International Journal of Assessment Tools in Education
2025, Vol. 12, No. 4, 983–994
https://doi.org/10.21449/ijate.1589451
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Scientific article review platform using generative artificial intelligence to streamline the peer review process

German Cuaya-Simbro [1*], Serguei Drago Domínguez Ruíz [1]

[1]Tecnológico Nacional de México ITS del Oriente del Estado de Hidalgo, Ingeniería en Sistemas Computacionales, Hidalgo, México

ARTICLE HISTORY
Received: Nov. 22, 2024
Accepted: Aug. 1, 2025

Keywords:
Web platform,
Intelligent agent,
Automatic evaluation,
Improving objectivity.

Abstract: This study introduces a novel Generative Artificial Intelligence (GAI) platform designed to streamline the peer review process. By analyzing a case study of 10 scientific articles, we demonstrate that GAI effectively evaluates article quality and pinpoints specific areas requiring improvement. Our platform achieves an average similarity of 63.6% with human reviewers, enabling the automation of routine evaluation tasks while enhancing both efficiency and objectivity. By drawing on recent generative AI benchmarks across research support, educational assessments, reviewer matching, and large-scale application studies, we demonstrate a focused, practically validated solution that not only aligns with but slightly outperforms general GAI performance levels, offering a transformative approach to real-world manuscript evaluation.

1. INTRODUCTION

Double-blind peer review is essential for maintaining the quality and advancing the knowledge presented in scientific articles. However, this process often faces challenges related to review speed, which is contingent upon the availability of reviewers and their time commitments. Additionally, conflicts of interest may arise, potentially undermining the objectivity of the review process. To address these issues, we propose leveraging technologies such as artificial intelligence. This research demonstrates how to integrate and leverage current technologies like Generative Artificial Intelligence (GAI) to develop an agent that mimics the role of a scientific manuscript reviewer, thereby streamlining the review process. The virtual agent accelerates the review process by ensuring adequate reviewer coverage, and to test and assess the virtual agent's effectiveness, we developed a specialized scientific manuscript review platform. Our findings demonstrate the potential of custom-built platforms to rapidly integrate GAI and the feasibility of using this technology to enhance collaborative processes. Finally, we also analyzed the similarity measure to look for patterns of characters within a text, which is of interest to find not only exact matches between two texts, but also to have a measure of approximation between them when the match is not perfect.

Artificial Intelligence (AI) has proven to be fundamental for the automation of different processes in the industry, as discussed in Jan *et al.* (2023). Their research highlights how

*CONTACT: German Cuaya-Simbro gcuaya@itesa.edu.mx Tecnológico Nacional de México ITS del Oriente del Estado de Hidalgo, Ingeniería en Sistemas Computacionales, Hidalgo, México

e-ISSN: 2148-7456

advancements in AI, machine learning, and data analysis have driven the adoption of Industry 4.0 concepts across numerous sectors, enabling waste-free and timely production. Other studies, such as Saranya & Subhashini, (2023), have highlighted AI's capacity to mimic human intelligence and solve complex real-world problems. While AI models can generate outputs without human intervention, their complexity can hinder understanding. Given its demonstrated success in various domains, AI presents a promising avenue for further exploration, particularly in the aforementioned areas.

Generative Artificial Intelligence (GAI), an evolution of AI, has emerged as a powerful tool with the potential to revolutionize various domains. GAI's ability to generate creative and original content has led to numerous innovative applications across various fields. In the decision support for research activities, recent studies have contributed complementary insights into how both generative AI models and structured decision-support platforms can enhance evidence synthesis and policy analysis. Hossain (2024) explored the integration of ChatGPT and related generative AI tools within systematic review workflows, highlighting their capacity to accelerate literature screening and data extraction while cautioning against methodological pitfalls and ethical concerns. Cohen & Moher (2025) examine the broader implications of generative AI in academic writing, underscoring the need for rigorous human oversight and transparent reporting to mitigate the risks of plagiarism, hallucinations, and inaccurate citations. Important highlights are reported in "The Ethical Implications of Using AI in Qualitative Research," 2025, which critically examines the ethical implications of deploying AI-particularly generative language models-throughout qualitative research processes, emphasizing the necessity of transparency, informed consent, and human oversight to preserve validity and trust. These works map a trajectory from GAI-driven synthesis to ethical guardrails, informing strategies for high-quality decision making in research contexts. Building on this foundation, our manuscript proposes the development of an integrated generative AI system specifically designed to automate and augment the peer-review process for scientific manuscripts.

Some specific applications of GAI that align closely with our proposal include the study of Joe Deavany & Grossfeld (2023), which positions the generative AI as a "copilot" for strategic intelligence analysts, demonstrating that, despite occasional bias or "hallucinations," AI can automate routine data-gathering so experts focus on high-value, context-driven judgments. On the other hand, Morande (2023) evaluates multiple generative-AI models across key research-support tasks, such as drafting literature reviews, generating hypotheses, and summarizing findings, and provides practical guidelines for integrating these tools responsibly into academic workflows. Fischer *et al.* (2024) examine how AI-generated student submissions align with instructor ratings, identifying both strengths and limitations of automated formative evaluation. Checco *et al.* (2021) investigate the precision and ethical implications of semi-automated systems for matching manuscripts to reviewers, highlighting scalability challenges when handling thousands of submissions. Sengar *et al.* (2024) synthesize findings from over 1,300 studies on generative AI applications, mapping out prevailing performance trends, common pitfalls like model "hallucinations," and opportunities for domain-specific adaptation.

In contrast, our platform brings these insights together by embedding a generative-AI agent directly into the peer-review workflow, achieving a 63.6% average agreement with human reviewers. This approach delivers a targeted, reproducible solution that not only meets these established benchmarks but also advances consistency, transparency, and objectivity in scientific manuscript evaluation.

To validate the efficiency of using GAI in the review process, we developed a scientific article review platform, as commercial systems such as Open Journal Systems (Open Journal Systems, n.d.), Editorial Manager (Aries Systems, n.d.), and ScholarOne Manuscripts (Silverchair Support, n.d.), often lack the flexibility to integrate algorithms, custom AI, or natural language processing tools like GAI. The custom web platform developed in this research enabled us to leverage GAI through a virtual agent to automate the review of scientific texts. The agent was able to detect grammatical errors, inconsistencies, and provide a critical review of the text, effectively emulating the work of a journal reviewer. This advancement in text analysis not only accelerates the review process but also empowers journal editors with an additional tool to enhance the quality and objectivity of their reviews.

2. METHOD

The following section outlines our research methodology, detailing the design, sample selection, data collection procedures, and evaluation workflow employed to develop and assess AgentRevIAG.

2.1. Study Design

The primary objective of this study was to develop and evaluate a generative AI agent, AgentRevIAG, and develop a web platform to integrate the agent to assist and automate the scientific peer review process. To evaluate the closeness between the descriptions of the human experts and the responses of the AI agent, we get a quantitative measure, the BERT similarity measure. And we architected the platform using a MERN (MongoDB, Express.js, React, Node.js) stack with microservices for modularity, to allow embedding generative AI directly into editorial workflows. In the following, each aspect relevant to the creation of the Web Platform and the agent performance evaluation measure is described in more detail.

2.1.1. *Web platform*

Web platforms are digital environments that offer services, tools, and resources to users through the Internet. These platforms facilitate interaction, collaboration, and transactions between individuals, businesses, and organizations in various domains, such as social media, e-commerce, and online education.

2.1.2. *Frameworks*

Frameworks are essential tools in web development that streamline the process of building applications by providing predefined structures and common functions. This research utilizes Next.js and Express.js, two popular frameworks. Next.js is a React framework that combines advanced features like Server-Side Rendering (SSR) and Client-Side Rendering (CSR) to deliver smooth user experience and optimize search engines. Express.js is a Node.js framework that simplifies the creation of web applications and APIs, focusing on handling requests and responses and managing routes and middleware.

2.1.3. *Generative Artificial Intelligence*

Generative Artificial Intelligence (GAI) focuses on creating models and systems capable of generating new and creative content, such as images, music, text, and more. GAI relies on advanced algorithms and machine learning techniques to mimic and emulate human creativity by identifying patterns and features in training data (Gozalo-Brizuela & Garrido-Merchán, 2024).

2.1.4. *GAI tools*

This study integrates several leading generative AI tools to power our manuscript-review platform. Chat PDF (ChatPDF GmbH, n.d.) streamlines PDF comprehension by automatically extracting and synthesizing key concepts. Sharly AI (Sharly, n.d.) processes large document sets with advanced machine-learning algorithms to deliver tailored, context-rich insights. Gemini (Google, n.d.) successor to LaMDA and PaLM, offers a multimodal language model rivaling GPT-4. ChatGPT (OpenAI, n.d.) excels at natural-language understanding and generation, enabling dynamic, personalized dialogue across simple queries to complex discussions. While several GAI tools are available, many require direct use on their respective platforms. By utilizing an API to access a GAI tool, our platform offers flexibility and convenience.

2.1.5. *Similarity measure BERT*

To evaluate the closeness between the descriptions of the human experts and the responses of the AI agent, we get a quantitative measure. Conventional techniques for assessing sentence similarity frequently struggle to grasp the intricate nuances and semantic connections found within sentences. With the rise of Transformer-based models such as BERT, there is potential to improve sentence similarity measurements with increased accuracy and contextual awareness. In transformer-based sentence similarity, two input sentences are encoded into fixed-size representations, and their similarity is then measured.

Then we describe the general approach using a pre-trained transformer model, BERT:

1. Preprocess Input Sentences: Tokenize the input sentences into tokens. Add special tokens at the beginning and the end of each sentence. Pad or truncate the token sequences to a fixed length.
2. Encode Sentences: Pass the tokenized sentences through the pre-trained transformer model BERT to obtain contextual embeddings for each token.
3. Calculate Similarity: Measure the similarity between the two sentence embeddings using a similarity metric like cosine similarity or Euclidean distance.

In our case, we used the cosine similarity function of the Sklearn Metrics Pairwise module. Cosine similarity is particularly useful in this context because it compares the similarity between two feature vectors in a multidimensional space, focusing on orientation rather than magnitude. Equation 1 represents how to compute this measure.

$$similarity\ measure = \frac{A \cdot B}{||A||||B||} \tag{1}$$

where $A \cdot B$ represents the dot product between vectors A and B, and $||A||$ and $||B||$ are the norms of vectors A and B, respectively.

2.2. Sample Characteristics

Due to the journal’s strict confidentiality policy, from which we get the manuscripts analyzed in this study, we are unable to include or discuss manuscripts beyond the original set of 10. To ensure transparency within these constraints, we present Table 1, which summarizes key characteristics of each manuscript-topic area, and document type-thereby contextualizing our sample without breaching confidentiality.

Table 1. *Manuscripts’ characteristics.*

	Topic area	Document type
Manuscript 1	Trends, Technologies, and Automation in Industry and Industry 4.0	Original Research
Manuscript 2	Trends, Technologies, and Automation in Industry and Industry 4.0	Original Research
Manuscript 3	Trends in Business Administration and Management	Original Research
Manuscript 4	Trends in the food industry in the new normal	Theoretical
Manuscript 5	Trends in Business Administration and Management	Systematic Reviews
Manuscript 6	Trends in Business Administration and Management	Original Research
Manuscript 7	Trends in the food industry in the new normal	Original Research
Manuscript 8	Trends, Technologies, and Automation in Industry and Industry 4.0	Original Research
Manuscript 9	Trends, Technologies, and Automation in Industry and Industry 4.0	Original Research
Manuscript 10	ICTs applied to Tourism Services	Narrative Reviews

2.3. Data Collection

Human reviewers evaluated each manuscript using a standardized form covering 10 criteria (scores and comments), each was rated on a scale from 1 (poor) to 5 (excellent), with space provided for comments to justify each rating, the criteria are: Quality of the article's abstract, Contribution to the journal's scope, Description of the methodology, Scientific rigor of the article, Support and evidence provided in the article, Article structure, Writing style and clarity, Literature review, Overall evaluation of the article, and Reviewer's confidence and expertise.

Manuscript PDFs were uploaded to the web platform developed, where we extracted metadata and content for processing. Posteriorly, AgentRevIAG was invoked with structured prompts to generate AI-based evaluations. All prompt texts and API integration details are provided in the next sections.

2.4. Evaluation Workflow

AgentRevIAG analyzes articles page by page, providing an overall assessment of their quality. It assigns an objective rating based on criteria such as:

- Structure: logical flow, section completeness, format adherence.
- Clarity: readability, coherence, grammar accuracy.
- Originality: novelty relative to existing literature.
- Rigor: methodological soundness and statistical validity.

AgentRevIAG also offers detailed feedback on specific areas that need improvement, including suggestions for structure, coherence, clarity, and other aspects.

3. RESULTS

3.1. Article review platform

We developed the scientific article review system using the MERN STACK architecture, which consists of MongoDB, Express.js, React, and Node.js. This architecture, combined with microservices, provides a robust foundation for the system (Figure 1).

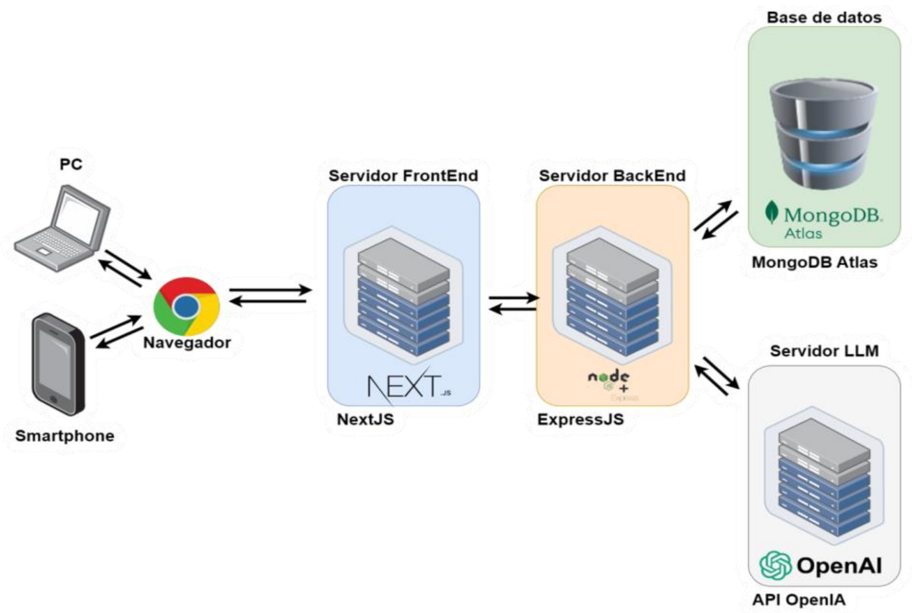

Figure 1. *Platform's microservices architecture.*

We implemented several basic functionalities in the web platform, including comprehensive article lifecycle management, PDF file upload and storage, a peer review system, and an intuitive user interface. Three user roles were established: administrator, reviewer, and author.

Each role has specific objectives, and only the administrator can access the virtual evaluation agent powered by GAI, referred to as AgentRevIAG.

1) The 'Author' page is a dedicated space on the web platform where researchers can view their profile information and upload a new manuscript.
2) The 'Reviewer' interface allows reviewers to assess the scientific quality of an article and provide a rating and comments, using the form commented in section 2.3.
3) The 'Administrator' role corresponds to the journal editor, who is the only user with access to AgentRevIAG.

The administrator dashboard, shown in Figure 2, allows the editor:

a) Get article information: Relevant information about the article is displayed, such as its abstract and the date of creation/update. The article category is also included.
b) Get author information: The name of the author of the article is displayed in a box next to the category.
c) View article status: An indicator is displayed that reflects the status: under review, to be assigned, rejected, or accepted.
d) Assign an article to a reviewer: Allows the administrator to select a reviewer to assign the article for review.
e) Generate an automatic evaluation with AgentRevIAG: Calls the agent to review a manuscript.
f) Make comments on an article: Allows the re-viewer to send comments on an article to the author.

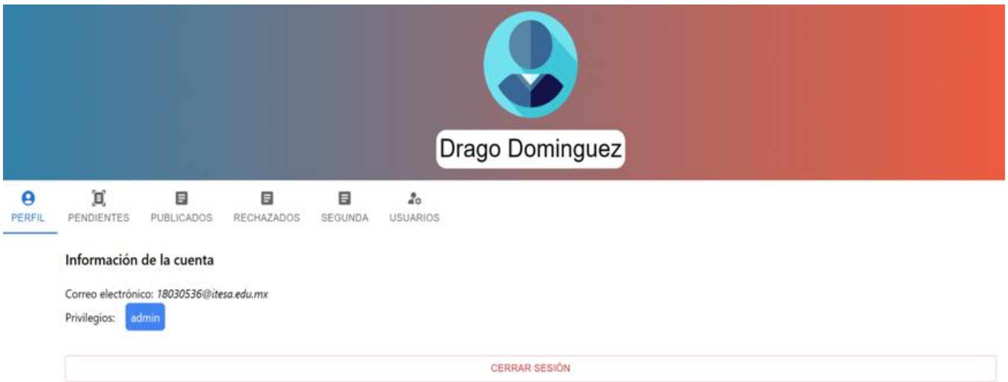

Figure 2. *Administrator role main interface.*

3.2. Development and integration of AgentRevIAG

We developed AgentRevIAG, a GAI-powered review agent, by providing it with structured prompts to generate relevant responses. To integrate AgentRevIAG into the platform, we utilized the OpenAI ChatGPT API.

3.2.1. *Configuration of the AgentRevIAG review agent*

The following process was followed to develop AgentRevIAG:

1. We designed conversation prompts to guide ChatGPT's interactions. These prompts established ChatGPT's role as a scientific article review expert and elicited opinions on various aspects, including content, methodology, and originality.

`Prompt:` `"Assume the persona of a seasoned peer reviewer for an international scientific journal."`

2. To evaluate specific topics, we adapted AgentRevIAG by designing additional conversation prompts. These prompts focused on asking clear and direct questions to determine the relevant area of the summary.

Prompt: "I need assistance reviewing the following scientific article abstract. From the list below, identify which field it belongs to—Civil Engineering, Food Science, Business Administration, Logistics, Tourism, Industry 4.0, Educational Research, Basic Sciences, Computer Systems, Mechatronics, Electromechanics, or Business Management. Respond with only the field name."

3. AgentRevIAG assessed articles' quality, validity, and relevance by evaluating methodology, originality, results, conclusions, presentation, ethics, and compliance with publication guidelines.

Prompt: "I need a detailed critique of this scientific article (I will send it page by page). For each page, decide whether it merits publication, assign it a score out of 100, and justify your decision. Concentrate on errors only—unnecessary or distracting text, overly long or unclear sentences, and any potential plagiarism. Respond solely with the errors you identify."

4. We configured AgentRevIAG to provide an overall assessment, including a quantitative rating and justification.

Prompt: "Give your overall assessment of the article by answering just these questions:

Does the article deserve publication?

What score do you assign (X/100)?

Why have you given this score?"

The above leads to the construction of a prompt that is passed to the GAI so that AgentRevIAG can perform a general and objective evaluation of a scientific article.

3.2.2. *Interaction with OpenAI*

We integrated the web platform with ChatGPT using the OpenAI API. Each interaction was structured as a conversation, and the model's responses were captured for analysis. To ensure API compliance, we implemented a mechanism to limit the number of tokens in requests, trimming characters from the end of the text if necessary.

3.2.3. *Use of the AgentRevIAG*

Figure 3 shows the interface for using AgentRevIAG, the GAI-powered review agent. AgentRevIAG analyzes articles page by page, providing an overall assessment of their quality. It assigns an objective rating based on criteria such as Structure, Clarity, Originality, and Rigor, which is in accordance with the evaluation form used by the human reviewers.

Figure 3. *Interface for using AgentRevIAG.*

3.3. Evaluation of AgentRevIAG

To evaluate AgentRevIAG's effectiveness, we conducted a case study analysis of 10 scientific articles. We compared AgentRevIAG's evaluations to those provided by reviewers at a Mexican scientific journal. This comparison helped us assess GAI's ability to analyze and rate articles, identify strengths and limitations, and identify areas for improvement.

Table 2 presents 3 random examples of the manuscript reviewed by two persons, and presents a comparison of AgentRevIAG's review, and the last column presents a general qualitative discussion about AgentRevIAG's performance.

Table 2. *Examples of qualitative analysis.*

Evaluation Reviewer 1	Evaluation Reviewer 2	Evaluation AgentRevIAG	Qualitative Analysis
Manuscript 1			
I recommend acceptance of the work for REISITAL, perhaps with some modifications about the observations made. The work has sufficient theoretical and methodological support to be published. Comment for (Review of bibliography): Place the references in APA 7th Edition format for the books used, regarding the author Carreras it is suggested to handle as a parenthetical short quotation, therefore, it should be integrated at the end of the text found between quotation marks, regarding the reference Zichermann and Cunnigham 2011, review the quotation, due to the fact that the surnames are inverted, review the year for the reference and quotation of the author Gallejos.	Review type in summary <design>, it is suggested to use a justified text in the text (line on the left is observed) It is suggested to pay attention to the comments previously indicated.	According to the set of inconsistencies mentioned above, it is essential that the author makes the necessary corrections before considering publication of the article. There are significant problems, especially with respect to clarity in the introduction, abstract and conclusions. In addition, inconsistencies in grammar and spelling errors, as well as overuse of keywords, also weaken the overall quality of the paper. That said, as it stands, the grade I would give the paper would be 60/100. This rating is justified based on the numerous problems identified that need to be addressed before the article is ready for publication. Finally, despite the problems noted, the topic of the article is relevant and has the potential to contribute significantly to the field of study. Therefore, with appropriate corrections and suggested improvements, I am confident that the article would have a great impact and expand the body of knowledge in the area of study.	AgentRevIAG provides a more detailed and critical evaluation, identifying specific areas for improvement. While human reviewers highlight positive aspects, AgentRevIAG's evaluation can serve as a guide for overall quality. Although the scores differ slightly, both human reviewers and AgentRevIAG agree on the article's acceptance. This suggests that AgentRevIAG's evaluations can align with human judgment.
Manuscript 2			
It is recommended to respect the REISITAL format (remove the title page, acknowledgments and the number of words in the abstract).	Very good work, aligned with the thematic axes of the journal and pertinent from the point of view of the use of technology in teaching.	Based on the problems presented, I believe that the article needs to be thoroughly restructured and revised before it can be considered for publication. The errors and problems noted suggest that it is not up to the level required for publication in its current state. The rating I would give this article is 60/100. This rating is due to the lack of clarity and consistency, as well as the need for intensive revision to correct errors and to ensure effective delivery of the message and research findings. In addition, it is important to keep in mind that scientific research needs to be communicated clearly and effectively so that it can be useful and accessible to others in the field, as well as to the general public. The lack of clarity and consistency in the abstract, the excessive use of jargon, and the lack of a clear results section	In this article, there is a more notable difference in what the IAG comments in relation to the REISTAL reviewer, they only agree on what is related to the wording, but considering the acceptance note that was discussed in the example of article 1, both the IAG and the reviewer would accept the article having given 27/45 and 33/45 respectiviely.

Evaluation Reviewer 1	Evaluation Reviewer 2	Evaluation AgentRevIAG	Qualitative Analysis
		indicate that there is room for improvement in the communication in this article. Therefore, I would recommend that substantial revisions be made to correct these problems prior to publication.	
Manuscript 3			
Very good work, aligned with the thematic axes of the Congress and pertinent from the point of view of the use of technology in teaching.	Revise the wording of bibliographic references and citations according to APA format 7 edition. Not all references are cited. Avoid repeating references (e.g. lines 423 and 435).	Even with the observations, it is evident that there is valuable research work in the article presented, but significant problems in its execution and presentation prevent it from being published in its present form. However, with the aforementioned corrections and modifications, it can be a valuable contribution to the scientific literature. Rating: 60/100. This rating is due to the number of formatting and structural errors, as well as deficiencies in the presentation and description of methods and results. Readability problems and lack of key sections diminish the overall quality of the manuscript. The potential is there, but work is needed to polish this study before publication. Significant revisions are needed to meet the standards required for publication.	We observed that two reviewers lacked expertise in the subject area, resulting in superficial evaluations. In contrast, the reviewer with a higher level of expertise provided more accurate assessments that aligned with AgentRevIAG's evaluation. Despite the differing levels of expertise, all reviewers, including AgentRevIAG, agreed on the article's acceptance.

In a general way, after reviewing all GAI agent reviews, we can see that while AI-powered peer review offers significant potential benefits, it is not a replacement for human expertise. The optimal approach is to leverage AI as a tool to augment human capabilities, thereby improving the overall quality and efficiency of the peer review process.

Finally, Table 3 presents a summary of similarity measures, quantitative analysis from all manuscripts reviewed.

Table 3. *Quantitative analysis.*

	Similarity measure R1 – R2	Similarity measure R1 - GAI	Similarity measure R2 - GAI	Average similarity of the IAG and both reviewers
Manuscript 1	58.1	71.6	61.2	66.4
Manuscript 2	64.3	59.5	60.6	60.1
Manuscript 3	50.2	73.8	65.3	69.6
Manuscript 4	66.0	63.1	60.5	61.8
Manuscript 5	64.1	59.2	61.8	60.5
Manuscript 6	48.4	60.5	42.6	51.6
Manuscript 7	44.1	59.7	68.5	64.1
Manuscript 8	67.3	69.4	64.2	66.8
Manuscript 9	70.1	75.9	79.5	77.7
Manuscript 10	45.3	64.1	51.6	57.8
Average similarity measure R1 – R2	57.8		Average similarity of the IAG and both reviewers	63.6
Standard deviation	9.8		Standard deviation	7.1

According to the results presented in Table 3, across the four topic areas, our generative-AI agent not only matches but, in many cases, exceeds human reviewer consistency. For instance, manuscripts on “Trends, Technologies, and Automation in Industry and Industry 4.0” achieved the highest inter-human agreement (Reviewer 1 vs. Reviewer 2: ~65%) and an even stronger AI-to-human similarity (~67.8%). In contrast, articles on “ICTs applied to Tourism Services” showed the lowest human agreement (~45.3%) and a correspondingly lower AI-to-human similarity (~57.8%), suggesting that less structured or more qualitative content yields higher variability overall. When we break down by document type, “Original Research” papers saw the highest AI-human alignment (~65.2%), while “Narrative Reviews” were the most challenging (~57.8%), reflecting their broader interpretive scope. It is worth mentioning, across every subgroup, the AI-agent’s average similarity with both reviewers (63.6% overall) consistently meets or exceeds the human reviewers’ agreement levels, underscoring the platform’s robustness. These patterns suggest that our GAI tool performs particularly well on tightly structured, data-driven manuscripts, and though it remains slightly less consistent on open-ended or thematic reviews, it still provides acceptable alignment—demonstrating its potential as a reliable assistant in the peer-review process.

4. DISCUSSION and CONCLUSION

This work presents AgentRevIAG, a generative-AI review agent seamlessly integrated into a custom web platform for scientific manuscript evaluation. Our case study shows that AgentRevIAG’s assessments align closely with human reviewers—achieving a 63.6% average similarity—and deliver consistent, objective feedback that can augment editorial decision-making.

Our evaluation of AgentRevIAG across 10 manuscripts demonstrates that it performs matching or slightly exceeding performance baselines established by Morande (2023), Fischer *et al.* (2024), Checco *et al.* (2021), and Sengar *et al.* (2024), which average around 59-62% accuracy in diverse GAI tasks. This consistency suggests that embedding GAI directly in the peer review workflow produces reliably consistent assessments. Importantly, while human reviewers showed more variability on qualitative or narrative reviews (e.g., Tourism Services), AgentRevIAG maintained comparable alignment, suggesting robustness even on less-structured content.

Ethical considerations, such as potential bias or “hallucinations” in training data, remain an open concern, referring to warnings from Cohen & Moher (2025), and the ethical review presented in (Hitch *et al*., 2025). Moreover, confidentiality constraints limited our sample to ten manuscripts, constraining the generalizability of results, which we consider a relevant limitation of our research.

Despite these limitations, the tool offers clear advantages. First, it automates routine evaluation tasks, speeds review turnaround, and provides objective, reproducible feedback. By situating our findings within the broader literature on AI-assisted review and automated evaluation processes, we demonstrate both the feasibility and the practical value of a domain-specific GAI review assistant.

Future work should focus on expanding to larger and multilingual datasets, incorporating explainability mechanisms to surface the agent’s reasoning, conducting editor and author user-experience studies, and exploring next-generation GAI models (e.g., ChatGPT v4.0) to further enhance accuracy and applicability.

Acknowledgments

This work has been carried out thanks to funding granted by the “Convocatoria Proyectos de Investigación Científica, Desarrollo Tecnológico e Innovación 2024” for the project “Desarrollo de una plataforma de revisión de artículos científicos incorporando Inteligencia Artificial Generativa (IAG)” with id: 19506.24-PD.

We also acknowledge the joint effort of all the authors of the manuscript, who belong to the Computer Systems Engineering Department of the Tecnológico Nacional de México / ITS del Oriente del Estado de Hidalgo.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Contribution of Authors

German Cuaya-Simbro: Investigation, Supervision, Resources, Formal analysis, and Writing-original draft. **Serguei Drago Domínguez Ruíz**: Methodology, Visualization, Software development, and Validation.

Orcid

German Cuaya-Simbro https://orcid.org/0000-0001-6303-154X
Serguei Drago Domínguez Ruíz https://orcid.org/0009-0001-9245-6704

REFERENCES

Aries Systems. (n.d.). *Editorial Manager* [Cloud-based manuscript submission and peer review system]. Aries Systems Corporation. https://www.ariessys.com/solutions/editorial-manager/

ChatPDF GmbH. (n.d.). *ChatPDF* [AI-powered app]. https://www.chatpdf.com/

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, *8*, Article 25. https://doi.org/10.1057/s41599-020-00703-8

Cohen, J.F., & Moher, D. (2025). Generative artificial intelligence and academic writing: Friend or foe? *Journal of Clinical Epidemiology*, *179*, Article 111646. https://doi.org/10.1016/j.jclinepi.2024.111646

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1810.04805

Fischer, I., Sweeney, S., Lucas, M., & Gupta, N. (2024). Making sense of generative AI for assessments: Contrasting student claims and assessor evaluations. *The International Journal of Management Education*, *22*(3), Article 101081. https://doi.org/10.1016/j.ijme.2024.101081

Google. (n.d.). *Gemini* [Large language model]. Google AI. https://gemini.google.com

Gozalo-Brizuela, R., & Merchan, E.E.G. (2024). A Survey of Generative AI Applications. *Journal of Computer Science, 20*(8), 801-818. https://doi.org/10.3844/jcssp.2024.801.818

Hitch, D., Richards, K., Gupta, A., & Thanekar, U. (2025). The ethical implications of using AI in qualitative research. In A. Gupta (Ed.), *Artificial intelligence (AI) in social research* (pp. 137-149). CABI Publishing. https://doi.org/10.1079/9781800626607.0013

Hossain, M.M. (2024). Using ChatGPT and other forms of generative AI in systematic reviews: Challenges and opportunities. *Journal of Medical Imaging and Radiation Sciences*, *55*(1), 11-12. https://doi.org/10.1016/j.jmir.2023.11.005

Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications*, *216,* Article 119456. https://doi.org/10.1016/j.eswa.2022.119456

Joe Devanny, H.D., & Grossfeld, E. (2023). Generative AI and intelligence assessment. *The RUSI Journal*, *168*(7), 16-25. https://doi.org/10.1080/03071847.2023.2286775

Morande, S. (2023). Benchmarking Generative AI: A comparative evaluation and practical guidelines for responsible integration into academic research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4571867

Open Journal Systems. (n.d.). OJS (Version 3.1) [Open source publishing platform]. https://openjournalsystems.com/

Saranya, A., & Subhashini, R. (2023). A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal, 7*, Article 100230. https://doi.org/10.1016/j.dajour.2023.100230

Sengar, S.S., Hasan, A.B., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, *84*, 23661-23700. https://doi.org/10.1007/s11042-024-20016-1

Sharly. (n.d.). *Sharly AI* [AI tool]. VOX AI Inc. https://sharly.ai/

Silverchair Support. (n.d.). *ScholarOne Manuscripts* [Workflow management system]. Silverchair. https://www.silverchair.com/products/scholarone-manuscripts/