

Agentes de IA na Avaliação Científica: Uma revisão automatizada de publicações em Periódicos Brasileiros

Felipe Pereira Amaro¹, Ana Cristina Bicharra Garcia

¹Programa de Pós-Graduação em Informática
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

felipe.amaro@edu.unirio.br, cristina.bicharra@uniriotech.br

Abstract. *This paper investigates the use of multi-agent architectures based on Artificial Intelligence as a support mechanism for scientific evaluation in the context of Brazilian journals. The study is motivated by structural limitations of the peer-review process, such as reviewer overload, heterogeneity of reports, and lack of structured feedback for authors. To address these challenges, the research proposes the design of a multi-agent architecture capable of analyzing scientific articles, producing structured evaluations, and generating review reports to support authors, reviewers, and editors. The study follows the Design Science Research approach, including literature review, artifact development, and empirical evaluation through a case study with researchers. The expected outcome is to improve the consistency of scientific assessments, enhance manuscript quality, and contribute to strengthening the Brazilian editorial process.*

Resumo. *Este trabalho investiga o uso de arquiteturas multiagentes baseadas em Inteligência Artificial (IA) como instrumento de apoio à avaliação científica no contexto de periódicos brasileiros. Parte-se do reconhecimento de limitações estruturais do processo de revisão por pares, como sobrecarga de revisores, heterogeneidade dos pareceres e escassez de feedback estruturado aos autores. Para enfrentar esse cenário, propõe-se o projeto de uma arquitetura multiagente capaz de analisar artigos científicos, produzir avaliações estruturadas e gerar relatórios de revisão que apoiem autores, revisores e editores. A pesquisa segue a abordagem de Design Science Research (DSR), envolvendo revisão de literatura, desenvolvimento do artefato e avaliação empírica por meio de estudo de caso com pesquisadores. Espera-se que a solução contribua para ampliar a consistência das análises científicas, melhorar a qualidade dos manuscritos e fortalecer o processo editorial brasileiro.*

1. Introdução

O crescimento exponencial da produção científica nas últimas décadas tem intensificado desafios estruturais do sistema de revisão, como sobrecarga de revisores, ciclos editoriais prolongados e variabilidade na qualidade dos pareceres. Esses fatores causam um impacto na eficiência do processo científico e a formação de novos pesquisadores, que frequentemente enfrentam dificuldades para obter *feedback* estruturado e acionável sobre seus trabalhos. Como contraponto novos pesquisadores elevam a qualidade dos seus trabalhos após estas revisões.

Paralelamente, avanços recentes em IA, especialmente no desenvolvimento de modelos de linguagem de grande escala e sistemas baseados em agentes, têm ampliado as possibilidades de automação de tarefas cognitivas complexas. Tais tecnologias vêm sendo exploradas em diversas etapas do ciclo científico, incluindo análise de literatura, geração de hipóteses, apoio à escrita acadêmica e avaliação de manuscritos. Estudos recentes como [Jin et al. 2024] indicam que sistemas baseados em IA já são capazes de produzir avaliações científicas com níveis de concordância comparáveis aos de revisores humanos em determinados critérios, sugerindo potencial para apoiar o processo editorial e reduzir o tempo de iteração entre submissões.

Nesse contexto, surge a oportunidade de investigar arquiteturas de agentes de IA como instrumentos de apoio à revisão científica, não com o objetivo de substituir revisores humanos, mas de ampliar a qualidade, a consistência e a escalabilidade do processo avaliativo. Esta pesquisa busca contribuir para esse debate ao propor e investigar uma arquitetura multiagente voltada à análise automatizada de artigos científicos, com foco no contexto de periódicos brasileiros e na qualificação da produção científica nacional.

2. Contexto e Problema

O sistema de comunicação científica brasileiro apresenta características próprias que influenciam o processo de avaliação acadêmica. Periódicos nacionais frequentemente operam com recursos editoriais limitados, dependem de redes restritas de revisores e enfrentam dificuldades para acompanhar o crescimento do número de submissões. Como consequência, observa-se demora na tramitação dos manuscritos, heterogeneidade na qualidade dos pareceres e escassez de *feedback* estruturado aos autores.

Esses desafios impactam de maneira significativa pesquisadores em início de carreira, que dependem do processo de revisão como espaço de aprendizado e aprimoramento metodológico. Além disso, a limitação de ferramentas automatizadas de apoio à revisão científica reduz a capacidade de editores de realizar triagens eficientes e dificulta a identificação precoce de problemas metodológicos ou lacunas na contextualização dos trabalhos. Embora pesquisas recentes explorem o uso de modelos de linguagem na geração de *feedback* acadêmico e na análise automatizada de artigos, a maioria dessas iniciativas concentra-se em conferências internacionais ou em áreas específicas do conhecimento. Ainda há pouca investigação sobre como arquiteturas multiagentes podem ser projetadas e avaliadas para apoiar o processo editorial brasileiro, considerando suas especificidades organizacionais e sociotécnicas.

Dessa forma, identifica-se uma lacuna científica relacionada à compreensão de como agentes de IA podem ser estruturados para produzir avaliações científicas confiáveis, contextualizadas e úteis no ambiente acadêmico nacional. Assim, o problema de pesquisa deste trabalho consiste em investigar como arquiteturas multiagentes baseadas em IA podem apoiar a avaliação sistemática de artigos científicos e contribuir para a qualificação das publicações em periódicos brasileiros.

A questão de pesquisa que orienta o estudo é:

Agentes de IA podem produzir análises científicas estruturadas e úteis para autores, revisores e editores no contexto de periódicos brasileiros?

3. Metodologia

A pesquisa será conduzida com base na abordagem de DSR [Hevner et al. 2004, Peffers et al. 2007], que orienta o desenvolvimento e a avaliação de artefatos tecnológicos voltados à solução de problemas organizacionais e científicos.

Inicialmente, será realizada uma revisão da literatura sobre sistemas multiagentes e aplicações de IA no ciclo de produção do conhecimento, com o objetivo de fundamentar conceitualmente o artefato e identificar critérios relevantes para a avaliação de artigos acadêmicos. Em seguida, será projetada a arquitetura multiagente proposta, incluindo a definição dos papéis dos agentes, do fluxo de interação entre eles e dos critérios analíticos adotados. A partir desse modelo conceitual, será implementado um protótipo funcional capaz de analisar artigos científicos e gerar relatórios estruturados de revisão.

A avaliação empírica do artefato será conduzida por meio de um estudo de caso envolvendo pesquisadores da área de Sistemas de Informação, incluindo estudantes de pós-graduação, autores e revisores. O estudo adota uma abordagem mista (mixed methods), integrando métodos quantitativos e qualitativos de forma complementar. Essa escolha se justifica pela natureza do objeto de avaliação: enquanto aspectos como concordância entre avaliações humanas e automatizadas, tempo de análise e cobertura dos critérios científicos podem ser mensurados quantitativamente, a percepção de utilidade, clareza e confiabilidade dos relatórios gerados depende de interpretação qualitativa por parte dos participantes. Como premissa final todos os resultados serão validados por humanos para garantir a confiabilidade e assertividade.

4. Solução Proposta

A solução parte da premissa de que a revisão científica pode ser decomposta em dimensões analíticas distintas, cada uma tratada por agentes especializados. Essa estratégia é respaldada por evidências recentes: [Jin et al. 2024] demonstraram, no framework AgentReview, que a simulação do processo de revisão por pares com agentes baseados em LLMs permite investigar fatores que influenciam decisões editoriais, evidenciando a viabilidade de arquiteturas multiagentes para tarefas avaliativas em contextos acadêmicos. Complementarmente, [Zhu et al. 2025] evidenciaram, no framework DeepReview, que um pipeline de raciocínio estruturado em múltiplos estágios, com verificação de novidade, avaliação multidimensional e verificação de confiabilidade.

4.1. Visão Geral da Arquitetura

O fluxo arquitetural inicia-se com a ingestão do artigo submetido e sua conversão para uma representação estruturada em formato textual. Em seguida, agentes especializados analisam diferentes aspectos do manuscrito de forma independente, produzindo avaliações parciais. Essas avaliações são integradas por um agente orquestrador, responsável por consolidar os resultados, resolver contradições entre agentes e gerar um relatório final estruturado contendo síntese crítica do trabalho.

A arquitetura deste trabalho é composta por quatro agentes principais, conforme ilustrado na Figura 1.

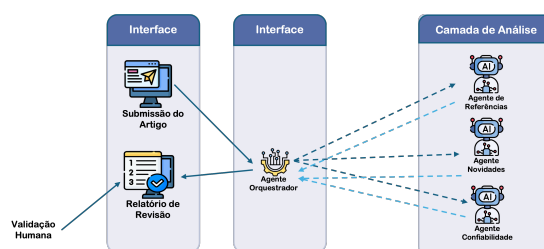


Figura 1. Arquitetura de Referência

4.2. Agentes Especializados

Agente de Verificação de Referências: Verifica a existência e a validade das referências bibliográficas do manuscrito, consultando bases como Semantic Scholar e CrossRef. Identifica referências não localizadas, inconsistências entre títulos e autores, e discrepâncias entre o conteúdo citado e as fontes reais. Essa funcionalidade se alinha ao estágio z do DeepReview [Zhu et al. 2025], que utiliza a API do Semantic Scholar para recuperar e validar trabalhos relacionados. Por envolver consultas repetitivas e estruturadas, este agente será executado em modelo local [Jiang et al. 2023], reduzindo custos operacionais.

Agente de Verificação de Novidade: Avalia a originalidade da contribuição do manuscrito em relação à literatura existente, operando em três etapas inspiradas no estágio z do DeepReview [Zhu et al. 2025]: (i) extração das alegações de contribuição; (ii) recuperação e ranqueamento de trabalhos relacionados; e (iii) análise comparativa com o estado da arte.

Agente de Verificação de Confiabilidade: Examina a coerência interna e o rigor metodológico do manuscrito por meio de uma cadeia de evidências inspirada no estágio z do DeepReview [Zhu et al. 2025], percorrendo três dimensões: verificação metodológica, verificação experimental e análise abrangente. Para cada fragilidade identificada, o agente coleta evidências textuais do manuscrito e atribui um nível de confiança.

4.3. Agente Orquestrador

O agente orquestrador é o componente central da arquitetura, responsável por coordenar a execução dos demais agentes, consolidar as avaliações parciais e produzir o relatório final de revisão.

4.4. Estratégia de Uso de Modelos

A arquitetura adota uma estratégia híbrida de uso de LLMs, combinando chamadas a APIs comerciais para análises que exigem raciocínio complexo e modelos executados em ambiente local para tarefas procedimentais e repetitivas. Essa abordagem visa equilibrar qualidade das análises e viabilidade econômica, considerando que a avaliação de um único manuscrito pode demandar dezenas de milhares de tokens.

4.5. Saída do Sistema

Ao final do processamento, o artefato produz um relatório de revisão estruturado para apoio à avaliação acadêmica, contendo: 1. Resumo do trabalho avaliado, 2. Análise de novidade com posicionamento frente à literatura, 3. Verificação de confiabilidade com

evidências textuais, 4. Validação das referências bibliográficas, 5. Identificação de pontos fortes e fracos, 6. Recomendações acionáveis de melhoria.

5. Trabalhos Relacionados

O objetivo deste trabalho é investigar o papel de agentes baseados em LLMs no processo de revisão por pares. O *framework* AgentReview utiliza agentes baseados em LLMs para simular interações entre autores, revisores e editores, permitindo investigar fatores sociológicos e organizacionais que influenciam decisões editoriais.

[Jin et al. 2024]. Há também esforços voltados especificamente à geração automática de pareceres científicos. O método MARG (Multi-Agent Review Generation) propõe o uso de múltiplos agentes especializados que analisam diferentes aspectos do artigo, como experimentos, clareza textual e impacto científico. [D'Arcy et al. 2024].

Avanços recentes têm buscado aprofundar a qualidade das revisões automatizadas por meio de processos de raciocínio estruturado. O *framework* DeepReview propõe uma abordagem em múltiplos estágios que emula o processo de revisores especialistas,

onde os autores enfatizam que o sistema não pretende substituir revisores humanos, mas explorar o potencial de assistência de LLMs no processo de revisão por pares [Zhu et al. 2025].

Apesar dos avanços na capacidade dos modelos de linguagem, estudos empíricos têm revelado vulnerabilidades importantes em sistemas baseados em LLMs. [Li et al. 2025] propõem um *framework* de perturbação multinível guiado por aspectos para diagnosticar sistematicamente as fraquezas de LLMs na revisão automatizada.

Em conjunto, esses estudos sugerem que agentes de IA podem desempenhar papel relevante no suporte à produção científica, mas evidenciam também limitações significativas que precisam ser consideradas no projeto de arquiteturas multiagentes. Ainda há lacunas quanto à adaptação dessas tecnologias a contextos editoriais específicos, especialmente fora do eixo das grandes conferências internacionais, como é o caso dos periódicos brasileiros que constituem o foco desta pesquisa.

6. Resultados Esperados

Como resultado este trabalho visa explorar o potencial de arquiteturas multiagentes de IA para apoiar o processo de revisão científica e contribuir para a qualificação de artigos acadêmicos.

Espera-se que o artefato produza análises consideradas úteis pelos usuários, capazes de identificar lacunas metodológicas, inconsistências argumentativas e oportunidades de melhoria nos manuscritos. Também se espera observar convergência parcial entre avaliações humanas e automatizadas, indicando viabilidade do uso do sistema como instrumento de apoio ao processo científico. As métricas de avaliação incluirão percepção de utilidade do sistema, clareza dos relatórios gerados, cobertura dos aspectos científicos analisados, tempo necessário para a análise e grau de concordância entre avaliações humanas e automatizadas. Esses indicadores permitirão investigar a eficácia, aplicabilidade e limitações da solução proposta no contexto acadêmico.

Espera-se, por fim, que os resultados contribuam para ampliar o conhecimento

sobre o papel de arquiteturas de agentes de IA no processo científico, oferecendo subsídios para futuras pesquisas e aplicações no sistema de comunicação científica brasileiro.

Referências

- D'Arcy, M., Hope, T., Birnbaum, L., and Downey, D. (2024). Marg: Multi-agent review generation for scientific papers.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Q.*, 28(1):75–105.
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. (2023). Mistral 7b. *arXiv preprint arXiv:2310.06825*.
- Jin, Y., Zhao, Q., Wang, Y., Chen, H., Zhu, K., Xiao, Y., and Wang, J. (2024). AgentReview: Exploring peer review dynamics with LLM agents. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1208–1226, Miami, Florida, USA. Association for Computational Linguistics.
- Li, J., Li, Y., Hu, X., Gao, M., and Wan, X. (2025). Where do llms go wrong? diagnosing automated peer review via aspect-guided multi-level perturbation. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, page 1572–1581, New York, NY, USA. Association for Computing Machinery.
- Peppers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. (2007). A design science research methodology for information systems research. *J. Manage. Inf. Syst.*, 24(3):45–77.
- Zhu, M., Weng, Y., Yang, L., and Zhang, Y. (2025). DeepReview: Improving LLM-based paper review with human-like deep thinking process. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T., editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 29330–29355, Vienna, Austria. Association for Computational Linguistics.