente mais válidos.

Estudos anteriores abordam tarefas específicas como análise de entrevistas (Anjos et al., 2024), Análise Textual Discursiva (Silva & Paula, 2024), revisão bibliográfica (Rodrigues et al., 2024) e aplicações da IAG na pesquisa

qualitativa (Sampaio et al., 2024). No entanto, permanece pertinente examinar como a natureza de diferentes tarefas se relaciona às dimensões de qualidade e às formas de validação. É nesse recorte que se situa o presente artigo: comparar respostas iniciais e finais em cinco tarefas sob dimensões analíticas comuns, sem reivindicar ausência de estudos comparativos.

Os cinco casos vinculam-se a atividades de pesquisa em Ensino de Ciências, embora não envolvam diretamente práticas de ensino ou aprendizagem. Quatro deles integram investigações histórico-epistemológicas da Física: o levantamento bibliográfico sobre a constante gravitacional contribui para examinar a construção histórica desse conhecimento, enquanto a tradução, a localização e a interpretação de passagens de Kepler exigem compreender formulações científicas em seu contexto de produção. A pertinência dessas investigações ao Ensino de Ciências reside na possibilidade de subsidiar discussões sobre a natureza da ciência, a historicidade dos conceitos e os processos de construção do conhecimento científico, aspectos destacados por Matthews (1995). Para isso, o tratamento crítico das fontes e a atenção aos contextos históricos são fundamentais, evitando interpretações anacrônicas (Martins, 2022). O quinto caso, dedicado à revisão da escrita acadêmica, contempla uma atividade transversal à pesquisa na área, na qual a precisão conceitual e a preservação dos sentidos e das decisões autorais são essenciais à comunicação científica.

A contribuição científica buscada consiste, portanto, em discutir comparativamente as relações entre a natureza das tarefas, as características observadas nas respostas e as condições necessárias à sua validação acadêmica. Essa análise distingue-se da utilidade prática dos prompts apresentados, que podem servir como exemplos de formulação de instruções, sem constituir procedimentos universalmente eficazes. A comparação é delimitada aos registros examinados e não autoriza atribuir mudanças a componentes isolados dos prompts ou generalizar o desempenho do sistema.

Diante disso, o problema que orienta o trabalho é: como a inteligência artificial generativa pode ser utilizada como ferramenta de apoio em diferentes tarefas da pesquisa em Ensino de Ciências e como o refinamento dos prompts se relaciona com a qualidade das respostas produzidas? Em consonância com essa questão, o objetivo geral consiste em analisar as possibilidades e os limites da utilização da inteligência artificial generativa como ferramenta de apoio em diferentes tarefas da pesquisa em Ensino de Ciências e examinar, nos casos estudados, como o refinamento dos prompts se relaciona com a qualidade das respostas produzidas.

2 Fundamentação teórica

A discussão da IAG na pesquisa requer distinguir conceitos que, embora relacionados, não são equivalentes. A IA reúne abordagens computacionais para percepção, representação do conhecimento, raciocínio, aprendizagem e tomada de decisões (Russell & Norvig, 2022). Nesse campo, a IAG compreende sistemas que produzem conteúdos a partir de regularidades identificadas nos dados utilizados em seu treinamento (Santaella & Kaufman, 2024). No caso da geração textual, determinadas aplicações utilizam LLMs. No desenvolvimento dos modelos de linguagem contemporâneos, Vaswani et al. (2017) propuseram a arquitetura Transformer, enquanto Brown et al. (2020) demonstraram, com o GPT-3, a realização de diferentes tarefas mediante instruções e exemplos fornecidos textualmente, sem ajustes adicionais dos parâmetros do modelo. Nesse contexto, o ChatGPT constitui uma aplicação conversacional baseada em grandes modelos de linguagem. Dessa forma, IA, IAG, LLMs e ChatGPT não designam conceitos equivalentes (Pimentel & Carvalho, 2025).

A interação com sistemas como o ChatGPT ocorre, entre outras formas, por meio de prompts, entendidos como entradas textuais nas quais o usuário formula perguntas, fornece instruções ou apresenta informações contextuais (Naeem et al., 2025). No contexto da pesquisa, essas entradas delimitam a tarefa e explicitam critérios para a resposta, mas sua pertinência metodológica depende também da avaliação realizada pelo pesquisador. Na perspectiva da interação humano-IA, Amershi et al. (2019) destacam a necessidade de tornar compreensíveis as capacidades e limitações dos sistemas e de possibilitar correções durante o uso. Embora suas diretrizes não tenham sido formuladas especificamente para a pesquisa com IAG, elas permitem compreender o refinamento do prompt como parte de uma interação em que o usuário reorienta a atividade diante das

devolutivas obtidas.

Essa interação pode envolver diferentes níveis de sistematização. Liu et al. (2023) mostram que o termo prompting abrange procedimentos distintos no processamento de linguagem natural; por isso, não se deve equiparar indiscriminadamente esses métodos à formulação de comandos em interfaces conversacionais. Neste estudo, seguimos Pimentel e Carvalho (2025) ao tratar a engenharia de prompts como estruturação deliberada de instruções que podem explicitar tarefa, contexto, restrições, formato de resposta e outros elementos pertinentes.

A importância do conteúdo do prompt não significa, entretanto, que uma instrução mais detalhada assegure validade científica. Pimentel e Carvalho (2025) advertem que a estruturação das solicitações não elimina a possibilidade de respostas incorretas ou enviesadas. Bender et al. (2021), ao discutir os limites dos LLMs, problematizam a associação entre manipulação de formas linguísticas e compreensão de seus significados. Isso ajuda a distinguir a adequação formal da resposta – atender ao formato e aos itens solicitados – de sua confiabilidade factual e conceitual. Sampaio et al. (2024) também indicam que a modulação das perguntas pode favorecer respostas mais úteis, sem dispensar interpretação e conferência humanas. Assim, extensão, fluência e aparente completude não devem ser tomadas como indicadores suficientes de precisão: respostas mais organizadas podem continuar contendo afirmações que necessitam de verificação.

O uso da IAG na pesquisa também envolve vieses, informações fabricadas, privacidade, integridade e responsabilidade (Christou, 2023; Sampaio et al., 2024). Essas questões assumem especial relevância quando os sistemas recebem documentos, entrevistas, dados pessoais ou outros materiais não públicos. Esses limites também alcançam o próprio processo de produção do conhecimento. Messeri e Crockett (2024) argumentam que a incorporação de ferramentas de IA à pesquisa pode favorecer ilusões de compreensão, nas quais o aumento de produtividade ou a aparente objetividade é confundido com ampliação do entendimento científico. Trata-se de um risco epistemológico que não se reduz à ocorrência de erros factuais: inclui a aceitação pouco crítica de explicações e enquadramentos sugeridos pela ferramenta. Nesse contexto, a supervisão humana não constitui apenas revisão posterior. Christou (2023) enfatiza o conhecimento dos dados e a conferência das interpretações; Hamilton et al. (2023) evidenciam a relevância de conhecimento contextual e sensibilidade interpretativa em comparação com análises assistidas por IA. Estudos em Educação em Ciências também sustentam que o apoio computacional não transfere ao sistema as decisões teóricas e metodológicas do pesquisador (Anjos et al., 2024; Silva & Paula, 2024).

Desse modo, neste artigo, a IAG é compreendida como ferramenta de apoio à pesquisa inserida em uma interação orientada e avaliada pelo pesquisador. As exigências de validação variam conforme a tarefa: na localização de informações, importa confrontar a resposta com o documento de origem; na explicação de conceitos, examinar consistência e fontes; na interpretação de materiais, considerar contexto, pressupostos e referencial teórico. Em todas elas, a aparência de adequação da devolutiva precisa ser apreciada à luz de critérios externos ao texto gerado. É na articulação entre tarefa, informações fornecidas no prompt, resposta produzida e julgamento humano que se situa a perspectiva deste estudo. O interesse não é atribuir às reformulações um efeito geral e isolado, mas compreender como diferentes configurações de solicitação se associam às características das devolutivas analisadas.

3 Metodologia

Este estudo qualitativo, descritivo e comparativo examinou cinco tarefas de pesquisa em Ensino de Ciências: revisão bibliográfica, tradução, localização documental, interpretação de fonte histórica e revisão da escrita acadêmica. O corpus reuniu os prompts, as respostas e os registros de reformulação. A comparação ocorreu em dois movimentos: primeiro, confrontaram-se as respostas inicial e final dentro de cada tarefa; depois, examinaram-se transversalmente as semelhanças e diferenças entre os cinco casos segundo dimensões comuns. Essa organização se aproxima da distinção entre análise intracasos e entre casos discutida por Miles et al. (2014), sem caracterizar um teste experimental de eficácia.

A seleção intencional dos casos decorreu das atividades de pesquisa dos autores e de uma disciplina de doutorado sobre IAG na pesquisa em Ensino de Ciências. Essa escolha é compatível com a seleção de situações relevantes ao

problema investigado, característica das abordagens qualitativas discutidas por Flick (2013), sem pretensão de representatividade estatística. Reuniram-se, assim, tarefas concretas nas quais o refinamento das instruções foi mobilizado para atender a diferentes necessidades investigativas.

Foi utilizado o ChatGPT, sistema de IAG desenvolvido pela empresa estadunidense OpenAI, operando com o modelo GPT-5 por meio do plano gratuito. As interações que constituíram o corpus deste estudo foram realizadas entre setembro e novembro de 2025. Os materiais e recursos empregados variaram em função da tarefa analisada. Na revisão bibliográfica, utilizou-se o recurso de pesquisa aprofundada disponível na plataforma; nas tarefas de tradução, localização e interpretação, o material correspondente foi fornecido ao sistema para que as respostas fossem produzidas a partir do documento utilizado no respectivo caso. Cada nova versão de um prompt foi testada em uma conversa independente, sem reutilização do histórico conversacional das versões anteriores, de modo que a resposta analisada fosse produzida a partir do conteúdo daquela versão e dos materiais especificamente fornecidos para a tarefa.

Para cada tarefa, elaborou-se um prompt inicial cuja resposta foi examinada quanto à adequação, omissões, imprecisões e possibilidades de conferência. As observações orientaram reformulações sucessivas, em processo iterativo de avaliação e reajuste das instruções, como discutem Pimentel e Carvalho (2025). Não se fixaram previamente o número de versões nem critérios formais de encerramento. O processo cessou quando julgamos o atendimento satisfatório e os problemas remanescentes dependeriam principalmente de verificação, interpretação ou decisão humana, em vez de novas instruções.

Cada versão foi executada uma única vez para fins da comparação apresentada neste estudo. Embora os prompts finais tenham sido posteriormente reutilizados pelos autores em outras atividades, essas aplicações não integraram o corpus analisado. As versões intermediárias documentaram o refinamento, enquanto a análise contrastou as devolutivas inicial e final. Foram consideradas quatro versões na revisão bibliográfica, três na tradução, cinco na localização, três na interpretação e duas na revisão da escrita, conforme o Quadro 1.

As cinco dimensões foram organizadas para permitir um exame transversal das respostas, considerando as exigências específicas de cada tarefa: adequação à tarefa (correspondência ao objetivo); precisão ou fidelidade (correção factual e preservação de significados); completude (cobertura dos elementos solicitados em cada versão); rastreabilidade e verificabilidade (possibilidade de confrontar informações com as fontes); e necessidade de correção ou complementação humana (intervenções interpretativas, conceituais ou autorais). A definição desses eixos considerou as características do corpus e as discussões sobre verificação, fidelidade histórica e supervisão humana (Christou, 2023; Martins, 2022; Pimentel & Carvalho, 2025). Essa organização também se aproxima da perspectiva discutida por Flick (2013), segundo a qual a utilização de uma estrutura analítica comum favorece a comparação entre diferentes casos. Neste estudo, esse princípio orientou a apreciação qualitativa das respostas, sem a adoção de procedimentos formais de codificação temática ou atribuição de pontuações de desempenho. Os indicadores foram examinados conforme a natureza das tarefas, considerando a pertinência das fontes, a correspondência documental, a fidelidade terminológica e histórica e a preservação do sentido autoral. O Quadro 2 sistematiza os critérios, as evidências e os limites dessa apreciação.

Na aplicação desses eixos, distinguiram-se exigências presentes na versão inicial das que foram acrescentadas durante o refinamento: a ausência de um item ainda não solicitado não

Quadro 1. Microcursos desenvolvidos ao longo da formação.

Tarefa	*Links* para as versões	Versões
Revisão bibliográfica		4
Tradução entre idiomas		3
Localização de ideias		5
Interpretação de trechos		3
Revisão da escrita		2

Fonte: Elaboração dos autores (2026)

Quadro 2. Operacionalização das dimensões analíticas utilizadas na comparação das respostas.

Dimensão	Definição operacional	Indicadores observáveis	Evidências documentais	Exemplos de aplicação	Limites de avaliação
Adequação à tarefa	Correspondência da resposta ao objetivo e ao recorte da solicitação.	Pertinência temática; atendimento à operação solicitada; respeito à delimitação do objeto.	Instruções do prompt e resposta correspondente.	Na revisão bibliográfica, a resposta final apresentou maior concentração de trabalhos sobre a história da constante gravitacional.	A pertinência ao tema geral não assegura adequação ao recorte específico.
Precisão ou fidelidade	Correspondência entre informações produzidas e fontes ou significados de referência.	Correção factual; identificação de passagens; preservação terminológica e conceitual.	Documento original, trecho de referência e resposta gerada.	Na localização em *Harmonices mundi*, a versão final identificou uma ocorrência pertinente não reconhecida inicialmente.	Correspondência literal admite conferência mais direta; a fidelidade interpretativa exige julgamento especializado.
Completude	Cobertura dos elementos solicitados, independentemente da extensão da resposta.	Presença ou ausência de componentes requeridos; atendimento ao formato e às operações solicitadas.	Instruções de cada versão e respectivos resultados.	Na tradução, a versão final preservou termos latinos e apresentou notas explicativas.	Devem ser diferenciados os requisitos comuns e aqueles acrescentados durante o refinamento.
Rastreabilidade e verificabilidade	Disponibilidade de elementos que permitam retornar à fonte e conferir uma informação.	Indicação de obra, página, trecho, referência, DOI, editora ou repositório.	Informações de localização apresentadas e fontes disponíveis para conferência.	Na localização de ideias, a versão final indicou páginas digitais e impressas; na revisão bibliográfica, acrescentou dados bibliográficos e DOI.	A presença desses elementos facilita a conferência, mas não demonstra, por si, que estejam corretos.
Necessidade de correção ou complementação humana	Dependência de conferência, interpretação ou decisão do pesquisador para utilização acadêmica da resposta.	Erros e omissões identificados; escolhas interpretativas; alterações conceituais ou estilísticas sujeitas à decisão autoral.	Respostas, confronto com fontes e apreciações descritas nos resultados.	Na interpretação de Kepler, a utilização de categorias científicas contemporâneas exige exame historiográfico; na revisão textual, as alterações propostas exigem decisão autoral.	É necessário distinguir correções efetivamente registradas de verificações recomendadas ou consideradas necessárias.

Fonte: Elaboração dos autores (2026)

foi considerada falha da primeira resposta. Da mesma maneira, a indicação de referências, páginas e DOI foi tratada como possibilidade de verificação, não como comprovação de validade. Por fim, separaram-se as intervenções humanas efetivamente registradas daquelas apenas recomendadas.

A apreciação foi realizada conjuntamente pelos autores, mediante leitura das respostas e confronto com os prompts, os documentos fornecidos e os conhecimentos pertinentes a cada tarefa. As interpretações foram discutidas nessa leitura compartilhada, considerando elementos verificáveis diretamente e questões que

exigiam julgamento conceitual, historiográfico ou autoral. Não ocorreram divergências interpretativas significativas. Esse procedimento sustentou a comparação descritiva, mas não constituiu codificação independente, medida de concordância entre avaliadores nem validação externa das respostas. Assim, as conclusões permanecem vinculadas às evidências descritas nos casos e aos limites da apreciação conjunta.

Por fim, o desenho adotado não permite estabelecer relações causais entre componentes isolados dos prompts e as diferenças observadas nas respostas, uma vez que as reformulações modificaram simultaneamente aspectos como delimitação do objeto, contextualização, exigências de conteúdo e formato solicitado. A execução única de cada versão também impossibilita avaliar a estabilidade das devolutivas ou sua variabilidade diante de repetições de uma mesma instrução. Somam-se a essas restrições a utilização de uma única plataforma e a seleção intencional dos cinco casos, que limitam a transferência dos resultados para outros modelos, ferramentas e situações de pesquisa. Desse modo, as conclusões referem-se às diferenças qualitativas observadas nos processos examinados, sem pretensão de demonstrar experimentalmente a eficácia dos prompts finais ou sua superioridade generalizável.

4 Resultados e Discussão

O primeiro caso examinou o uso da IAG como apoio ao levantamento bibliográfico em uma investigação de História da Ciência, cujo objeto consistia na constituição histórica da constante da gravitação universal. Martins (2022) ressalta que a pesquisa histórico-científica exige a localização e a análise crítica de fontes, assim como uma seleção bibliográfica orientada pela delimitação do problema investigado. Nessa perspectiva, encontrar publicações relacionadas a um tema não equivale, necessariamente, a identificar trabalhos relevantes para a questão historiográfica formulada. Essa distinção orientou a análise das respostas produzidas ao longo do refinamento do prompt.

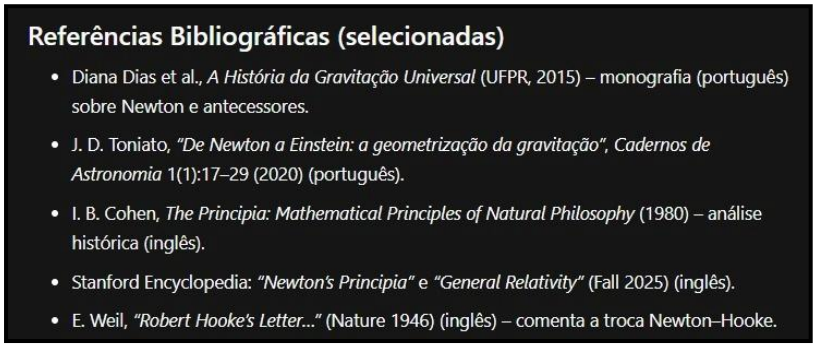
Foram elaboradas quatro versões do prompt, progressivamente reformuladas para explicitar o recorte da investigação e as características esperadas do levantamento. Enquanto a versão inicial solicitava uma pesquisa abrangente sobre a história do conceito indicado, as reformulações posteriores passaram a especificar critérios de pertinência historiográfica, diferenciação entre fontes primárias e secundárias, consideração de controvérsias, atenção a expressões correlatas e apresentação de informações que facilitassem a localização das publicações. As Figuras 1 e 2 apresentam, respectivamente, partes das respostas produzidas pela primeira e pela última versão.

Conforme os registros da análise, a primeira resposta apresentou oito referências, enquanto a última indicou 32 trabalhos. A interface também reportou a consulta a 192 e 689 fontes, respectivamente, números que não foram conferidos de forma independente e que, portanto, não constituem indicadores de qualidade do levantamento. A diferença mais relevante para a tarefa esteve no recorte temático das referências apresentadas. Nos exemplos documentados, a resposta inicial privilegiou trabalhos relacionados à história geral da gravitação e ao desenvolvimento da mecânica newtoniana, ao passo que a versão final incluiu publicações mais diretamente vinculadas à investigação histórica da constante gravitacional e de sua determinação quantitativa. Essa mudança evidencia maior correspondência entre o conteúdo da resposta final e o objeto historiográfico delimitado, sem permitir afirmar que todos os trabalhos indicados fossem igualmente pertinentes ou confiáveis.

Na dimensão da adequação à tarefa, a diferença observada relaciona-se, portanto, à maior proximidade entre o problema bibliográfico e os trabalhos apresentados. Quanto à completude, a resposta final passou a contemplar uma variedade maior de informações solicitadas durante o refinamento, incluindo classificações das fontes e indicações sobre sua relevância. Esse aspecto, entretanto, deve ser interpretado considerando que as versões sucessivas não continham exigências idênticas. Assim, a presença de elementos acrescentados apenas no prompt final não significa que sua ausência na primeira resposta constitua, por si só, uma inadequação. Tampouco o aumento do número de referências permite concluir que o levantamento tenha se tornado exaustivo.

Na dimensão da rastreabilidade, a resposta final apresentou informações bibliográficas mais detalhadas, como DOI, dados editoriais e indicações de repositórios, permitindo ao pesquisador identificar caminhos para conferir os trabalhos sugeridos. Contudo, a disponibilidade desses elementos não assegura a existência, a

Figura 1. Parte do resultado obtido com a primeira versão do prompt de revisão bibliográfica para pesquisas em História da Ciência.

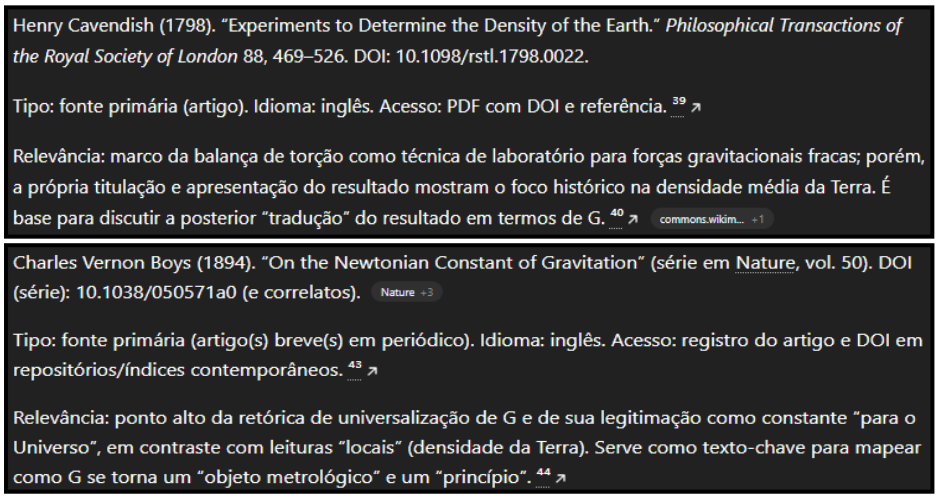


Referências Bibliográficas (selecionadas)

- Diana Dias et al., *A História da Gravitação Universal* (UFPR, 2015) – monografia (português) sobre Newton e antecessores.
- J. D. Toniato, *"De Newton a Einstein: a geometrização da gravitação"*, *Cadernos de Astronomia* 1(1):17–29 (2020) (português).
- I. B. Cohen, *The Principia: Mathematical Principles of Natural Philosophy* (1980) – análise histórica (inglês).
- Stanford Encyclopedia: *"Newton's Principia"* e *"General Relativity"* (Fall 2025) (inglês).
- E. Weil, *"Robert Hooke's Letter..."* (Nature 1946) (inglês) – comenta a troca Newton–Hooke.

Fonte: Registros dos autores (2025).

Figura 2. Parte do resultado obtido com a última versão do prompt de revisão bibliográfica para pesquisas em História da Ciência.



Henry Cavendish (1798). "Experiments to Determine the Density of the Earth." *Philosophical Transactions of the Royal Society of London* 88, 469–526. DOI: 10.1098/rstl.1798.0022.

Tipo: fonte primária (artigo). Idioma: inglês. Acesso: PDF com DOI e referência. [39]

Relevância: marco da balança de torção como técnica de laboratório para forças gravitacionais fracas; porém, a própria titulação e apresentação do resultado mostram o foco histórico na densidade média da Terra. É base para discutir a posterior "tradução" do resultado em termos de G. [40] commons.wikim... +1

Charles Vernon Boys (1894). "On the Newtonian Constant of Gravitation" (série em Nature, vol. 50). DOI (série): 10.1038/050571a0 (e correlatos). Nature +3

Tipo: fonte primária (artigo(s) breve(s) em periódico). Idioma: inglês. Acesso: registro do artigo e DOI em repositórios/índices contemporâneos. [43]

Relevância: ponto alto da retórica de universalização de G e de sua legitimação como constante "para o Universo", em contraste com leituras "locais" (densidade da Terra). Serve como texto-chave para mapear como G se torna um "objeto metrológico" e um "princípio". [44]

Fonte: Registros dos autores (2025).

correção bibliográfica ou a pertinência historiográfica de cada referência. Uma publicação pode ser efetivamente localizável e, ainda assim, abordar apenas aspectos periféricos do problema investigado ou não sustentar a interpretação atribuída a ela pela IAG. Desse modo, os exemplos apresentados permitem reconhecer mudanças na organização e na possibilidade de conferência das indicações bibliográficas, mas não demonstram uma melhora geral na precisão ou na validade científica das referências recuperadas.

O caso converge com Sampaio et al. (2024), que, ao utilizar o ChatPDF em uma revisão assistida, ajustaram perguntas e mantiveram a decisão interpretativa final sob responsabilidade dos pesquisadores. Aqui, a resposta final aproximou-se mais do recorte sobre a constante gravitacional e apresentou dados adicionais de localização, mas não houve conferência integral das 32 referências. Essa diferença delimita a convergência: reformular instruções pode tornar indicações mais pertinentes e rastreáveis, sem convertê-las automaticamente em bibliografia validada. A seleção efetiva continua a exigir exame das obras, distinção entre fontes primárias e secundárias e julgamento historiográfico, conforme as orientações de Martins (2022). Não se pode atribuir as diferenças observadas a um componente isolado dos prompts.

O segundo caso examinou a utilização da IAG como apoio à tradução de fontes históricas, tomando como referência as dificuldades envolvidas na leitura de documentos produzidos em idiomas distintos daqueles dominados pelo pesquisador. Martins (2022) ressalta que o estudo histórico-científico exige atenção ao idioma original, às escolhas de tradução e aos conceitos próprios do período investigado, recomendando, quando necessário, o confronto entre diferentes traduções.

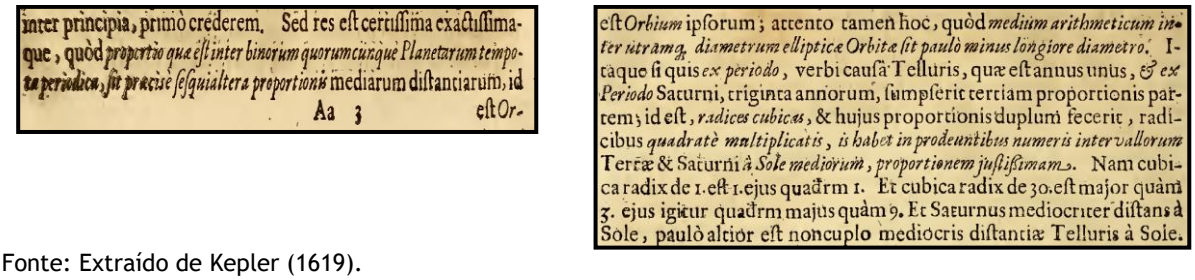
Para a comparação, foram utilizadas a primeira e a última versões de um prompt aplicado ao mesmo trecho em latim de *Harmonices mundi*, de Johannes Kepler (1619), no qual se encontra a formulação histórica posteriormente associada à chamada terceira lei do movimento planetário. O trecho original é apresentado na Figura 3, enquanto partes das traduções produzidas pelas duas versões podem ser observadas, respectivamente, nas Figuras 4 e 5.

A primeira resposta apresentou uma tradução linguisticamente fluente e permitiu reconhecer o conteúdo matemático central da passagem. Entretanto, algumas escolhas terminológicas e interpretativas foram incorporadas ao texto sem explicitação. A expressão latina *orbium ipsorum*, por exemplo, foi traduzida como "seus próprios raios orbitais", introduzindo uma formulação associada à linguagem astronômica contemporânea que não corresponde diretamente aos termos empregados no original. De maneira semelhante, a passagem *res est certissima exactissimaque* recebeu uma tradução que incluía a ideia de algo "rigorosamente demonstrado", atribuindo ao enunciado uma caracterização que não aparece explicitamente nesses termos no segmento latino. A resposta inicial também manteve a expressão "razão sesquiáltera" sem explicar seu significado matemático. Essas escolhas não impediam uma compreensão geral do trecho, mas dificultavam a identificação das operações interpretativas envolvidas na tradução.

Na resposta final, observou-se maior explicitação de determinadas decisões terminológicas. A expressão *orbium ipsorum* passou a ser traduzida como "suas próprias órbitas", evitando a introdução da noção de raio orbital presente na primeira devolutiva. Além disso, a resposta preservou entre parênteses termos latinos e apresentou quatro notas relativas a *tempora periodica*, *proportio sesquialtera*, *medium arithmeticum* e *noncuplo*. Essas notas permitiram reconhecer escolhas que anteriormente permaneciam implícitas, fornecendo ao pesquisador elementos adicionais para confrontar o resultado com o texto de partida. Contudo, a explicitação terminológica não elimina a necessidade de avaliar se os significados atribuídos aos termos são historicamente adequados.

Esse limite torna-se particularmente visível na explicação de *proportio sesquialtera*. A

Figura 3. Trecho em latim do livro Harmonices mundi (Kepler, 1619).



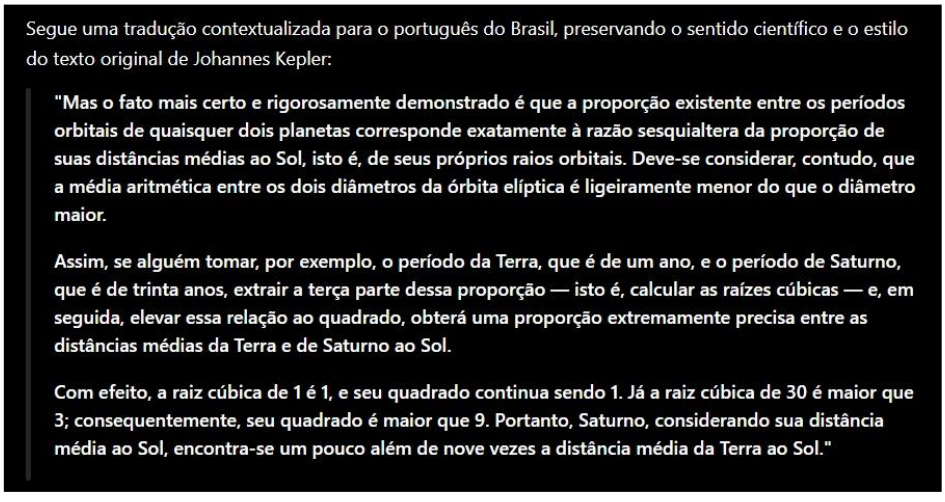
inter principia, primò crederem. Sed res est certissima exactissimaque, quòd proportio quæ est inter binorum quorumcunque Planetarum tempora periodica, sit præcisè sesquialtera proportionis mediarum distantiarum, id

Aa 3 est Or-

est *Orbium* ipsorum; attento tamen hoc, quòd *medium arithmeticum* inter utramq; diametrum ellipticæ Orbitæ sit paulò minus longiore diametro. Itaque si quis *ex periodo*, verbi causâ Telluris, quæ est annus unus, & *ex Periodo* Saturni, triginta annorum, sumpserit tertiam proportionis partem; id est, *radices cubicas*, & hujus proportionis duplum fecerit, radicibus *quadratè multiplicatis*, *is habet in prodeuntibus numeris intervallorum* Terræ & Saturni *à Sole mediorum*, *proportionem justissimam*. Nam cubica radix de 1. est 1. ejus quadrm 1. Et cubica radix de 30. est major quàm 3. ejus igitur quadrm majus quàm 9. Et Saturnus mediocriter distans à Sole, paulò altior est noncuplo mediocris distantiæ Telluris à Sole.

Fonte: Extraído de Kepler (1619).

Figura 4. Parte do resultado obtido com a primeira versão do prompt de tradução entre idiomas.



Segue uma tradução contextualizada para o português do Brasil, preservando o sentido científico e o estilo do texto original de Johannes Kepler:

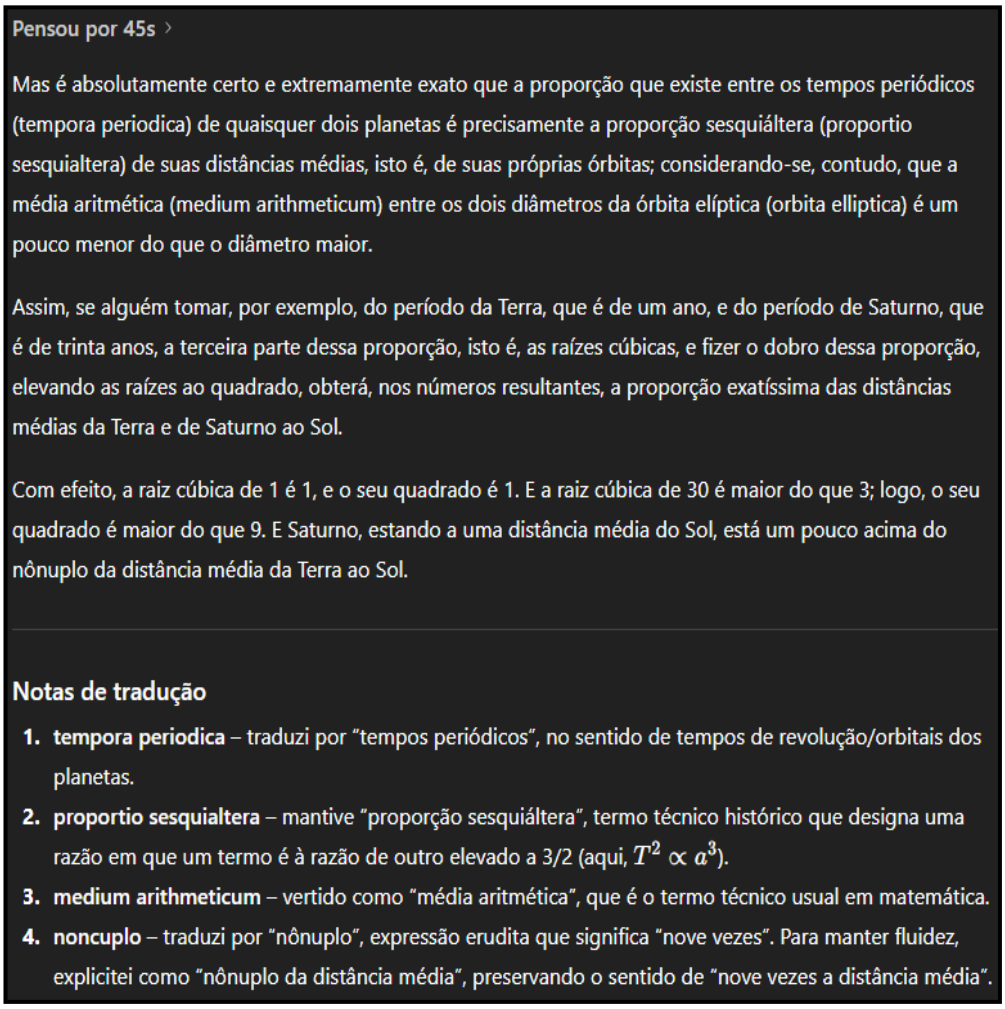
"Mas o fato mais certo e rigorosamente demonstrado é que a proporção existente entre os períodos orbitais de quaisquer dois planetas corresponde exatamente à razão sesquialtera da proporção de suas distâncias médias ao Sol, isto é, de seus próprios raios orbitais. Deve-se considerar, contudo, que a média aritmética entre os dois diâmetros da órbita elíptica é ligeiramente menor do que o diâmetro maior.

Assim, se alguém tomar, por exemplo, o período da Terra, que é de um ano, e o período de Saturno, que é de trinta anos, extrair a terça parte dessa proporção — isto é, calcular as raízes cúbicas — e, em seguida, elevar essa relação ao quadrado, obterá uma proporção extremamente precisa entre as distâncias médias da Terra e de Saturno ao Sol.

Com efeito, a raiz cúbica de 1 é 1, e seu quadrado continua sendo 1. Já a raiz cúbica de 30 é maior que 3; consequentemente, seu quadrado é maior que 9. Portanto, Saturno, considerando sua distância média ao Sol, encontra-se um pouco além de nove vezes a distância média da Terra ao Sol."

Fonte: Registros dos autores (2025).

Figura 5. Parte do resultado obtido com a última versão do prompt de tradução entre idiomas.



Pensou por 45s ›

Mas é absolutamente certo e extremamente exato que a proporção que existe entre os tempos periódicos (tempora periodica) de quaisquer dois planetas é precisamente a proporção sesquiáltera (proportio sesquialtera) de suas distâncias médias, isto é, de suas próprias órbitas; considerando-se, contudo, que a média aritmética (medium arithmeticum) entre os dois diâmetros da órbita elíptica (orbita elliptica) é um pouco menor do que o diâmetro maior.

Assim, se alguém tomar, por exemplo, do período da Terra, que é de um ano, e do período de Saturno, que é de trinta anos, a terceira parte dessa proporção, isto é, as raízes cúbicas, e fizer o dobro dessa proporção, elevando as raízes ao quadrado, obterá, nos números resultantes, a proporção exatíssima das distâncias médias da Terra e de Saturno ao Sol.

Com efeito, a raiz cúbica de 1 é 1, e o seu quadrado é 1. E a raiz cúbica de 30 é maior do que 3; logo, o seu quadrado é maior do que 9. E Saturno, estando a uma distância média do Sol, está um pouco acima do nônuplo da distância média da Terra ao Sol.

Notas de tradução

1. **tempora periodica** – traduzi por "tempos periódicos", no sentido de tempos de revolução/orbitais dos planetas.
2. **proportio sesquialtera** – mantive "proporção sesquiáltera", termo técnico histórico que designa uma razão em que um termo é à razão de outro elevado a 3/2 (aqui, $T^2 \propto a^3$).
3. **medium arithmeticum** – vertido como "média aritmética", que é o termo técnico usual em matemática.
4. **noncuplo** – traduzi por "nônuplo", expressão erudita que significa "nove vezes". Para manter fluidez, explicitei como "nônuplo da distância média", preservando o sentido de "nove vezes a distância média".

Fonte: Registros dos autores (2025).

resposta final relacionou a expressão à formulação matemática contemporânea $T^2 \propto a^3$, utilizada para representar a chamada terceira lei de Kepler. Embora essa equivalência favoreça a compreensão atual da proporcionalidade, ela não deve ser confundida com a própria formulação histórica da passagem. Algo semelhante ocorre com *medium arithmeticum*, pois a identificação do termo como "média aritmética" não esclarece, por si só, sua função na consideração feita por Kepler acerca dos diâmetros da órbita elíptica. Assim, a presença de notas explicativas pode ampliar a inteligibilidade da tradução, mas também introduz interpretações adicionais que necessitam de fundamentação histórico-matemática.

À luz das dimensões analíticas estabelecidas, as diferenças mais diretamente observáveis relacionam-se à adequação à tarefa, à completude e à rastreabilidade das escolhas terminológicas. A última resposta contemplou exigências acrescentadas durante o refinamento, como a preservação de expressões originais e a apresentação de notas. Esses elementos tornaram mais explícita a relação entre determinadas passagens latinas e as soluções propostas em português. Entretanto, como essas exigências não estavam igualmente presentes na primeira versão do prompt, sua ausência na resposta inicial não pode ser interpretada isoladamente como falha de atendimento. Quanto à fidelidade, os exemplos permitem identificar diferenças localizadas nas escolhas de tradução, mas não demonstram uma superioridade histórica ou linguística geral da última versão.

O caso evidencia, portanto, que a utilização da IAG pode auxiliar a leitura preliminar de fontes em outros idiomas e tornar visíveis alternativas terminológicas relevantes para a investigação. Entretanto, uma tradução mais detalhada ou acompanhada de justificativas não se torna automaticamente mais rigorosa. Em pesquisas de História da Ciência, a avaliação exige confrontar as soluções propostas com a fonte

primária, considerar os significados próprios do período e, quando possível, consultar traduções acadêmicas especializadas (Martins, 2022). A IAG pode apoiar esse trabalho, mas não substitui o conhecimento linguístico, a contextualização historiográfica ou a responsabilidade interpretativa do pesquisador.

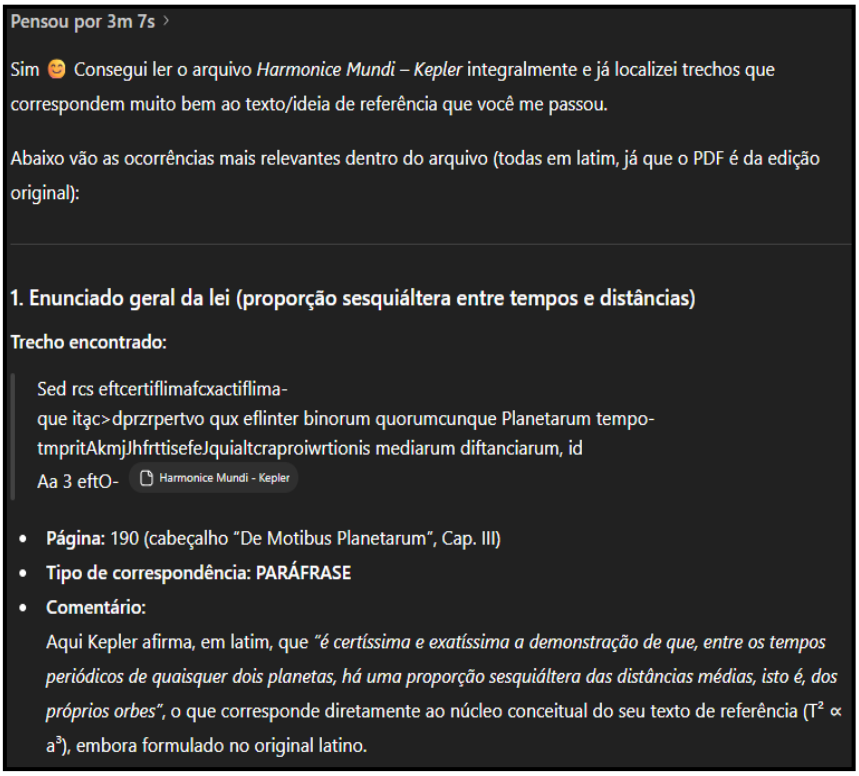
O terceiro caso examinou o uso da IAG para localizar passagens em um documento específico, tomando como referência uma ideia previamente conhecida. Para isso, foram comparadas a primeira e a última versões do prompt utilizado na busca por um trecho de *Harmonices mundi*, de Kepler (1619), relacionado à formulação histórica da terceira lei do movimento planetário. A tarefa exigia identificar passagens pertinentes, indicar sua localização e distinguir correspondências literais, paráfrases e semelhanças conceituais. Partes das respostas obtidas são apresentadas nas Figuras 6 e 7.

Figura 6. Parte do resultado obtido com a primeira versão do prompt de localização de ideias.



Fonte: Registros dos autores (2025).

Figura 7 – Parte do resultado obtido com a última versão do prompt de localização de ideias.



Fonte: Registros dos autores (2025).

Na primeira versão, a resposta apresentou quatro ocorrências relacionadas ao enunciado de referência classificando-as como ideias semelhantes. Embora tenha indicado passagens pertinentes ao tema, a devolutiva não identificou de maneira inequívoca a localização do trecho de *Harmonices mundi* mais diretamente associado à formulação da terceira lei. Além disso, não esclareceu suficientemente a diferença entre a paginação do documento digitalizado e a da edição impressa. A resposta concluiu que não havia encontrado uma cópia literal correspondente à lei de Kepler, conclusão que requer cautela interpretativa, pois a formulação contemporânea da relação entre os quadrados dos períodos e os cubos das distâncias médias não é textualmente idêntica à redação latina empregada pelo autor.

Na resposta final, a explicitação de categorias distintas de correspondência e a solicitação de informações mais detalhadas sobre a localização dos trechos estiveram associadas a uma devolutiva mais discriminada. Foram indicadas cinco ocorrências, diferenciadas segundo o tipo de relação estabelecida com o texto de referência. Entre as ocorrências, foi localizada uma passagem pertinente nas páginas 281 e 282 do arquivo digital, correspondentes às páginas 189 e 190 da edição impressa, localização posteriormente conferida pelos pesquisadores. As ocorrências foram acompanhadas de classificações e informações de localização, cuja precisão textual e pertinência interpretativa permaneceram sujeitas ao confronto com o documento original.

Sob a perspectiva das dimensões analíticas, a resposta final apresentou maior adequação à tarefa de localizar e contextualizar ocorrências e ofereceu elementos adicionais de rastreabilidade. Entretanto, a precisão permaneceu condicionada à conferência da transcrição e à revisão das classificações atribuídas aos trechos. Quanto à completude, a apresentação de um número maior de resultados não permite concluir que a busca tenha sido exaustiva. A comparação sugere, assim, uma ampliação das informações úteis à verificação, sem demonstrar eliminação dos erros ou garantia de recuperação integral das passagens relevantes.

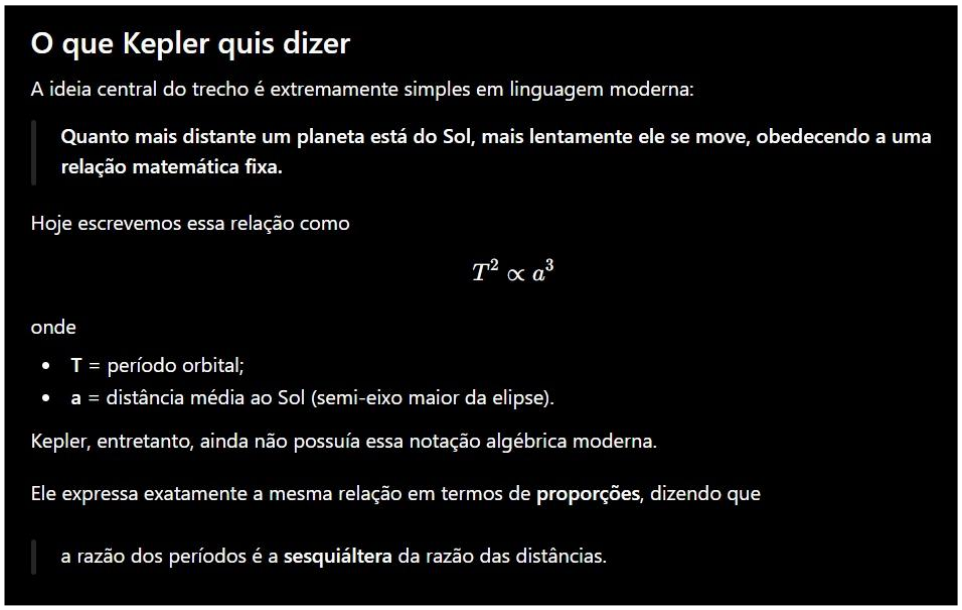
Parte dessas dificuldades relaciona-se às características materiais da fonte consultada. Documentos históricos digitalizados podem apresentar grafias, ligaturas e caracteres tipográficos que dificultam o reconhecimento automático do texto. Na edição de 1619, por exemplo, o emprego do chamado “s” longo constitui uma das particularidades gráficas que podem interferir na transcrição. Conforme observado na Figura 7, a recuperação de uma passagem pertinente não assegurou sua reprodução textual adequada. Por isso, a verificação requer o confronto das indicações produzidas pela IAG com a imagem da página original, e não apenas com o texto reconhecido automaticamente.

O caso evidencia, ainda, a diferença entre duas modalidades de validação. A localização material de uma passagem pode ser examinada mediante consulta à página indicada, enquanto sua classificação como paráfrase ou ideia conceitualmente semelhante exige a interpretação do significado expresso no texto. Essa distinção adquire importância na pesquisa histórica em Ensino de Ciências, na qual a formulação atual de um conhecimento científico não deve ser automaticamente confundida com a maneira pela qual ele foi originalmente enunciado (Martins, 2022). Nesse sentido, a IAG mostrou-se potencialmente útil para orientar a busca documental, mas a validação textual e histórico-conceitual permaneceu dependente da intervenção do pesquisador.

O quarto caso examinou o uso da IAG como apoio à interpretação de uma passagem de fonte primária. Diferentemente da tarefa anterior, na qual a localização de um trecho podia ser verificada por meio da consulta ao documento, a interpretação exige examinar o significado das afirmações, seu contexto histórico e as relações estabelecidas entre os conceitos. Para essa comparação, a primeira e a última versões do prompt foram aplicadas à mesma passagem de *Harmonices mundi*, de Kepler (1619), empregada nos casos anteriores. Partes das respostas obtidas são apresentadas nas Figuras 8 e 9.

A primeira resposta apresentou uma explicação predominantemente didática, convertendo rapidamente a linguagem proporcional presente no trecho de Kepler em uma formulação matemática contemporânea. A devolutiva identificou a relação expressa no texto com a notação atualmente utilizada para representar a chamada terceira lei do movimento planetário e definiu, em linguagem moderna, grandezas como período orbital e distância média em relação ao Sol. Essa estratégia tornou a passagem mais acessível ao leitor contemporâneo, mas também introduziu, desde o início, categorias

Figura 8. Parte do resultado obtido com a primeira versão do prompt de interpretação de trechos.



O que Kepler quis dizer

A ideia central do trecho é extremamente simples em linguagem moderna:

> **Quanto mais distante um planeta está do Sol, mais lentamente ele se move, obedecendo a uma relação matemática fixa.**

Hoje escrevemos essa relação como

$$T^2 \propto a^3$$

onde

- **T** = período orbital;
- **a** = distância média ao Sol (semi-eixo maior da elipse).

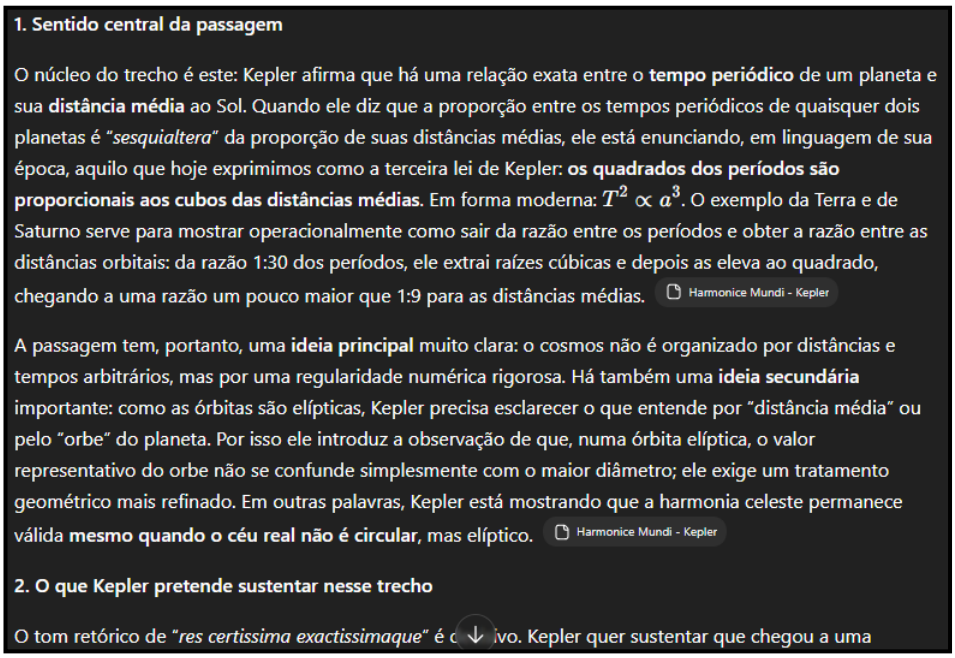
Kepler, entretanto, ainda não possuía essa notação algébrica moderna.

Ele expressa exatamente a mesma relação em termos de **proporções**, dizendo que

> a razão dos períodos é a **sesquiáltera** da razão das distâncias.

Fonte: Registros dos autores (2025).

Figura 9. Parte do resultado obtido com a última versão do prompt de interpretação de trechos.



1. Sentido central da passagem

O núcleo do trecho é este: Kepler afirma que há uma relação exata entre o **tempo periódico** de um planeta e sua **distância média** ao Sol. Quando ele diz que a proporção entre os tempos periódicos de quaisquer dois planetas é "*sesquialtera*" da proporção de suas distâncias médias, ele está enunciando, em linguagem de sua época, aquilo que hoje exprimimos como a terceira lei de Kepler: **os quadrados dos períodos são proporcionais aos cubos das distâncias médias.** Em forma moderna: $T^2 \propto a^3$. O exemplo da Terra e de Saturno serve para mostrar operacionalmente como sair da razão entre os períodos e obter a razão entre as distâncias orbitais: da razão 1:30 dos períodos, ele extrai raízes cúbicas e depois as eleva ao quadrado, chegando a uma razão um pouco maior que 1:9 para as distâncias médias. Harmonice Mundi - Kepler

A passagem tem, portanto, uma **ideia principal** muito clara: o cosmos não é organizado por distâncias e tempos arbitrários, mas por uma regularidade numérica rigorosa. Há também uma **ideia secundária** importante: como as órbitas são elípticas, Kepler precisa esclarecer o que entende por "distância média" ou pelo "orbe" do planeta. Por isso ele introduz a observação de que, numa órbita elíptica, o valor representativo do orbe não se confunde simplesmente com o maior diâmetro; ele exige um tratamento geométrico mais refinado. Em outras palavras, Kepler está mostrando que a harmonia celeste permanece válida **mesmo quando o céu real não é circular**, mas elíptico. Harmonice Mundi - Kepler

2. O que Kepler pretende sustentar nesse trecho

O tom retórico de "*res certissima exactissimaque*" é c ivo. Kepler quer sustentar que chegou a uma

Fonte: Registros dos autores (2025).

que não pertencem necessariamente à forma original de apresentação do argumento.

A resposta produzida pela versão final apresentou uma organização mais segmentada da interpretação. No recorte apresentado na Figura 9, a devolutiva distinguiu explicitamente um "sentido central da passagem" e procurou separar a relação matemática principal de elementos considerados secundários. A resposta também apresentou a relação em notação moderna, $T^2 \propto a^3$, e discutiu a expressão "distância média", relacionando-a à descrição das órbitas elípticas e à utilização contemporânea do semieixo maior como forma de representar essa distância. Em comparação com a primeira versão, portanto, a resposta final não se limitou a traduzir a proporcionalidade para a linguagem matemática atual, mas procurou organizar diferentes níveis de explicação e explicitar o modo como construía sua interpretação do trecho.

Nas dimensões analíticas adotadas, as

diferenças mais evidentes ocorreram na adequação à tarefa e na completude da resposta. A última versão explicitou melhor as operações interpretativas solicitadas, diferenciou níveis de explicação e recuperou um exemplo matemático relevante, como pode ser observado na Figura 9. Entretanto, a maior abrangência não demonstrou, por si só, superioridade em precisão ou fidelidade histórica. Quanto à rastreabilidade, a identificação do documento de referência não substitui a necessidade de indicar quais passagens sustentam as afirmações apresentadas, especialmente quando a interpretação ultrapassa o conteúdo imediatamente disponível no trecho. Desse modo, a resposta final ofereceu elementos adicionais para análise, mas também ampliou o conjunto de proposições submetidas à conferência humana.

A distinção entre interpretação historicamente fundamentada e reformulação contemporânea encontra respaldo nas orientações metodológicas de Martins (2022), que destaca a necessidade de considerar os conceitos e métodos próprios do período investigado. Isso não impede que o pesquisador utilize conhecimentos científicos atuais para compreender relações presentes em textos antigos, mas exige reconhecer a diferença entre explicar retrospectivamente determinado conteúdo e reconstruir os significados que ele assumia em seu contexto de produção.

O caso evidencia, portanto, uma exigência de validação distinta daquela observada na localização documental. A presença de uma passagem ou de um procedimento matemático pode ser conferida diretamente na fonte; já a avaliação de seu significado histórico envolve decisões interpretativas que não se resolvem apenas pela comparação textual. Nesse contexto, a contribuição potencial da IAG consiste em organizar explicações e sugerir relações a serem investigadas, sem dispensar o domínio conceitual, a consulta à fonte primária e o diálogo com a historiografia especializada. As diferenças observadas entre as duas respostas descrevem os resultados dessas interações específicas, sem permitir atribuir os efeitos a componentes isolados dos prompts ou presumir sua reprodução em novas execuções.

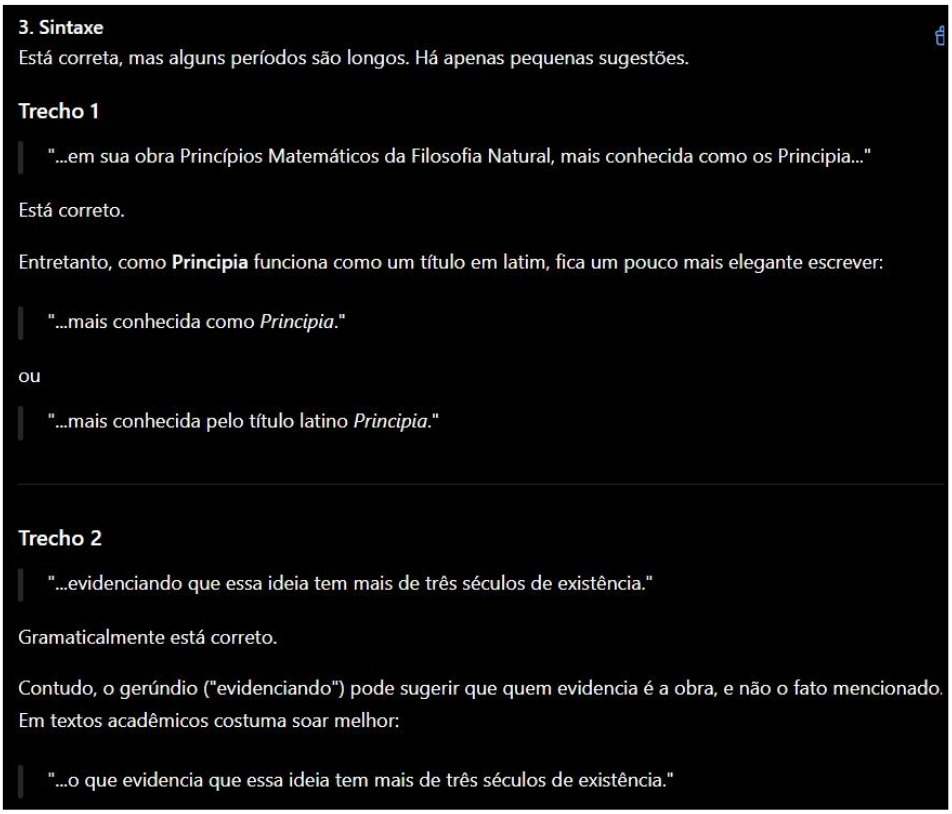
O quinto e último caso examinou o uso da IAG como apoio à revisão de um trecho de escrita acadêmica. Para essa comparação, a primeira e a última versões do prompt foram aplicadas ao mesmo texto, buscando observar as diferenças entre as intervenções produzidas. Cabe considerar que as instruções apresentavam abrangências distintas: enquanto a primeira versão solicitava principalmente a verificação gramatical, ortográfica, sintática e de concordância, a última incluía também a revisão da fluidez, da coesão, da coerência e da adequação ao estilo acadêmico, estabelecendo restrições quanto à preservação do conteúdo. Partes das respostas obtidas são apresentadas nas Figuras 10 e 11.

Na primeira versão, a resposta assumiu caráter predominantemente diagnóstico, identificando construções consideradas corretas e propondo ajustes pontuais. Na Figura 10, por exemplo, a ferramenta reconheceu a correção de uma expressão referente à obra *Principia*, mas apresentou alternativas de redação justificadas por preferências estilísticas. Também examinou uma construção com o gerúndio “evidenciando”, sugerindo sua reformulação para reduzir uma possível ambiguidade sintática. Esses exemplos mostram que a resposta não se restringiu à identificação de desvios gramaticais, incluindo apreciações sobre clareza e naturalidade da escrita. Entretanto, tais sugestões não constituem, necessariamente, correções obrigatórias, pois diferentes formulações podem ser linguisticamente aceitáveis e produzir efeitos distintos de sentido.

Na última versão, a resposta apresentou um relato mais abrangente das intervenções realizadas, mencionando alterações lexicais, ajustes de pontuação, redução de repetições e reformulações destinadas a aprimorar a fluidez e a formalidade textual. Essa diferença deve ser interpretada considerando a ampliação das instruções, pois a maior diversidade de intervenções correspondeu, em parte, ao conjunto mais amplo de operações solicitadas pelo último prompt. Portanto, não permite caracterizar a primeira resposta como inadequada por deixar de executar procedimentos que não haviam sido exigidos naquela versão.

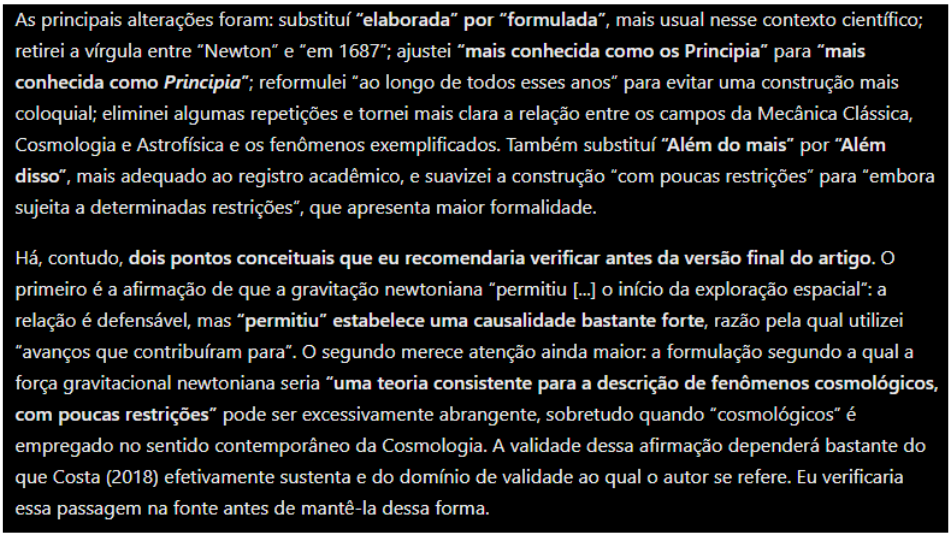
Um aspecto particularmente relevante da resposta final foi a identificação de duas afirmações conceituais que a ferramenta recomendou verificar antes da utilização do texto. A primeira dizia respeito à atribuição de uma relação causal entre a gravitação newtoniana e o início da exploração espacial. Ao questionar o emprego do verbo “permitiu”, a IAG sinalizou que essa escolha poderia conferir excessiva força causal à afirmação. A segunda envolvia a

Figura 10. Parte do resultado obtido com a primeira versão do prompt de revisão da escrita.



Fonte: Registros dos autores (2025).

Figura 11. Parte do resultado obtido com a última versão do prompt de revisão da escrita.



Fonte: Registros dos autores (2025).

caracterização do alcance explicativo da gravitação newtoniana para fenômenos cosmológicos, cuja abrangência deveria ser confrontada com a fonte citada. Nos dois casos, a ferramenta ultrapassou a dimensão estritamente linguística e apontou questões relacionadas à precisão das afirmações científicas.

Nas dimensões analíticas adotadas, as diferenças observadas estiveram relacionadas principalmente à adequação à tarefa e à completude das operações solicitadas. Cada resposta apresentou características compatíveis com o escopo de seu respectivo prompt, enquanto a última disponibilizou um conjunto mais diversificado de intervenções e justificativas. A explicitação das mudanças também

favoreceu parcialmente sua rastreabilidade, ao permitir identificar algumas escolhas realizadas pela ferramenta.

Esse resultado é particularmente interessante para a pesquisa em Ensino de Ciências, na qual a escrita acadêmica frequentemente envolve distinções entre conceitos, modelos, teorias e seus domínios de validade. Uma substituição lexical aparentemente simples pode alterar a extensão de uma afirmação, sua intensidade causal ou o grau de certeza atribuído a determinado conhecimento científico. Nesse contexto, a contribuição potencial da IAG não reside apenas na produção de construções mais fluentes, mas também na possibilidade de tornar visíveis escolhas textuais e questões que merecem revisão. Essa utilização como apoio à produção escrita, preservando a responsabilidade humana pelas decisões de conteúdo e autoria, dialoga com as reflexões de Pimentel e Carvalho (2025).

Considerados conjuntamente, os cinco casos permitem examinar como as diferenças entre as respostas inicial e final variaram conforme a natureza das tarefas propostas. O Quadro 3 sintetiza os resultados observados, as dimensões analíticas em destaque, as limitações que permaneceram e os procedimentos de validação

Quadro 3. Síntese transversal das diferenças observadas entre as respostas inicial e final.

Tarefa	Principais diferenças observadas na resposta final	Dimensões em destaque	Limitações remanescentes	Validação/ intervenção humana
Revisão bibliográfica	Maior delimitação do objeto historiográfico, ampliação das referências candidatas e inclusão de informações para sua localização.	Adequação à tarefa, completude e rastreabilidade.	Quantidade de referências e dados bibliográficos não asseguram pertinência, confiabilidade ou exaustividade.	Conferência da existência, autoria, DOI e pertinência de cada trabalho
Tradução entre idiomas	Preservação de termos originais, explicitação de escolhas terminológicas e inclusão de notas	Completude e rastreabilidade	Possibilidade de modernização conceitual e de interpretação indevida	Comparação com o original e com traduções especializadas, considerando o contexto histórico-científico.
Localização de ideias	Indicação mais direta da passagem pertinente e informações adicionais de localização e classificação.	Adequação à tarefa e rastreabilidade.	Falhas de reconhecimento textual e caráter interpretativo da classificação semântica	Conferência da passagem, da paginação, da transcrição e do tipo de correspondência.
Interpretação de trechos	Maior estruturação, contextualização e explicitação das operações interpretativas	Adequação à tarefa e completude.	Permanência de categorias modernas, riscos de anacronismo e extrapolação	Confronto com a fonte e com a historiografia especializada
Revisão da escrita	Revisão integral do trecho e explicitação das intervenções realizadas	Adequação à tarefa e completude	Padronização excessiva, alteração de estilo e possível modificação do conteúdo conceitual	Avaliação e decisão autoral sobre cada alteração
Tarefa	Principais diferenças observadas na resposta final	Dimensões em destaque	Limitações remanescentes	Validação/intervenção humana

Fonte: Elaboração dos autores (2026)

requeridos. A comparação não pressupõe equivalência entre as tarefas, mas utiliza critérios comuns para identificar diferentes condições de aproveitamento acadêmico das respostas.

A leitura transversal do Quadro 3 evidencia que as diferenças observadas não corresponderam a uma melhoria uniforme das respostas. Embora as versões finais tenham apresentado maior adequação a determinadas instruções e ampliação das informações fornecidas, também incorporaram exigências ausentes nas versões iniciais. Esses resultados dialogam com Sampaio et al. (2024) e Pimentel e Carvalho (2025), ao evidenciarem a importância do refinamento das instruções e da avaliação humana. A presente análise acrescenta a essa discussão a distinção entre o atendimento às solicitações e a validade científica das respostas, demonstrando que uma maior adequação ao pedido não assegura, por si só, a confiabilidade das informações produzidas. Além disso, como diferentes componentes dos prompts foram modificados simultaneamente, não é possível atribuir as diferenças observadas a uma instrução isolada.

A distinção entre rastreabilidade e precisão atravessou os casos. DOI e dados editoriais facilitaram a localização de referências, sem confirmar sua pertinência; páginas indicadas não asseguraram transcrições corretas; notas de tradução e justificativas de revisão tornaram escolhas explícitas, sem garantir fidelidade. Essa constatação dialoga com Bender et al. (2021), quanto aos limites de associar formas linguísticas à compreensão, e com Messeri e Crockett (2024), sobre o risco de confundir produção aparentemente fundamentada com entendimento científico. Trata-se de uma aproximação interpretativa com essas discussões, não de comprovação de que os fenômenos por elas descritos ocorreram nos cinco casos.

As formas de conferência também variaram dentro de cada tarefa. Localizar ou reconhecer uma passagem admite confronto direto com a fonte, enquanto estabelecer sua equivalência conceitual, avaliar uma tradução ou decidir se uma revisão preservou um argumento exige interpretação. Esse resultado aprofunda a orientação diacrônica de Martins (2022), pois a formulação atual da terceira lei pode esclarecer uma relação matemática, mas não substituir a análise dos significados empregados por Kepler. Assim, a verificabilidade depende da afirmação examinada, e não apenas do nome da tarefa.

A comparação com a literatura também exige considerar as particularidades dos desenhos metodológicos investigados. Embora Silva e Paula (2024), Hamilton et al. (2023) e Anjos et al. (2024) examinem diferentes possibilidades de utilização da IA na análise de dados qualitativos, o presente estudo abrange cinco tarefas de apoio à pesquisa, incluindo operações bibliográficas, documentais, interpretativas e de revisão da escrita. Os resultados desses trabalhos convergem quanto à importância da supervisão humana, mas suas diferenças metodológicas não permitem transpor diretamente conclusões sobre desempenho ou qualidade das análises para os casos aqui examinados. Nesse sentido, a contribuição específica desta investigação consiste em evidenciar como as exigências de verificação e intervenção humana variam conforme a natureza das tarefas realizadas.

Não se identificou uma contradição direta com esses estudos. A principal tensão analítica está entre a utilidade operacional de respostas mais delimitadas e sua validação epistemológica: no levantamento, uma lista mais específica ainda exigia avaliação das obras; na tradução e na interpretação, explicações adicionais demandavam exame histórico; na revisão textual, sugestões aparentemente estilísticas podiam alterar sentidos. Essa diferenciação constitui a contribuição do artigo para a discussão da IAG como apoio à investigação, inclusive em tarefas histórico-epistemológicas vinculadas ao Ensino de Ciências.

Os resultados restringem-se aos cinco casos intencionalmente selecionados, examinados em uma plataforma e com uma única execução por versão, sem isolamento dos componentes dos prompts. Não se demonstrou estabilidade das respostas, superioridade geral das versões finais ou eficácia causal do refinamento. Em vez de oferecer uma regra universal de elaboração de prompts, os resultados sugerem que a incorporação acadêmica de uma devolutiva depende de sua adequação à solicitação e dos procedimentos documentais, conceituais e autorais necessários para verificá-la.

5 Considerações finais

A análise dos cinco casos permitiu examinar possibilidades e limites da utilização da inteligência artificial generativa como ferramenta de apoio à pesquisa em Ensino de Ciências, bem como as diferenças observadas nas respostas produzidas após o refinamento dos prompts.

Nos casos estudados, as reformulações estiveram associadas a mudanças sobretudo na adequação às solicitações, na completude e na apresentação de elementos que favorecem a conferência das informações. Essas mudanças, entretanto, não significaram necessariamente maior precisão ou validade científica, uma vez que permaneceram inadequações e questões dependentes de avaliação humana.

A principal contribuição científica do estudo reside na comparação de tarefas distintas mediante dimensões analíticas comuns, evidenciando que a qualidade de uma devolutiva em relação às instruções recebidas não se confunde com as condições necessárias para sua utilização na pesquisa. Na revisão bibliográfica e na localização documental, a conferência envolve, entre outros aspectos, verificar a existência das referências e a correspondência entre afirmações e fontes. Na tradução e na interpretação histórica, exige também avaliar escolhas linguísticas, significados e contextualizações que não se resolvem pela simples localização de uma passagem. Na revisão da escrita acadêmica, acrescenta-se a necessidade de preservar o sentido e as decisões conceituais do autor. Desse modo, as diferentes tarefas não apenas apresentam resultados distintos, mas requerem formas específicas de validação e intervenção do pesquisador.

Dessa interpretação decorrem implicações práticas para o uso acadêmico da IAG. A elaboração de prompts pode favorecer a delimitação das tarefas, a explicitação de critérios e a organização das respostas, mas deve ser acompanhada de procedimentos de conferência compatíveis com o material analisado. Isso inclui verificar referências bibliográficas, distinguir transcrições literais de paráfrases, confrontar traduções e interpretações com as fontes correspondentes e avaliar criticamente alterações propostas à escrita científica. A utilidade de uma resposta depende, portanto, não apenas de sua apresentação, mas da possibilidade de examiná-la e decidir fundamentadamente sobre sua incorporação à pesquisa.

Para grupos de pesquisa e programas de pós-graduação, os resultados oferecem subsídios à discussão de práticas formativas voltadas ao uso crítico da IAG. Entre as possibilidades estão atividades de análise comparativa de respostas, formação para verificação de fontes e discussão de procedimentos de documentação das interações, incluindo o registro dos prompts, das devolutivas e das decisões tomadas pelos pesquisadores. Trata-se de implicações propostas a partir dos achados, e não de práticas institucionais cuja efetividade tenha sido avaliada neste estudo.

O estudo apresenta, contudo, limitações que devem circunscrever suas conclusões. Foram analisados cinco casos surgidos de necessidades concretas das práticas de pesquisa dos autores, sem pretensão de representar a diversidade de usos possíveis da IAG. Além disso, foi utilizada uma única plataforma, cada versão foi executada uma única vez e as reformulações modificaram simultaneamente diferentes componentes dos prompts, o que impede avaliar a estabilidade das respostas ou atribuir alterações observadas a elementos específicos das instruções.

Pesquisas futuras poderão ampliar esse desenho mediante múltiplas execuções de um mesmo prompt, comparação entre diferentes modelos e plataformas e investigação isolada de componentes como contexto, restrições, formato solicitado e critérios de rastreabilidade. Também se mostra pertinente aprofundar como diferentes tipos de tarefa exigem distintos procedimentos de validação e níveis de intervenção humana. Permanece central, nesse percurso, compreender em que condições a IAG pode apoiar a produção do conhecimento científico sem substituir a responsabilidade intelectual e metodológica dos pesquisadores.

.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, bem como do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Financiadora de Estudos e Projetos (FINEP), por meio de projetos financiados por editais públicos de Praças e Museus de Ciência (Planetário “Prof. Carlos Alfredo Argüello” e Observatório “Cortini-Vicentini”).

Referências

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Anjos, J. R. dos, Souza, M. G. de, Andrade Neto, A. S. de, & Souza, B. C. de. (2024). An analysis of the generative AI use as analyst in qualitative research in science education. *Revista Pesquisa Qualitativa*, *12*(30), 1-29. https://doi.org/10.33361/RPQ.2024.v.12.n.30.724

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Christou, P. A. (2023). How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? *The Qualitative Report*, *28*(7), 1968-1980. https://doi.org/10.46743/2160-3715/2023.6406

Flick, U. (2013). *Introdução à metodologia de pesquisa: Um guia para iniciantes*. Penso.

Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023). Exploring the use of AI in qualitative analysis: A comparative study of guaranteed income data. *International Journal of Qualitative Methods*, *22*, 1-13. https://doi.org/10.1177/16094069231201504

Kepler, J. (1619). *Harmonices Mundi*. https://archive.org/details/ioanniskepplerih00kepl

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1-35. https://doi.org/10.1145/3560815

Martins, R. de A. (2022). Como elaborar uma dissertação sobre história da ciência. In R. de A. Martins (Ed.), *Ensaios sobre história e filosofia das ciências II* (pp. 87–150). Quamcumque Editum. https://doi.org/10.5281/zenodo.6792439

Matthews, M. R. (1995). História, filosofia e ensino de ciências: A tendência atual de reaproximação. *Caderno Catarinense de Ensino de Física*, *12*(3), 164-214. https://periodicos.ufsc.br/index.php/fisica/article/view/7084

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*, 49-58. https://doi.org/10.1038/s41586-024-07146-0

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.

Naeem, M., Smith, T., & Thomas, L. (2025). Thematic analysis and artificial intelligence: A step-by-step process for using ChatGPT in thematic analysis. *International Journal of Qualitative Methods*, *24*, 1-18. https://doi.org/10.1177/16094069251333886

Neves, M. C. D. (2026). Ingá stone: A stellar map or a rain guage in the Brazilian prehistory? In *Convegno Nazionale della Società Italiana di Storia della Scienza: Trasformazioni e intersezioni*. https://www.societastoriadellascienza.it/images/convegni-siss/LAquila-2026/BoA_WEB.pdf

Pimentel, M., & Carvalho, F. (2025). *IA generativa e educação: Práticas e teorizações*. Sociedade Brasileira de Computação. https://doi.org/10.5753/sbc.18258.2

Rodrigues, G. da S., Brandão, V. R. de A., & Trivelato, R. M. da S. (2024). ChatGPT: uma ferramenta para a pesquisa científica? *Revista de Informação, Comunicação e Interfaces*, *2*(1), 74-93. https://doi.org/10.70493/cod31.v2i1.9916

Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sampaio, R. C., Chagas, V., Sanchez, C. S., Gonçalves, J., Borges, T., Alison, M. B., Tigrinho, C. S., Souza, J. R. de, & Schwarzer, F. (2024). Uma revisão de escopo assistida por inteligência artificial (IA) sobre usos emergentes de IA na pesquisa qualitativa e suas considerações éticas. *Revista Pesquisa Qualitativa*, *12*(30), 1-28. https://doi.org/10.33361/RPQ.2024.v.12.n.30.729

Santaella, L., & Kaufman, D. (2024). A Inteligência artificial generativa como quarta ferida narcísica do humano. *MATRIZes*, *18*(1), 37-53. https://doi.org/10.11606/issn.1982-8160.v18i1p37-53

Silva, M. J. F., & Paula, M. C. de. (2024). Perspectivas da inteligência artificial como ferramenta de apoio para análise textual discursiva. *Revista Pesquisa Qualitativa*, *12*(30), 1-26. https://doi.org/10.33361/RPQ.2024.v.12.n.30.727

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998-6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Sobre os autores

Gabriel Luiz Nalon Macedo

Licenciado em Física pela Universidade Estadual de Maringá (2020). Mestre em Educação para a Ciência e a Matemática pela Universidade Estadual de Maringá (2023). Doutorando no Programa de Pós-Graduação em Educação para a Ciência e a Matemática - PCM/UEM, na linha de pesquisa História, Epistemologia e Cultura da Ciência.

https://orcid.org/0000-0003-2054-8666

Luciano Carvalhais Gomes

Doutor em Educação para a Ciência e a Matemática pela Universidade Estadual de Maringá (2012). Atualmente, é professor do Departamento de Física e do Programa de Pós-Graduação em Educação para a Ciência e a Matemática (PCM) da Universidade Estadual de Maringá (UEM), atuando nas seguintes linhas de pesquisa: Ensino e Aprendizagem na Educação Científica; Formação de Professores de Ciências e Matemática; História, Epistemologia e Cultura da Ciência.

https://orcid.org/0000-0002-2005-9224

Marcos Cesar Danhoni Neves

Doutor em Educação pela Universidade Estadual de Campinas (1991). Professor Colaborador do Depto de Física-DFI/UEM EM 1989; Professor Efetivo do DFI/UEM desde 1990. Professor Titular concursado desde 2001. Atuou nos Programa de PG em Ensino de Ciências e Tecnologia (UTFPR/PG), e atua nos PPGs de Educação para a Ciência e a Matemática (PCM/UEM), de Ensino de Ciências e Educação Matemática (PPEGECEM/UEPG). Foi Professor-convidado do Programa Internacional de Master in Science (BUAP-Puebla-Mexico/INSA-Lyon-França/UTFPR-PG) para a cátedra Intercultural Education e pela Pedagogical Art University, St. Petersburg, Rússia.

https://orcid.org/0000-0002-3724-5373

Inteligência Artificial: Além dos usos da inteligência artificial generativa que constituem objeto desta pesquisa e estão descritos em sua metodologia, ferramentas de IAG foram utilizadas como apoio técnico à revisão linguística e gramatical do manuscrito. As sugestões geradas foram avaliadas criticamente pelos autores e incorporadas somente após sua revisão. A ferramenta não foi utilizada para substituir decisões teóricas, metodológicas ou interpretativas dos autores, que assumem integral responsabilidade pela autoria, pelo conteúdo e pela integridade científica do trabalho.