


Dossier: Ética e Integridad, Inteligencia Artificial, Tecnologías de la Información y La comunicación (TIC) y desinformación en la investigación en Humanidades

Are we reading the researchers or the artificial intelligence? Transformations in academic writing after the expansion of generative AI*

Estamos lendo os pesquisadores ou a inteligência artificial? Transformações na escrita acadêmica após a expansão da IA generativa

¿Leemos a los investigadores o a la inteligencia artificial? Transformaciones en la escritura académica tras la expansión de la IA generativa

Daniel Pattier**

 <https://orcid.org/0000-0003-3426-922X>

Abstract: The rapid expansion of generative artificial intelligence (AI) has intensified debates on authorship and authenticity in academic discourse, yet empirical evidence remains limited. This ex post facto study examines how AI has transformed textual quality in education research. A corpus of 1,000 open access articles indexed in Google Scholar was analysed, comparing papers published up to 2021 (pre-AI) with those from 2024 onwards (AI era). Textual quality was assessed using a validated 25-item instrument covering five dimensions: orthographic and grammatical accuracy, cohesion and coherence, adequacy to academic register, style and readability, and formal conventions. Results reveal significant differences. Pre-2021 articles scored higher in accuracy, cohesion, register adequacy, and formal conventions, while post-2024 articles excelled in style and readability. Findings indicate a discursive shift: AI enhances accessibility and fluency but may compromise rigour and authorial distinctiveness. These results highlight the need to reassess academic writing standards in digitalised contexts.

Keywords: Generative artificial intelligence. Academic integrity. Academic writing.

Resumo: A rápida expansão da inteligência artificial (IA) generativa tem intensificado os debates sobre autoria e autenticidade no discurso acadêmico, embora as evidências empíricas ainda sejam limitadas. Este estudo ex-post facto examina como a IA transformou a qualidade textual na pesquisa educacional. Foi analisado um corpus de 1.000 artigos de acesso aberto indexados no Google Scholar, comparando artigos publicados até 2021 (pré-IA) com aqueles publicados a partir de 2024 (era da IA). A qualidade textual foi avaliada por meio de um instrumento validado de 25 itens, abrangendo cinco dimensões: correção ortográfica e gramatical, coesão e coerência, adequação ao registro acadêmico, estilo e legibilidade, e

* This work is the result of the R&D&I project 'The imperative of educational innovation: analysis of its reception and articulation in the Spanish educational system (IMP-NOVA)', with reference PID2022-138878NA-I00, funded by the Spanish Ministry of Science and Innovation.

** PhD in Education. Universidad Carlos III de Madrid. Email: <daniel.pattier@uc3m.es> .

convenções formais. Os resultados revelam diferenças significativas. Os artigos anteriores a 2021 obtiveram pontuações mais altas em correção, coesão, adequação ao registro e convenções formais, enquanto os artigos posteriores a 2024 destacaram-se em estilo e legibilidade. Os achados indicam uma mudança discursiva: a IA melhora a acessibilidade e a fluência, mas pode comprometer o rigor e a distintividade autoral. Esses resultados ressaltam a necessidade de reavaliar os padrões de escrita acadêmica em contextos digitalizados.

Palavras-chave: Inteligência artificial generativa. Integridade acadêmica. Escrita acadêmica.

Resumen: La rápida expansión de la inteligencia artificial (IA) generativa ha intensificado los debates sobre autoría y autenticidad en el discurso académico, si bien la evidencia empírica sigue siendo limitada. Este estudio ex post facto examina cómo la IA ha transformado la calidad textual en la investigación educativa. Se analizó un corpus de 1.000 artículos de acceso abierto indexados en Google Scholar, comparando trabajos publicados hasta 2021 (pre-IA) con aquellos publicados a partir de 2024 (era de la IA). La calidad textual se evaluó mediante un instrumento validado de 25 ítems que abarca cinco dimensiones: corrección ortográfica y gramatical, cohesión y coherencia, adecuación al registro académico, estilo y legibilidad, y convenciones formales. Los resultados revelan diferencias significativas. Los artículos anteriores a 2021 obtuvieron puntuaciones más altas en corrección, cohesión, adecuación al registro y convenciones formales, mientras que los artículos posteriores a 2024 destacaron en estilo y legibilidad. Los hallazgos indican un cambio discursivo: la IA mejora la accesibilidad y la fluidez, pero puede comprometer el rigor y la distintividad autoral. Estos resultados subrayan la necesidad de reevaluar los estándares de escritura académica en contextos digitalizados.

Palabras clave: Inteligencia artificial generativa. Integridad académica. Escritura académica.

Introduction

General context and relevance of the study

Over the past two decades, academic writing has undergone an unprecedented process of transformation. Digital tools, reference managers, and automatic proofreading software had already altered scientific writing routines. However, the emergence of generative artificial intelligence (AI)—particularly since 2022, with the release of large language models such as ChatGPT, Gemini, and Claude—has introduced a qualitative shift that transcends mere technical assistance.

These technologies not only correct, summarise, or rephrase: they generate original language, imitate academic styles, and can compose entire sections of scientific papers. This capacity creates new opportunities but also raises ethical, epistemological, and linguistic questions about authorship, originality, reliability of knowledge, and the very nature of academic discourse (Bozkurt, 2024; Kasani et al., 2024; Currie, 2023).

In this context, the relationship between knowledge and the use of AI and ethical practices in academia has become a growing concern. Elom et al. (2025) demonstrated a significant inverse relationship between exposure to AI and ethical behaviour, accompanied by a positive association with academic misconduct. Moreover, they found that current educational policies are only marginally effective in preventing these practices.

The speed at which AI has been integrated into research practice has outpaced academia's capacity to assess its consequences. Although multiple studies have examined its potential in teaching or in automating scientific tasks, empirical evidence regarding its actual impact on academic writing—and particularly on the textual quality of articles produced in the AI era—remains scarce and fragmented.

AI in scientific production and evaluation

The integration of AI into research has been rapid and wide-ranging. From automated literature searches to abstract generation, generative models are used to accelerate writing processes, improve readability, and reduce formal errors. However, this adoption introduces a fundamental tension between technological efficiency and academic integrity, understood as an ethical stance towards others (Mejía & Garcés-Flórez, 2025).

Several authors (Lund & Naheem, 2024; Raitskaya & Tikhonova, 2023) note that journals and universities are revising their policies to require transparency in the use of AI tools, as the boundary between “assistance” and “partial authorship” is increasingly blurred. Nonetheless, one must also consider the limitations of AI detection tools (Odri & Yoon, 2023), which often misclassify human-written texts as AI-generated (Yoo, 2025) or produce inconsistent detection rates depending on intermediary processes such as translation software (Bisi et al., 2023). Furthermore, studies such as Palmisani et al. (2025) claim that the quality of AI-generated text and code has become so high that it may soon be impossible to determine whether an author has used these tools. Consequently, we may already be encountering academic texts where the true authorship lies, in part, with the AI itself (Öztürk et al., 2025).

This debate goes beyond regulation or the technical accuracy of detection tools (Alawad et al., 2025; Saqib & Zia, 2025) and reaches the epistemological realm: to what extent does the mediation of a language model affect the construction and communication of knowledge? In this vein, Güneş and Kaban (2025) emphasise that the rapid integration of AI into higher education has not only revolutionised scientific production and teaching but also generated increasingly complex ethical dilemmas. Through a Delphi study involving international experts, they identified key concerns such as data privacy, algorithmic bias, transparency in decision-making, and authorship in AI-assisted research. These issues are particularly relevant to contemporary academic production, where the boundary between technological support and algorithmic co-authorship is increasingly blurred. Their findings highlight the urgent need for clear ethical guidelines, robust institutional oversight, and the integration of AI ethics into university curricula to ensure responsible and transparent use of these technologies. From this perspective, the incorporation of AI into academic writing and evaluation must be understood not merely as a technical advance but as a process that requires active academic governance and continuous ethical reflection to safeguard integrity, trust, and sustainability within the research ecosystem.

Despite the growing interest, most studies focus on perceptions, ethical frameworks, or case studies, rather than measurable linguistic evidence on AI’s effect on textual production. As a result, empirical research on formal quality, cohesion, readability, and appropriateness of academic discourse remains at an early stage.

Discursive transformations: from authorial to algorithmic text

Academic writing is traditionally characterised by precision, formality, and conceptual density. The broader context of algorithmic digital culture—characterised by attentional fragmentation and the pursuit of instant gratification—may also be reshaping cognitive dispositions relevant to academic writing and reading (Pattier, 2026). However, texts generated or assisted by AI display a distinct discursive morphology marked by fluency, uniformity, and stylistic predictability (Deep & Chen, 2025; Zhou et al., 2025). Recent studies suggest that generative AI can enhance creative performance at the cost of content diversity, producing outputs that tend to converge toward similar forms. This trend towards homogenisation (defined as the extent to which AI-assisted outputs reduce lexical, semantic, or conceptual diversity) presents a paradox: while it

may improve the formal appearance or readability of the text, it can diminish the author's expressive individuality (Zhou et al., 2025).

Researchers have raised concerns that reliance on AI writing tools may lead to a decline in students' critical thinking skills and proficiency levels (Deep & Chen, 2025; Gustilo et al., 2024). Excessive use of generative AI may limit students' opportunities to develop creative and innovative ideas, as they may bypass the cognitive effort required to construct original arguments (Gustilo et al., 2024). Furthermore, AI cannot be held responsible for generated texts as an author, which raises fundamental questions about authorship and accountability in academic writing (Raitskaya & Tikhonova, 2023).

Consequently, contemporary academic discourse faces a tension between clarity and authenticity. On the one hand, generative models can make writing more efficient and accessible (Deep & Chen, 2025). On the other, the homogenising imprint of generative AI appears to be durable and accumulative, persisting even after AI assistance is withdrawn (Zhou et al., 2025). This phenomenon suggests that AI may be reshaping the texture of scientific language, even when the human author retains control over the production process (Deep & Chen, 2025; Zhou et al., 2025).

To understand these transformations more precisely, it is necessary to conceptualise the constructs underlying this study. The notion of *textual quality* in academic writing is a multidimensional construct shaped by disciplinary conventions, rhetorical traditions, and evolving publishing standards. Empirical research on text processing supports the relevance of readability and clarity as key dimensions of textual quality. Fernández-Silva and Núñez-Cortés (2024), in their study on plain language in administrative texts, demonstrated that textual quality can be meaningfully assessed through reader-oriented variables such as processing time and comprehension accuracy. Their findings reinforce the importance of assessing stylistic and readability features in academic writing, as they directly influence how efficiently readers can process and understand the text.

For the purposes of this study, textual quality is operationalised through five empirically verifiable dimensions (accuracy, cohesion, register adequacy, readability, and formal conventions). Together, these dimensions capture both the micro-linguistic precision and the macro-discursive organisation of scientific texts. Similarly, *authorial voice* (the distinctive presence through which researchers position themselves in relation to their claims and readers) has become a central concern in discussions of AI-mediated writing. While AI-generated content may offer well-structured responses, it often lacks the writer's unique voice and genuine insight (Deep & Chen, 2025).

In the context of generative AI, the risk of voice erosion arises not only from grammatical incorrectness but from the tendency of language models to produce standardised outputs that may smooth over individualised expression. This study approaches authorial voice indirectly through indicators of lexical diversity, syntactic variety, and register adequacy; these proxies offer insight into the stylistic distinctiveness of academic prose.

Finally, the concept of *AI mediation* refers to the process by which generative language models intervene between the author's communicative intentions and the final written product. In this study, AI mediation is understood not as a passive tool but as an active intermediary that filters, reformulates, and potentially transforms the original discourse. While AI mediation can enhance fluency, clarity, and accessibility, it may also attenuate authorial distinctiveness and contribute to discursive homogenisation (Deep & Chen, 2025; Zhou et al., 2025). This conceptualisation aligns with studies on technology-mediated communication, where the mediating agent is seen as shaping

the interaction between the communicator and the audience (Fernández-Silva & Núñez-Cortés, 2024).

Ethical, institutional, and educational challenges

The impact of AI on academic writing extends beyond linguistic issues. It also touches upon fundamental values such as integrity, authorship, and research responsibility. Balalle and Pannilage (2025) warn that using generative tools without critical supervision may compromise transparency and yield texts that are academically plausible yet conceptually hollow. Similarly, Flores Morales et al. (2025) note that the boundary of honesty becomes blurred due to the inherent accessibility of information provided by AI tools.

In the same line, Karkoulian et al. (2025) found that both teachers and students recognize the formative potential of tools such as ChatGPT to enhance productivity and learning, yet they also express ethical concerns related to authorship and academic integrity. The authors emphasize the need for universities to establish clear policies and training programmes to ensure responsible AI use, highlighting that the real challenge lies in balancing innovation with academic honesty within contemporary educational culture. As Wan et al. (2025) point out, it is essential to develop strategies that train educators and students to navigate the ethical landscape governing the use of generative AI.

Concurrently, authors such as Coates et al. (2025) and Kangwa et al. (2025) underscore the urgency of establishing institutional governance frameworks to regulate the ethical use of AI in scientific production. These should be complemented by training in digital competence and AI-assisted academic writing, aimed at integrating technology without replacing critical thinking.

From an educational perspective, Bozkurt (2024) argues that AI not only transforms how writing is produced but also how it is taught and assessed. Traditional evaluation systems—focused on error detection or stylistic appraisal—are no longer sufficient to identify texts produced with algorithmic mediation. Therefore, empirical analysis of textual quality becomes an essential tool for understanding the magnitude of the ongoing transformation.

Gaps in the literature and justification of the study

Despite the proliferation of theoretical works and essays on the relationship between AI and academic writing, quantitative empirical evidence remains limited. Some studies explore teachers' perceptions or pedagogical uses of AI, yet few provide objective measurements of linguistic and discursive impact on real academic texts.

To date, most investigations analyse artificially generated texts or experimental essays, not authentic scientific papers. Hence, it is crucial to compare real outputs of the academic community across periods to determine whether the spread of generative AI has effectively altered textual quality traits such as accuracy, cohesion, readability, or formal adequacy in scholarly writing.

This empirical gap is precisely what the present study seeks to address. Its contribution lies in applying a validated instrument for textual analysis to a large sample of authentic articles, thereby enabling a rigorous statistical examination of measurable changes in the formal quality of contemporary academic writing.

Are we reading the researchers or the artificial intelligence? Transformations in academic writing...

Objectives and hypothesis

The general objective of this study is to analyse the influence of generative AI tools on the textual quality of academic articles in the field of education, comparing those published up to 2021 with those published from 2024 onwards—the period in which AI use became widespread in scientific writing.

Specific objectives:

1. To evaluate differences across the five dimensions of textual quality (accuracy, cohesion, adequacy, style and readability, and formal conventions) between the two periods;
2. To analyse whether the article's thematic focus (educational technology, innovative education, AI in education, digital education, and educational methodologies) significantly influences textual quality;
3. To examine the interaction between year of publication and thematic area to identify differential discursive patterns;
4. To verify the reliability and validity of the evaluation instrument used;
5. To identify potential linguistic or discursive markers associated with the use of generative AI.

Research hypothesis:

H1: The expansion of generative AI tools has significantly transformed the textual quality of academic articles in the field of education, producing statistically significant differences between texts published before and after their widespread adoption.

Method

Study design

A descriptive and comparative ex post facto design was employed, aimed at quantitatively analysing differences in the textual quality of academic articles according to year of publication and thematic area. The approach was non-experimental and cross-sectional, as the texts were already published and no independent variable was manipulated. Analyses were conducted using data obtained through an evaluation instrument specifically designed for this study.

Instrument

To assess the quality of texts produced by scholars in international academic literature, a textual quality analysis instrument was developed to examine various aspects of writing. The instrument includes five evaluation dimensions and a total of 25 items rated on a 5-point Likert scale (1 = poor performance, 5 = excellent performance). The dimensions are as follows:

1. Orthographic and grammatical accuracy: analyses spelling, accentuation, and agreement errors, as well as appropriate use of verb tenses and vocabulary;

2. Cohesion and coherence: examines the overall structure of the text, the appropriateness of connectors, the absence of contradictions, thematic progression, and informational density;
3. Adequacy to academic register: assesses adherence to academic writing conventions, including the use of specialised terminology, formal tone, and absence of colloquialisms;
4. Style and readability: considers expressive clarity, average sentence length, overall readability, lexical variety, absence of ambiguities, and syntactic diversity;
5. Formal conventions: reviews compliance with citation standards, reference formatting, use of headings and subheadings, proper handling of tables and figures, and typographical and layout consistency.

Each article was assessed by ChatGPT-5 according to these 25 items, without being provided with information about the study's hypothesis or objectives, in order to ensure objectivity in the responses. The decision to use ChatGPT-5 as the evaluator was based on its capacity to process full-length articles without segmentation, its consistency in applying predefined criteria under standardised prompting, and the deterministic nature of its responses under identical input conditions, which ensured reproducibility across the corpus.

The instrument provided both quantitative evaluations (average scores by item and dimension) and qualitative observations (descriptive comments on writing features identified in low-performing texts).

Reliability and validity of the instrument

To ensure the psychometric quality of the instrument, procedures for content validation, internal reliability, and structural validity were conducted.

Firstly, content validity was established through the judgement of six experts in applied linguistics, academic writing, and text didactics. All experts held a doctoral degree and had extensive experience in research and teaching in higher education, with publications on academic discourse and textual assessment. They were selected based on their expertise in the field and were invited to participate voluntarily. The experts evaluated the relevance, clarity, and representativeness of each item in relation to its theoretical dimension. They agreed on the adequacy of the 25 items to the construct of academic textual quality, suggesting only minor wording adjustments to improve terminological precision.

Subsequently, internal reliability was analysed using Cronbach's alpha coefficient, which reached a value of $\alpha = .939$, indicating excellent internal consistency. This result demonstrates a high inter-item correlation and confirms that the instrument coherently measures the overall construct.

Finally, an exploratory factor analysis (EFA) was performed using the principal components method with Varimax rotation to examine the structural validity of the instrument. Sampling adequacy indices were excellent (KMO = .934), and Bartlett's test of sphericity was significant ($\chi^2(300) = 2450.12$, $p < .001$), confirming the suitability of the analysis. The EFA revealed five factors with eigenvalues greater than 1, jointly explaining 74.3% of the total variance. This factorial structure fully corresponded to the five theoretical dimensions proposed in the original design: orthographic and grammatical accuracy, cohesion and coherence, adequacy to academic register, style and readability, and formal conventions. Factor loadings (ranging from .48 to .86) showed a clear grouping of items within their respective dimensions, demonstrating a solid

and coherent structural validity consistent with the theoretical model. The highest loadings were found in the dimensions of formal conventions (.74 - .86), orthographic accuracy (.49 - .82), and register adequacy (.56 - .79), while the lowest loadings remained above the acceptable threshold of .40 (Hair et al., 2010), confirming the robust structural validity of the instrument.

Overall, the results of the reliability, factorial validity, and expert judgement analyses confirm that the instrument constitutes a reliable and valid tool for the systematic evaluation of textual quality in academic writing.

Sample

The study sample consisted of a total of 1,000 scientific articles. Selection was carried out through a Google Scholar search conducted in September 2025 using five keyword sets: *educational technology*, *innovative education*, *artificial intelligence in education*, *digital education*, and *educational methodologies*. The search was performed in English to maximise the international representativeness of the sample.

From the results obtained in each keyword set, the highest-ranking articles on the platform were selected, considering only those published in open-access scientific journals and excluding book chapters, conference papers, or other academic document types. The decision to select the first 100 results for each category and year was based on Google Scholar's ranking algorithm, which prioritises relevance and citation impact (Gusenbauer & Haddaway, 2020). While Google Scholar has limitations for systematic reviews, its broad coverage across disciplines and its ability to capture a wide range of scholarly outputs make it a useful tool for exploratory and comparative studies (Martín-Martín et al., 2018). This method provided a transparent and reproducible selection criterion, minimising researcher bias and facilitating future replication.

To enable temporal comparison, the sample was divided into two groups:

1. Group 1 (up to 2021): the first 100 articles in each thematic area published before the emergence of publicly available generative AI tools such as ChatGPT. This group comprises 500 articles representative of the pre-AI period;
2. Group 2 (from 2024 onwards): the first 100 articles in each thematic area published from 2024, when, considering typical review and publication times, texts could plausibly have been produced with AI support. This group comprises 500 articles corresponding to the post-AI period.

Thus, the sample was intentionally and evenly configured, based on both relevance in search results and temporal pertinence to the phenomenon under study.

Procedure

Articles were evaluated independently by ChatGPT-5, which acted as an automated evaluator following a standardised protocol. For each article, the full text was input into the model along with the previously validated analysis instrument, and a score for each of the 25 items was requested on a 1-to-5 scale. The same prompt was used for every article, instructing the model to act as an expert in academic writing assessment, to apply the criteria strictly without introducing external interpretations or value judgements, and to structure the output by dimension and item following a predefined format. Each article was evaluated once, as the model's responses were deterministic under identical input conditions. No additional human raters were involved, as the

primary objective was to test the feasibility and consistency of using a generative AI model as an evaluation tool in a large-scale corpus analysis.

To prevent confirmation bias, the model was not informed of the study's hypothesis or objectives; it was only asked to perform an objective and systematic evaluation of each text.

The resulting scores for the 25 items were recorded in a database and subsequently analysed using IBM SPSS Statistics, version 30. Descriptive analyses (means and standard deviations), independent-sample t-tests, univariate and multivariate analyses of variance (ANOVA and MANOVA), and Tukey post hoc tests were conducted to identify significant differences between groups. The level of significance adopted was $p < .05$ for all tests.

Ethical considerations

Although the study did not involve human participants, principles of transparency, integrity, and responsible AI use were followed. All analysed texts were open-access publications; therefore, no copyright or confidentiality rights were infringed. The use of ChatGPT-5 was restricted exclusively to objective linguistic evaluation, with no human intervention in the final scoring process.

Results

Differences by individual items

To identify more precisely the specific aspects that account for the differences observed between articles published before and after the emergence of generative AI, a detailed analysis was conducted of the 25 items comprising the textual quality evaluation instrument.

Each item was subjected to an independent-samples t-test to compare the mean scores of articles published up to 2021 and those published from 2024 onwards. Table 1 presents the means, standard deviations, t values, and significance levels (p) corresponding to each indicator.

Table 1 - Statistical Comparison of Textual Quality Indicators in Academic Articles by Year of Publication

Item	Articles up to 2021		Articles from 2024		t	p
	\bar{x}	SD	\bar{x}	SD		
1. Absence of spelling errors	4.98	.133	4.91	.339	4.176	<.001
2. Absence of accentuation errors	4.99	.118	4.94	.288	3.165	<.001
3. Grammatical agreement (subject-verb, gender-number)	4.97	.159	4.84	.444	6.253	<.001
4. Appropriate use of verb tenses	4.97	.181	4.78	.464	8.265	<.001
5. Absence of calques or evident lexical errors	4.92	.318	4.66	.575	8.851	<.001
6. Clarity of overall structure (introduction, development, conclusion)	4.95	.222	4.87	.376	3.993	<.001
7. Appropriate use of discourse connectors (causal, consecutive, adversative, etc.)	4.69	.472	4.65	.542	1.307	.096
8. Absence of internal contradictions	4.81	.393	4.78	.445	.980	.164

(Continued)

(Continued)

Item	Articles up to 2021		Articles from 2024		t	p
	\bar{x}	SD	\bar{x}	SD		
9. Clarity in thematic progression (each paragraph introduces a new idea connected to the previous one)	4.71	.463	4.56	.568	4.516	<.001
10. Balanced informational density (neither redundant nor overly terse)	4.35	.569	4.31	.651	.879	.190
11. Predominant use of third person and impersonal/passive voice	4.98	.133	4.88	.365	6.108	<.001
12. Appropriate use of disciplinary terminology	4.97	.165	4.86	.399	5.695	<.001
13. Absence of colloquialisms or inappropriate subjective expressions	4.98	.147	4.84	.444	6.784	<.001
14. Precision in the use of definitions and concepts	4.89	.354	4.62	.598	8.691	<.001
15. Consistency in formal tone throughout the text	4.96	.216	4.75	.518	8.373	<.001
16. Average sentence length (neither excessively long nor fragmented)	4.17	.541	4.09	.611	2.077	.019
17. Readability index	3.67	.738	4.07	.629	-9.177	<.001
18. Lexical variety (lexical diversity index)	4.07	.808	4.38	.597	-6.858	<.001
19. Absence of semantic ambiguities	4.56	.539	4.27	.662	7.595	<.001
20. Variety in syntactic structure	4.18	.452	4.16	.670	.498	.309
21. Citation standards consistently applied (APA, MLA, Chicago, etc.)	4.92	.346	4.79	.557	4.365	<.001
22. Correct formatting of bibliography and references	4.91	.374	4.77	.579	4.544	<.001
23. Use of clear and hierarchically organised headings and subheadings	4.94	.278	4.82	.434	5.397	<.001
24. Inclusion of tables, figures, or appendices correctly referenced	4.79	.604	4.52	.802	6.013	<.001
25. Typographical and formatting consistency (homogeneous use of italics, bold, capitals, section numbering, and other formal elements)	4.93	.299	4.80	.472	5.120	.393

Source: Created by the author.

Overall, the results show that articles published before 2021 achieved significantly higher scores in most items, particularly those related to linguistic accuracy (items 1–5), adequacy to academic register (items 11–15), and formal conventions (items 21–23). In contrast, texts published from 2024 onwards scored higher on items related to readability (item 17) and lexical variety (item 18), with these differences also proving statistically significant ($p < .001$). These findings suggest that, while recent texts exhibit greater fluency and readability, articles published before 2021 maintain greater precision and consistency in the formal aspects of academic discourse.

The standard deviations reveal interesting patterns: items related to basic linguistic accuracy showed very low dispersion ($SD = 0.1–0.3$), indicating that these formal aspects are consistently addressed across most academic texts, regardless of publication period. In contrast, items related

to stylistic and readability features exhibited substantially higher variability (SD often exceeding .700), reflecting greater diversity in writing styles and the varying influence of AI-mediated editing on these dimensions.

Differences by overall dimensions

To synthesise the results, means and standard deviations were calculated for the five dimensions of the instrument, and independent-samples t-tests were applied (Table 2).

Table 2 - Statistical Comparison of Textual Quality Dimensions in Academic Articles by Year of Publication

Dimension	Articles up to 2021		Articles from 2024		t	p
	\bar{x}	SD	\bar{x}	SD		
Orthographic and grammatical accuracy	4.97	.153	4.83	.355	7.953	<.001
Cohesion and coherence	4.70	.336	4.64	.410	2.770	.003
Adequacy to academic register	4.96	.157	4.79	.395	8.823	<.001
Style and readability	4.13	.400	4.22	.796	-2.255	.012
Formal conventions	4.90	.325	4.76	.760	3.599	<.001

Source: Created by the author.

The results indicate that articles published before 2021 obtained significantly higher scores in orthographic and grammatical accuracy ($t = 7.953$, $p < .001$), cohesion and coherence ($t = 2.770$, $p = .003$), adequacy to academic register ($t = 8.823$, $p < .001$), and formal conventions ($t = 3.599$, $p < .001$).

In contrast, texts published from 2024 onwards showed higher mean scores in style and readability ($t = -2.255$, $p = .012$), reflecting greater expressive clarity and ease of reading. Taken together, the results confirm that articles published prior to 2021 maintain greater accuracy and formal adequacy, whereas those published from 2024 are characterised by greater readability and fluency, possibly associated with the increasing use of generative AI tools.

Differences according to article theme

To examine whether textual quality varies according to theme, one-way analyses of variance (ANOVA) were conducted for each of the five dimensions. The independent variable was the thematic category of the article (five categories: *educational technology*, *innovative education*, *artificial intelligence in education*, *digital education*, and *educational methodologies*). Additionally, Tukey post hoc tests were applied to identify between which thematic areas the significant differences occurred. Table 3 summarises the results obtained. The analyses revealed statistically significant differences across all dimensions ($p < .001$).

Table 3 - Results of the Analysis of Variance (ANOVA) by Thematic Area

Dimension	F	p	η^2	Significant differences (Tukey, $p < .05$)
Orthographic and grammatical accuracy	8.231	< .001	.032	<i>Educational technology</i> < <i>Innovative education</i> , <i>Digital education</i>
Cohesion and coherence	47.832	< .001	.161	<i>AI in education</i> , <i>Digital education</i> > others
Adequacy to academic register	8.699	< .001	.034	<i>Educational methodologies</i> < <i>Digital education</i> , <i>AI in education</i>
Style and readability	44.354	< .001	.151	<i>AI in education</i> > others; <i>Educational methodologies</i> lowest
Formal conventions	6.874	< .001	.027	<i>Educational technology</i> < others

Source: Created by the author.

Note: η^2 = effect size. Significant differences established at $p < .05$.

The largest discrepancies were found in cohesion and discourse coherence ($F(4, 995) = 47.832, p < .001, \eta^2 = .161$) and style and readability ($F(4, 995) = 44.354, p < .001, \eta^2 = .151$), both showing large effect sizes. The dimensions of accuracy ($F(4, 995) = 8.231, p < .001$), adequacy ($F(4, 995) = 8.699, p < .001$), and formal conventions ($F(4, 995) = 6.874, p < .001$) also revealed significant differences, although with small to moderate effect sizes. Post hoc Tukey tests indicated that the most notable differences were observed among the thematic areas *Educational technology*, *Artificial intelligence in education*, *Digital education*, and *Educational methodologies*, following these patterns:

1. Cohesion and style were significantly higher in *Artificial intelligence in education* and *Digital education*;
2. Adequacy and formal conventions were lower in *Educational methodologies*;
3. Accuracy showed lower scores in *Educational technology* compared with *Innovative education* and *Digital education*.

Combined effect of publication year and thematic area

Finally, to jointly examine the effect of publication year and thematic area, a multivariate analysis of variance (MANOVA) was performed, considering the five dimensions of textual quality as dependent variables.

The multivariate results (Table 4) showed significant effects for publication year, thematic area, and their interaction ($p < .001$ in all cases):

Publication year: Wilks' $\Lambda = .840, F(5, 986) = 37.67, p < .001, \eta^2 p = .160$

Thematic area: Wilks' $\Lambda = .444, F(20, 3271) = 45.43, p < .001, \eta^2 p = .184$

Year \times Thematic area interaction: Wilks' $\Lambda = .510, F(20, 3271) = 36.81, p < .001, \eta^2 p = .155$

Table 4 - Results of the Multivariate Analysis of Variance (MANOVA) by Publication Year and Thematic Area

Source	Wilks' Λ	F	df (Hypothesis)	df (Error)	p	η^2p
Publication year	.840	37.672	5	986	< .001	.160
Thematic area	.444	45.428	20	3271	< .001	.184
Year \times Thematic area	.510	36.813	20	3271	< .001	.155

Source: Created by the author.

Note: All multivariate tests were significant at the $p < .001$ level.

These results confirm that overall textual quality differs significantly according to both the publication period and the thematic area, and that the effect of publication year is not uniform across all research fields.

Complementary univariate analyses showed that differences associated with publication year were significant in all dimensions: accuracy ($F(1, 990) = 66.45, p < .001$), cohesion ($F(1, 990) = 11.05, p < .001$), adequacy ($F(1, 990) = 80.83, p < .001$), style and readability ($F(1, 990) = 6.52, p = .011$), and formal conventions ($F(1, 990) = 13.26, p < .001$).

Regarding the thematic area, the largest differences were found in cohesion ($F(4, 990) = 58.40, p < .001, \eta^2p = .191$) and style ($F(4, 990) = 48.64, p < .001, \eta^2p = .164$). The year \times thematic area interaction was significant for accuracy ($F(4, 990) = 5.72, p < .001$), cohesion ($F(4, 990) = 53.47, p < .001$), and style ($F(4, 990) = 23.68, p < .001$), indicating that the temporal evolution of academic discourse varies across thematic domains.

Discussion

The results confirm the research hypothesis: significant differences in textual quality exist between articles published before and after the widespread adoption of generative AI. These findings provide a more precise understanding of the impact of AI on recent academic writing, while also suggesting the existence of additional factors that may have influenced the differences observed between publication periods.

The comparative analysis revealed that articles published up to 2021 achieved higher scores on orthographic and grammatical accuracy, cohesion and coherence, adequacy to academic register, and formal conventions, whereas texts published from 2024 onwards performed better in style and readability. These statistically significant differences ($p < .001$ for most dimensions) reflect meaningful shifts in academic writing practices. The higher scores in accuracy, cohesion, and formal conventions among pre-2021 articles suggest that traditional academic writing—shaped by human editorial processes and peer review—tends to prioritise linguistic precision, structural consistency, and adherence to established scholarly conventions. In contrast, the superior performance of post-2024 articles in style and readability indicates that AI-assisted writing may enhance clarity and accessibility, potentially broadening the readership of academic texts. This trend partially aligns with the conclusions of Deep and Chen (2025) and Zhou et al. (2025), who note that AI-mediated texts tend to display cleaner syntax, more predictable structures, and greater discursive clarity.

The finding that more recent articles are more readable but less formally precise reinforces the hypothesis that AI is driving a discursive shift in academic writing—one oriented towards communicative efficiency and stylistic simplification. However, this shift also raises concerns about a potential trade-off: while AI tools may improve the fluency and surface-level quality of academic prose, they may also contribute to a reduction in the depth, conceptual precision, and authorial distinctiveness that characterise scholarly communication (Deep & Chen, 2025; Zhou et al.,

2025). The fact that pre-2021 articles scored higher across most dimensions calls for a more nuanced interpretation.

A possible explanation—beyond the temporal effect of technological adoption—lies in the positioning of texts within Google Scholar, from which the sample was extracted. Google’s algorithm prioritises relevance and accumulated citations, which may enhance the visibility of better-written, more influential papers over time. Consequently, articles published before 2021, with greater consolidation and higher ranking, may appear more frequently in top search results, thereby representing samples of higher formal quality. Conversely, more recent articles (from 2024 onwards) have not yet had sufficient time to achieve comparable visibility, which may partially explain their lower scores without necessarily implying inferior textual quality.

This methodological nuance is relevant, as it suggests that temporal positioning in academic search engines could act as an intervening variable in studies of this kind and should be considered in future research.

The differences observed across thematic areas also reinforce the notion that the adoption of AI is not homogeneous within the educational field. Articles on *Artificial Intelligence in Education* and *Digital Education* obtained the highest scores in cohesion and readability, whereas those on *Educational Methodologies* and *Educational Technology* exhibited a more traditional style and lower formal adequacy. This pattern is consistent with Bozkurt (2024), who argues that technology-oriented disciplines tend to integrate AI tools earlier—not only in teaching and research, but also in academic writing practices. Accordingly, the results suggest a differentiated evolution of discursive styles: technological domains show greater adaptation to AI-mediated writing models, while methodological areas retain more conventional structures.

From an ethical and epistemological perspective, these findings must be interpreted with caution. Authors such as Balalle and Pannilage (2025) and Coates et al. (2025) emphasise that the impact of AI goes beyond technical aspects of writing, affecting core values such as academic integrity and authorship itself. While AI mediation may enhance certain features of academic writing, it also poses the risk of diluting the author’s voice, reducing discursive diversity and critical thinking (Raitskaya & Tikhonova, 2023).

Although the present study does not empirically confirm a homogenisation of discourse, recent literature suggests that increasing automation may be leading to stylistic convergence between AI-generated and human-authored texts (Gustilo et al., 2024; Currie, 2023).

Furthermore, the factorial and reliability analyses confirm that textual quality can be measured rigorously through observable and quantifiable dimensions, providing a useful tool for empirical research on academic discourse. In this sense, the study offers a methodological contribution to applied linguistics and textual assessment by proposing a replicable model for investigating the influence of AI on scientific writing.

Overall, the findings indicate that the emergence of generative AI is reconfiguring academic style: more recent texts are clearer and more readable, yet they also show a trend towards formal simplification compared with pre-2021 articles. As noted by Kangwa et al. (2025) and Gulumbe et al. (2025), this scenario calls for ethical and institutional frameworks to ensure the responsible use of technology, promoting transparency and training in AI-assisted writing.

As Coates et al. (2025) warn, the integration of generative technologies requires a balance between communicative efficiency and the preservation of discursive authenticity. Likewise, the results of this study should also be interpreted in light of the “organic versus synthetic” paradox

highlighted by Bozkurt (2024). Academic writing—traditionally conceived as a cognitive and discursive process grounded in human creativity—is now mediated by algorithmic systems capable of reproducing its external forms. This tension between the *organic* (creativity, authorship, critical voice) and the *synthetic* (efficiency, correctness, textual optimisation) redefines the parameters of quality and originality in scholarly communication.

In this context, the use of generative AI tools raises profound ethical and pedagogical implications concerning academic integrity, critical competence, and research evaluation. As several authors observe, this transformation is intensified by the institutional “publish or perish” culture (Elbanna & Child, 2023), which privileges productivity over reflection and may encourage the instrumental use of AI to meet editorial demands rather than as a genuine intellectual aid. Thus, the organic–synthetic paradox describes not only a technological shift but also a cultural mutation in the way academic writing is conceived, where the pursuit of efficiency risks eclipsing the epistemic value of authorship and individual thought.

For all these reasons, academic institutions should promote training in AI-assisted writing, establish transparency policies regarding AI use, and develop regulatory frameworks that safeguard academic integrity and authorial responsibility (Gulumbe et al., 2025; Kangwa et al., 2025; Pattier & Redondo-Duarte, 2025). Far from being an inevitable threat, AI can become a pedagogical and scientific ally—provided that its integration is guided by critical reflection and authorial accountability.

Conclusions

The present study aimed to analyse the influence of generative AI tools on the textual quality of academic articles in the field of education, comparing papers published up to 2021 and from 2024 onwards, and examining the role of thematic area and its potential interaction.

The research hypothesis is confirmed: significant differences in textual quality exist between the two publication periods. However, these differences do not reflect a uniform improvement. Articles published from 2024 onwards show significant gains in style and readability, while pre-2021 articles scored higher in grammatical accuracy, cohesion and coherence, adequacy to academic register, and formal conventions. This suggests that AI enhances certain surface-level features of academic writing—such as fluency and accessibility—while potentially compromising others, including formal precision and conceptual rigour. The observed pattern points to a differential transformation of academic discourse rather than a straightforward enhancement.

In relation to the specific objectives, the analyses confirm that textual quality varies significantly not only by publication period but also by thematic area. Pre-2021 articles stand out in accuracy, cohesion, register adequacy, and formal conventions, while post-2024 articles excel in style and readability. Regarding thematic focus, the more technology-oriented fields (Artificial Intelligence in Education and Digital Education) obtained the highest scores in cohesion and readability, whereas more traditional areas (Educational Methodologies and Educational Technology) maintained more conventional discursive structures. The interaction between publication year and thematic area revealed that the temporal evolution of textual quality is not uniform: improvements in cohesion and readability are concentrated in the most digitalised domains. The instrument employed proved to be reliable and valid ($\alpha = .939$), with a factor structure consistent with the proposed theoretical model. Finally, readability and lexical variety emerge as potential discursive markers associated with AI use, while the decline in formal adequacy and conceptual precision in recent texts suggests a possible trade-off between fluency and rigour.

Beyond the influence of AI, various contextual factors may be contributing to the observed shifts in academic writing. The growing emphasis on accessibility and dissemination, driven by open access policies and the increased visibility of research through platforms such as Google Scholar, may encourage authors to prioritise readability and clarity over formal complexity. Likewise, the evolution of journal guidelines—many of which now encourage plain language summaries and broader audience engagement—may influence writing styles independently of AI use. The institutional "publish or perish" culture may also play a role, as researchers face mounting pressure to produce a high volume of outputs, potentially leading to more standardised and less distinctive writing. Finally, changes in authorship patterns, such as the rise of international collaborations and the increasingly widespread use of English as a lingua franca, could contribute to the homogenisation of academic discourse observed in recent texts. These factors, although not directly examined in this study, should be considered in future research to disentangle the specific effects of AI from broader transformations in academic publishing and research culture.

Taken together, these findings indicate that generative AI enhances clarity, fluency, and lexical variety, but human intervention remains essential to ensure linguistic precision, conceptual coherence, and academic rigour. The results point to a reconfiguration of academic style, in which technology optimises written expression and readability, but also appears to be associated with a partial loss of linguistic control and formal adequacy—posing new ethical and methodological challenges for scientific communication.

From an ethical and pedagogical perspective, the findings invite reflection on the role of AI in the production of knowledge. The growing use of these tools underscores the need to preserve authorship and intellectual responsibility in a context where technological mediation increasingly shapes writing processes. It is therefore essential to strengthen training in digital competence and academic ethics, so that researchers and educators can integrate AI critically—harnessing its potential to improve scientific communication without relinquishing authenticity, creativity, or independent thought.

Among the main limitations of the study are the non-probabilistic nature of the sample, selected from Google Scholar search results, which may introduce visibility and ranking biases. In addition, the analysis focused exclusively on indicators of textual quality, without addressing deeper semantic or argumentative dimensions; inferences about the influence of AI should therefore be interpreted with caution. Furthermore, the use of ChatGPT-5 as the sole evaluator, without cross-validation with human raters, may introduce systematic biases in the assessment of textual quality. Future studies could adopt mixed methods, combining corpus analysis with interviews with authors or editors, to explore perceptions and regulations surrounding generative AI in academic production. Likewise, it would be relevant to extend the research to other disciplinary fields and languages to determine whether the observed trends are specific to education or form part of a broader transformation in contemporary scientific discourse. Future investigations could also examine whether AI involvement in editorial review processes contributes to modifying the discursive patterns identified.

References

Alawad, E. A., Ayadi, H. H., & Alhinai, A. A. (2025). Guarding Integrity: A Case Study on Tackling AIGenerated Content and Plagiarism in Academic Writing. *Theory and Practice in Language Studies*, 15(6), 1730–1742. <https://doi.org/10.17507/tpls.1506.02>

- Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, 11, 1–12. <https://doi.org/10.1016/j.ssaho.2025.101299>
- Bisi, T., Risser, A., Clavert, P., Migaud, H., & Dartus, J. (2023). What is the rate of text generated by artificial intelligence over a year of publication in Orthopedics & Traumatology: Surgery & Research? Analysis of 425 articles before versus after the launch of ChatGPT in November 2022. *Orthopaedics & Traumatology: Surgery & Research*, 109(8), 1–7. <https://doi.org/10.1016/j.rcot.2023.09.014>
- Bozkurt, A. (2024). GenAI et al. Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis*, 16(1), 1–10. <https://doi.org/10.55982/openpraxis.16.1.654>
- Coates, H., Croucher, G. & Calderon, A. (2025). Governing Academic Integrity: Ensuring the Authenticity of Higher Thinking in the Era of Generative Artificial Intelligence. *Journal of Academic Ethics*, 23, 2015–2028. <https://doi.org/10.1007/s10805-025-09639-7>
- Currie, G. M. (2023). Academic integrity and artificial intelligence: Is ChatGPT hype, hero or heresy?. *Seminars in Nuclear Medicine*, 53(5), 719–730. <https://doi.org/10.1053/j.semnuclmed.2023.04.008>
- Deep, P. D., & Chen, Y. (2025). The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*, 15(9), 1–19. <https://doi.org/10.3390/soc15090247>
- Elbanna, S., & Child, J. (2023). From ‘publish or perish’ to ‘publish for purpose’. *European Management Review*, 20(4), 614–618. <https://doi.org/10.1111/emre.12618>
- Elom, C. O., Ayanwale, M. A., Ukeje, I. O., Offiah, G. A., Umoke, C. C., & Ogbonnaya, C. E. (2025). Does AI Knowledge Encourage Cheating? Investigating Student Perceptions, Ethical Engagement, and Academic Integrity in the Digital Age. *International Journal of Learning, Teaching and Educational Research*, 24(4), 708–729. <https://doi.org/10.26803/ijlter.24.4.33>
- Fernández-Silva, S., & Núñez-Cortés, J. A. (2024). Exploring the Effect of Plain Terminology on Processing and Comprehension of Administrative Texts in Spanish: A Self-Paced Reading Experiment. *International Journal of Applied Linguistics*, 35, 659–671. <https://doi.org/10.1111/ijal.12650>
- Flores Morales, J. A., Anglas La Torre, M. E., & Solorzano Arias, S. (2025). Inteligencia Artificial e integridad en los trabajos académicos de estudiantes en una Universidad de Lima. *European Public & Social Innovation Review*, 11, 1–20. <https://doi.org/10.31637/epsir-2026-1897>
- Gulumbé, B.H., Audu, S.M. & Hashim, A. M. (2025). Balancing AI and academic integrity: What are the positions of academic publishers and universities?. *AI & Society*, 40, 1775–1784. <https://doi.org/10.1007/s00146-024-01946-8>
- Güneş, A., & Kaban, A. L. (2025). A Delphi study on ethical challenges and ensuring academic integrity regarding AI research in higher education. *Higher Education Quarterly*, 79(4). <https://doi.org/10.1111/hequ.70057>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research synthesis methods*, 11(2), 181–217. <https://doi.org/10.1002/jrsm.1378>

Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in AI era. *International Journal for Educational Integrity*, 20, 1–43. <https://doi.org/10.1007/s40979-024-00153-8>

Kangwa, D., Msafiri, M. M., & Fute, A. (2025). Exploring the Factors That Promote a Balance Between Academic Integrity and the Effective Use of GenAI Tools in Higher Education: A Systematic Review. *Journal of Computer Assisted Learning*, 41(5). <https://doi.org/10.1111/jcal.70109>

Karkoulian, S., Sayegh, N., & Sayegh, N. (2025). ChatGPT Unveiled: Understanding Perceptions of Academic Integrity in Higher Education - A Qualitative Approach. *Journal of Academic Ethics*, 23, 1171–1188. <https://doi.org/10.1007/s10805-024-09543-6>

Kasani, P. H., Cho, K. H., Jang, J. W., & Yun, C. H. (2024). Influence of artificial intelligence and chatbots on research integrity and publication ethics. *Science Editing*, 11(1), 12–25. <https://doi.org/10.6087/kcse.323>

Lund, B. D., & Naheem, K. T. (2024). Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learned Publishing*, 37(1), 13–21. <https://doi.org/10.1002/leap.1582>

Mejía, A., Garcés-Flórez, M. F. (2025). What do we mean by academic integrity?. *International Journal for Educational Integrity*, 21, 1–20. <https://doi.org/10.1007/s40979-024-00176-1>

Odri, G. A., & Yoon, D. J. Y. (2023). Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity. *Orthopaedics & Traumatology: Surgery & Research*, 109(8), 1–6. <https://doi.org/10.1016/j.rcot.2023.10.007>

Öztürk, A., Karahan, A. T., Günay, S., Erdal, A. S., Komut, S., Komut, E., & Yiğit, Y. (2025). Artificial intelligence as author: Can scientific reviewers recognize GPT-4o-generated manuscripts?. *The American Journal of Emergency Medicine*, 97, 216–219. <https://doi.org/10.1016/j.ajem.2025.07.034>

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1120–1177. <https://doi.org/10.1016/j.joi.2018.09.002>

Palmisani, F., Segelcke, D., & Vollert, J. (2025). Navigating the light and shadow of scientific publishing faced with machine learning and generative AI. *European Journal of Pain*, 29(3), 1–8. <https://doi.org/10.1002/ejp.4736>

Pattier, D. (2026). Tiktonian Brain: Cognition and Learning in Algorithmic Digital Culture. *Revista Interuniversitaria*, 38(2), 211–229. <https://doi.org/10.14201/teri.32815>

Pattier, D., & Redondo-Duarte, S. (2025). La vida online, la inteligencia artificial y su lectura pedagógica. *Márgenes, Revista De Educación De La Universidad De Málaga*, 6(1), 28–45. <https://doi.org/10.24310/mar.6.1.2025.20784>

Raitskaya, L., & Tikhonova, E. (2023). Academic Integrity: Author-Related and Journal-Related Issues. *Journal of Language and Education*, 9(4), 5–10. <https://doi.org/10.17323/jle.2023.18489>

Saqib, M. B., & Zia, S. (2025). Evaluation of AI content generation tools for verification of academic integrity in higher education. *Journal of Applied Research in Higher Education*, 17(4), 1430–1440. <https://doi.org/10.1108/JARHE-10-2023-0470>

Wan, G., Grassi, M. H., Golden, T., Barnes, T., Kahveci, M., Wan, X., & Colacchio, B. (2025). Artificial Intelligence and Academic Integrity: Legislate or Educate?. *Journal of Scholarly Publishing*, 56(2), 320–376. <https://doi.org/10.3138/jsp-2024-1118>

Yoo, J. H (2025). Defining the Boundaries of AI Use in Scientific Writing: A Comparative Review of Editorial Policies. *Journal of Korean Medical Science*, 40(23), 1–13. <https://doi.org/10.3346/jkms.2025.40.e187>

Zhou, Y., Liu, Q., Huang, J., & Li, G. (2025). Creative scar without generative AI: Individual creativity fails to sustain while homogeneity keeps climbing. *Technology in Society*, 84. <https://doi.org/10.1016/j.techsoc.2025.103087>

Recibido: 19/03/2026

Versión corregida recibida: 22/06/2026

Aceptado: 24/06/2026

Publicado online: 29/06/2026