Journal of Intelligent Decision Making and Information Science

www.jidmis.org
eISSN: 3079-0875

Linguistic Features of AI-Generated Academic Texts and the Role of Human Editing

Tetiana Nedashkivska[1]*, Inna Varvaruk[2], Mykhailo Podoliak[3], Khrystyna Dzyubynska[4], Alla Yarova[5], Svitlana Lytvynska[6]

[1]Department of Slavic and Germanic Philology and Translation, Educational and Scientific Institute of Philology and Journalism, Zhytomyr Ivan Franko State University, Zhytomyr, Ukraine

[2]Department of Ukrainian Philology and Social Sciences, Higher Education Institution "King Danylo University", Ivano-Frankivsk, Ukraine

[3]Department of Philology named after Iakym Iarema, Faculty of Management, Business and Public Administration, Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies Lviv, Lviv, Ukraine

[4]Faculty of Management, Business and Public Administration, Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies Lviv, Lviv, Ukraine

[5]Department of Journalism and Philology, Faculty of Foreign Philology and Social Communications, Sumy State University, Sumy, Ukraine

[6]Department of Journalism and Linguistic Communication, Faculty of Humanities and Pedagogy, National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

ARTICLE INFO

Article history:
Received 10 April 2026
Received in revised form 01 May 2026
Accepted 19 May 2026

Keywords: Large Language Models; Academic Discourse; Human-AI Collaboration; Linguistic Features; Human Editing.

ABSTRACT

The rapid spread of large language models (LLMs) has significantly transformed academic writing practices and actualized discussions about authorship, language quality, and academic integrity. At the same time, diachronic changes in academic discourse during the active implementation of generative artificial intelligence remain insufficiently studied. The study combines a systematic literature review with a corpus diachronic analysis of authentic academic annotations, allowing us to trace long-term trends in the development of academic discourse. The aim of the work is to identify linguistic changes in academic writing during 2015–2025 and determine the role of human editing in quality assurance of AI-assisted scientific texts. The research material was a corpus of 870 English-language annotations of scientific articles indexed in the Scopus database in the field of arts and humanities. Quantitative linguistic analysis was carried out using Python tools. Indicators of lexical density, lexical diversity, syntactic complexity, average sentence length and frequency of cohesive markers were analyzed. A statistically significant increase in lexical density and frequency of cohesive markers has been revealed, indicating an increase in information compression and explicit discursive organization of texts. Indicators of traditional lexical diversity and syntactic complexity remained relatively stable. The observed trends are consistent with characteristics described in AI-assisted writing studies; however, the study design does not allow them to be directly related to the use of large language models. Human editing remains a key factor in ensuring factual accuracy, discursive coherence, lexical enrichment, and academic integrity. The results can be used to develop practices for the responsible use of generative AI in academic communication.

1. Introduction

The rapid spread of generative artificial intelligence, particularly large language models (LLMs) like GPT-4, has dramatically changed the landscape of academic writing. More and more scholars are using these tools to create drafts, paraphrase, and edit scholarly texts. This actualizes the question of the linguistic consequences of such use and the transformation of the boundaries between human authorship and machine assistance.

Although large language models demonstrate high grammatical correctness and compliance with the genre conventions of academic discourse, evidence is accumulating that they produce texts with reduced lexical variability, increased wording, and increased local coherence, often at the expense of global coherence and originality. Such transformations require a deep understanding of the differences between traditional human writing and AI-generated texts, as well as elucidating the role of the human editor in compensating for the limitations of machine generation.

This study has two interrelated goals. First, to analyze diachronic linguistic changes in the abstracts of scientific articles for the period 2015–2025, with a particular focus on the period after the widespread adoption of ChatGPT at the end of 2022. Second, to investigate the compensatory functions of human editing in the context of AI-assisted academic writing.

By integrating a systematic review of the literature with empirical corpus analysis, the paper seeks to find out whether the recorded linguistic shifts can be related to the influence of generative models, as well as to determine how human intervention contributes to the preservation of the scientific quality of texts.

The results of the study should contribute to the discussion on the implications of using generative AI for scientific communication, authorship and academic integrity, as well as provide practical recommendations for researchers, editors and higher education institutions in the context of the development of human-machine collaboration in writing.

2. Literature Review

The current stage of development of linguistics and applied linguistics is characterized by the significant influence of Large Language Models (LLMs), which transform not only the methods of text generation, but also the very nature of academic writing. Starting with the advent of transformer architectures [1] and the subsequent development of scale models such as GPT-3 and GPT-4 [2, 3], researchers are increasingly considering LLM as a new type of discursive agent capable of mimicking complex genres of scientific speech. In this context, the question arises about the linguistic boundaries between human and machine writing, as well as the degree of authenticity of academic texts generated by artificial intelligence.

One of the key characteristics of LLMs is their ability to generate grammatically correct and stylistically consistent texts that meet the formal requirements of academic discourse. However, research shows that such language production is often based on statistical prediction of subsequent tokens rather than a deep understanding of the content [4, 5]. This leads to the phenomenon of

fluency without grounding, where the text looks convincing but does not necessarily contain verifiable or cognitively sound information [6].

From the point of view of lexical organization, numerous studies demonstrate that LLM texts are characterized by limited variability and high repeatability of lexical patterns. In particular, analysis of academic text corpora suggests that large language models, in particular ChatGPT, show a tendency to use standardized academic constructs and repetitive discursive patterns, resulting in a decrease in individual stylistic variability between texts [7, 8]. Similar results are supported by corpus studies, which show that after the widespread introduction of large language models into scientific writing, there is a tendency towards stylistic convergence and a decrease in lexical diversity compared to human texts, which manifests itself in more standardized academic constructions and reduced variability of discursive markers [9, 10].

The syntactic structure of texts generated by large language models can show certain tendencies related to the simplification of the hierarchical organization of sentences. In particular, corpus studies show that LLMs in a number of corpus studies tend to have a greater proportion of coordinate constructions that provide linear organization of information, while human writing is characterized by a more frequent use of subordination, which forms a deeper syntactic hierarchy and more complex dependencies between clauses [11, 12]. This trend may be due to the fact that coordinate structures are cognitively easier to generate and process, but at the same time reduce the level of syntactic nesting and structural complexity of the text [13].

In addition, while LLM texts may exhibit high local coherence and formal grammatical correctness, they are often characterized by weaker global discourse coherence. Studies in the field of text generation show that models are able to maintain coherence at the level of individual sentences, but their argumentative structure between sentences and paragraphs is less stable compared to human texts, which manifests itself in the phenomena of semantic drift and violation of global coherence [14, 15].

Modern research shows that large language models are able to reproduce typical genre patterns of academic writing, including structural elements such as introductions, methods, and conclusions, but such organization often remains superficial and formulaic and does not guarantee deep argumentative coherence of the text [16, 17].

An important problem that is actively investigated is the phenomenon of hallucinations in large speech models. A study by Ji et al. [15] shows that LLMs can generate false or fictitious claims, often in the form of high confidence. Similarly, Maynez et al. [14] demonstrate that even in text summarization problems, models often violate the principles of actual source matching. This is seen as a potential challenge to academic integrity, especially in cases of automated scientific text generation [18].

In this context, much attention is paid to the influence of large language models on the stylistic characteristics of texts in the process of human–AI interaction. Research shows that the use of LLMs in the writing process can influence the choice of lexical and structural decisions of authors, forming certain patterns of stylistic adaptation within the co-creation of text [19]. This phenomenon is

considered as an important area of study of changes in written practices under the influence of generative models.

At the same time, the problem of automatic detection of AI texts remains open. Many studies demonstrate the low reliability of existing detectors, which often have both false positive and false negative results [20, 21]. This indicates the limitations of purely automatic identification methods and emphasizes the expediency of considering more complex linguistic and discursive approaches to text analysis.

A separate area of research is the role of human editing in the process of creating AI-assisted texts. Human-in-the-loop approaches are seen as an important element in ensuring the quality of academic writing, as they allow to compensate for the limitations of large language models, including factual inaccuracies and logical inconsistencies [16, 22]. In this context, editing acts not only as a technical stage, but also as a cognitive process of interpreting and restructuring the text generated by the algorithm.

Overall, the current literature demonstrates that LLMs are significantly changing the linguistic landscape of academic writing, shaping new models of textual production that combine machine generation and human interpretation. The result is a complex hybrid writing system in which the boundaries between author, tool, and editor are gradually blurred.

3. Methodology

For the study, a corpus of 870 English-language annotations of scientific articles published in 2015–2025 and indexed in the Scopus database was formed. The corpus included peer-reviewed scientific articles on issues of academic writing, academic discourse, and scientific annotations. In order to ensure genre and disciplinary homogeneity of the material, the selection was limited to publications in English in the field of arts and humanities. This approach made it possible to form a comparable corpus of academic annotations, suitable for diachronic analysis of linguistic changes.

Data processing procedure. All annotations have been pre-cleaned using a specially designed Python script. The procedure included: removing the text after the © character (copyright and license disclaimers); normalization of spaces and removal of newline characters; protection of standard scientific abbreviations (e.g., i.e., et al., Ph.D., etc.) for correct tokenization.

Linguistic analysis. Quantitative analysis was carried out using Python 3 program code using pandas, nltk, re libraries. For each text, the following indicators were calculated:

Lexical level: Word count – number of tokens after tokenization and filtering; Type-Token Ratio (TTR) is a classic indicator of lexical diversity; Measure of Textual Lexical Diversity (MTLD) is a more stable indicator for the length of the text; Lexical density is the ratio of content words to the total number of words.

Syntactic level: Average Sentence Length; Syntactic complexity is the average number of clauses per sentence, determined by counting subordinate conjunctions using regular expressions.

Discursive level: Frequency of cohesive markers per 1000 words – counting of key discursive markers (however, therefore, moreover, furthermore, thus, in addition, on the other hand, etc.). For the accuracy of counting phrases, the method of exact match was used, taking into account the boundaries of words.

All metrics were normalized in text length where necessary. To detect diachronic changes, a one-way analysis of variance (One-way ANOVA) with the factor "Year of publication" was used, as well as checking linear trends. The strength of the effect was estimated using η^2 (eta squared).

Software implementation. The calculations were carried out in the IDLE (Integrated Development and Learning Environment) environment for Python using open libraries. To ensure the reproducibility of the study, all parameters (list of stop words, cohesive markers, MTLD = 0.72 threshold) were fixed.

Research limitations: The analysis is limited to the genre of scientific annotations in the arts and humanities, which does not allow for direct extrapolation of the results to other academic genres or disciplines.

4. Results and Discussion

The analysis of modern literature allows us to identify a number of linguistic characteristics that are most often associated with academic texts generated by large language models. Previous research suggests that such texts demonstrate a relatively high level of grammatical correctness and formal coherence, but at the same time are characterized by a number of specific stylistic and discursive features that distinguish them from human academic writing [4, 16, 7].

Based on the generalization of the results of previous studies, a list of key linguistic parameters that can be used for comparative analysis of human and AI-generated annotations was formed (Table 1).

Table 1. Linguistic features frequently reported in AI-generated academic texts

Linguistic feature	Typical manifestation in AI-generated texts	Confirmatory studies
Lexical diversity	Decrease in lexical diversity indicators and increased frequency of formulaic and repeated lexical constructions	Martínez et al. [10]; Tudino and Qin [7]
Stylistic variability	Greater standardization of style and reduced author's personality	Liu and Demberg [8]; Lin et al. [9]
Academic phraseology	Predominance of formulaic academic language and standardized discursive patterns	Tudino and Qin [7]; Kasneci et al. [16]
Syntactic complexity	Possible tendency to lower syntactic complexity (lower proportion of subordinate constructions)	Liu and Liu [11]
Local coherence	Relatively high local coherence between adjacent sentences	Ji et al. [15]

Global coherence	Potential difficulties in maintaining global discursive coherence and factual consistency between sentences	Ji et al. [15]; Maynez et al. [14]
Actual reliability	Risk of factual errors and generation of non-factual content ("hallucinations")	Ji et al. [15]; Stokel-Walker and Van Noorden [18]
Genre imitation	Reproduction of standard academic genre structures at the surface level	Kasneci et al. [16]; Liang et al. [17]
Detectability	The increasing difficulty of distinguishing between AI and human writing, which reduces the accuracy of automatic detectors	Weber-Wulff et al. [20]; Sadasivan et al. [21]

Source: built by the authors

As shown in Table 1, the most consistently described characteristics of AI-generated academic texts in the literature are a decrease in lexical diversity, an increased level of stylistic standardization, and the use of template academic constructions [8, 10, 7]. Such features are explained by the statistical nature of large language models, which reproduce the most likely language sequences based on large learning corpora [4, 5].

At the same time, studies show conflicting results regarding the syntactic complexity of AI-generated texts. While some papers record a tendency to use less complex syntactic structures compared to human writing, others point to the ability of modern models to successfully reproduce complex academic constructions depending on the genre and task at hand [11, 16].

Particular attention is drawn to the issue of discursive organization of the text. Researchers note that large language models are able to provide high local coherence between adjacent sentences, but maintaining argumentative logic at the level of the entire text remains problematic, especially in complex analytical genres [14, 15]. That is why the assessment of global coherence is increasingly considered as one of the key criteria for distinguishing between human and AI-generated academic writing.

A separate aspect is the role of human editing. Previous research suggests that editorial intervention can largely compensate for typical shortcomings of LLM texts, including factual errors, excessive templates, and insufficient argumentative depth [16, 22]. Therefore, most modern approaches to the academic use of generative AI are based on the human-in-the-loop model, which combines automated text generation with human quality control.

The linguistic characteristics of AI-generated academic texts summarized in Table 1 demonstrate that the key differences between machine and human writing are manifested mainly at the level of lexical organization, syntactic complexity, and discursive coherence. At the same time, these features are not final or immutable markers. Since they can be largely modified in the process of post-editing and academic validation of the text.

In this context, special attention is paid to the role of human editing as an intermediate link between the automatically generated text and the final academic publication. Editing itself acts as a mechanism for correcting structural, stylistic, and factual shortcomings of LLM-generated content, as

well as a tool for its adaptation to the requirements of scientific discourse. A summary of the main functions of such an intervention is presented in Table 2.

Table 2. The Role of Human Editing in Improving AI-Generated Academic Texts

Human Editing Feature	Main contribution to the text	Typical editing results	Confirmatory sources
Factual verification	Identifying and correcting potentially false or unverified claims	Potential reduction of the risk of factual errors due to human verification	Ji et al. [15]; Stokel-Walker and Van Noorden [18]
Coherence Editing	Rearrangement of logical connections between sentences and paragraphs	Potential improvement in global discursive coherence	Maynez et al. [14]; Kasneci et al. [16]
Stylistic adaptation	Correction of pattern and excessive formality	Reduction of template and excessive formalization of style	Tudino and Qin [7]; Lin et al. [9]
Lexical enrichment	Expanding the dictionary variability of the text	Reducing the repetition of lexical units and formula expressions in the text	Martínez et al. [10]; Liu and Demberg [8]
Restructuring the argument	Rebuilding the logic of evidence and inference	Improving logical consistency and structured argumentation	Ouyang et al. [22]; Kasneci et al. [16]
Genre alignment	Adaptation of the text to the requirements of the academic genre	Ensuring structural compliance with academic genre models	Liang et al. [17]; Kasneci et al. [16]
Reduction of discursive pattern	Eliminate repetitive phrases and overly standard constructions	Reduction of discursive template and standardization of the text	Lin et al. [9]; Tudino and Qin [7]
Ethical and Academic Compliance	Control over the correctness of citations and prevention of plagiarism	Reducing the risk of incorrect citations and textual borrowings	Stokel-Walker and Van Noorden [18]; Ouyang et al. [22]
Human "alignment"	Interpretation and refinement of AI-generated ideas	Clarify the meaning and increase the contextual consistency of the interpretation of AI-generated text	Bender et al. [4]; Kasneci et al. [16]

Source: built by the authors

Table 2 systematizes the key functions of human editing in the process of working with AI-generated academic texts and demonstrates how editorial intervention transforms the initial machine-generated product into a scientifically relevant and academically valid text.

As can be seen from the table, human editing performs a multi-level function, covering both the micro-level (lexical enrichment, stylistic adaptation, elimination of patterning) and the macro-level of the text (restructuring of argumentation, ensuring global coherence and genre alignment). Especially important is the factual verification function, which compensates for the limitations of large language models associated with the generation of false or unverified information [15, 18].

In addition, editing plays a critical role in restoring the discursive integrity of the text, as it allows for the elimination of logical gaps and increased argumentative consistency, which is especially important for complex analytic genres [14]. At the same time, stylistic and lexical modification contributes to the reduction of excessive standardization characteristic of LLM-generated texts and the return of elements of authorial individuality [7, 9].

Thus, Table 2 demonstrates that human editing is not only a technical stage of correction, but acts as a cognitive intermediary between algorithmically generated text and the requirements of academic writing, providing a transition from formal linguistic correctness to meaningful scientific quality.

As part of the empirical part of the study, a corpus of 870 English-language scientific articles obtained from the Scopus database was analyzed. The main objective of the analysis was to identify diachronic changes in the linguistic characteristics of annotations (2015–2025) from the perspective of the hypothesis of gradual LLM-stylization of academic writing, i.e. a shift towards more standardized, cognitively "medium-optimized" textual structures similar to the output of large language models.

For each annotation, a set of quantitative linguistic indicators was calculated, which are conventionally divided into three groups: lexical, syntactic and discursive.

Lexical characteristics were represented by indicators of text volume, lexical diversity and density. Specifically, the total length of the annotation was defined as the number of tokens:

$$\text{Words} = N_{\text{tokens}}.$$

The size of the vocabulary was calculated as the number of unique tokens:

$$\text{Vocabulary size} = \mid V \mid.$$

The classic Lexical Diversity Index (Type-Token Ratio) was defined as:

$$TTR = \frac{\mid V \mid}{N}$$

where $\mid V \mid$ is the number of unique words, and N is the total number of tokens.

For a more stable assessment of lexical variability, the MTLD (Measure of Textual Lexical Diversity) index was used, which is based on the average length of token sequences before reaching the threshold level of variance:

$$MTLD = \frac{N_{\text{tokens}}}{\text{number of factors}}.$$

Lexical density was defined as the share of content words in the total volume of the text:

$$LD = \frac{N_{\text{content words}}}{N_{\text{total words}}}.$$

The syntactic level was represented by the average length of the sentence and the index of syntactic complexity. The average sentence length was calculated as:

$$ASL = \frac{N_{\text{words}}}{N_{\text{sentences}}}.$$

Syntactic complexity was approximated through the average number of clausal structures in the sentence:

$$SC = \frac{\sum_{i=1}^{S} C_i}{S},$$

where C_i the number of clausal markers in the th sentence, a S is the total number of sentences.

The discursive level was measured through the frequency of cohesive markers normalized per 1000 words:

$$CM_{1000} = \frac{N_{\text{cohesion markers}}}{N_{\text{words}}} \times 1000.$$

To detect changes over time, single-factor analysis of variance (ANOVA) with the Year factor was used, as well as checking linear trends (linear vs deviation from linearity) and estimating the strength of the effect through eta squared.

The most consistent changes were recorded at the lexical and discursive levels.

Lexical level. Lexical density shows a steady positive trend during the study period (Figure 1): the average value increased from 0.66 in 2015 to 0.70 in 2025. Analysis of variance revealed a highly significant linear effect of the year of publication ($F = 57.14$, $p < 0.001$), and the proportion of explained variation was $\eta^2 = 0.082$. This indicates a gradual increase in the specific weight of content units in the structure of annotations and is consistent with the results of modern corpus studies, which record a long-term increase in the information density of academic texts. In particular, on large corpora of academic writing, lexical density in scientific annotations has been shown to tend to increase in diachronic perspective, with this effect being relatively stable across disciplines [23]. Similar results have been obtained in studies of the long-term evolution of academic discourse, where a gradual increase in lexical density and general informational compression of texts has also been recorded [24].

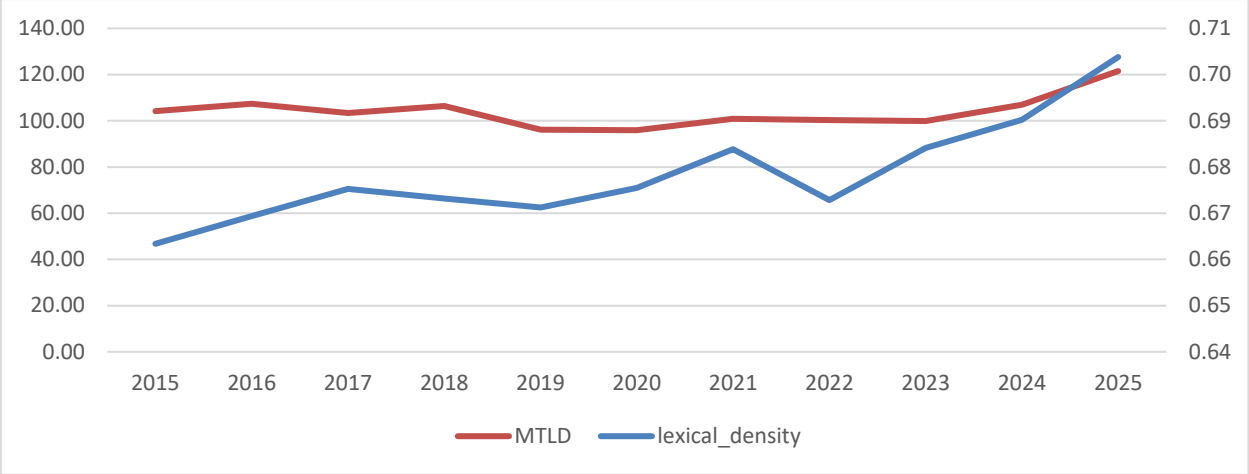

Fig. 1. Dynamics of lexical density indicators and Measure of Textual Lexical Diversity in annotations

Source: built by the authors

Thus, the results obtained can be interpreted as part of a broader trend in the development of academic writing, which is characterized by an increase in information density and stabilization of genre norms in international scientific discourse.

A similar trend is observed for the MTLD score, which reflects lexical diversity and is resistant to the influence of text length [25]. In the study corpus, the MTLD values show an overall increase during the observation period (especially in 2024-2025), and the linear trend check revealed a statistically significant effect ($F = 8.91$, $p = 0.003$). Although the magnitude of the effect remains small ($\eta^2 = 0.038$), the results are consistent with current corpus studies, which show that texts generated or partially processed by generative language models, in particular ChatGPT, exhibit increased lexical diversity compared to traditional academic writing [26]. The obtained results indicate a moderate increase in the lexical diversity of annotations during the last years of the study period. Contrary to assumptions about possible lexical unification or simplification of academic writing, the recorded dynamics do not indicate a decrease in the variability of vocabulary. At the same time, the small size of the effect indicates that lexical changes are gradual and can be associated with both the general evolution of academic discourse and transformations of modern academic writing practices.

Discursive level. The frequency of cohesive markers normalized per 1000 words increased from 1.40 in 2015 to 4.18 in 2024 ($F = 7.04$, $p = 0.008$, $\eta^2 = 0.024$). This increase indicates an increased explicitness of logical connections and is a characteristic feature of texts created or edited using large language models [15, 16, 9].

Syntactic level. At the same time, no systematic changes were found in the indicators of syntactic complexity ($F = 1.26$, $p = 0.252$; $\eta^2 = 0.014$), as well as in the classic index of lexical diversity TTR, for which the linear trend turned out to be statistically insignificant ($F = 2.42$, $p = 0.120$). Mean TTR values ranged from 0.58–0.63, which is consistent with the results of modern studies of academic texts and AI-generated writing [11, 26]. The absence of significant changes in the TTR may indicate the relative stability of the lexical diversity of annotations, despite the increase in other parameters (lexical density and cohesion markers).

Overall, the results do not show a simplification of academic writing, but a gradual increase in information density, lexical variability, and discursive coherence of annotations. Such changes are consistent with the general evolution of academic discourse towards a more standardized, reader-oriented and information-rich presentation. At the same time, the data obtained do not give grounds to assert the emergence of characteristic features of lexical or syntactic unification, which could be unambiguously interpreted as a consequence of the mass use of generative artificial intelligence systems. Rather, they reflect the continuation of long-term trends in the development of academic writing recorded in previous corpus studies of scientific discourse.

In general, the data obtained indicate a partial stylistic convergence of academic annotations after 2022–2023. An increase in lexical density and frequency of cohesive markers indicates an increase in the formal organization of texts, typical of the influence of generative models. At the same time, the stability of TTR and syntactic complexity indicates that the influence of LLMs in the genre of scientific annotations is manifested mainly at the surface level.

These results highlight the importance of human editing in the process of creating AI-assisted texts. Editorial intervention allows you to compensate for excessive standardization, restore argumentative depth, eliminate factual inaccuracies, and return individual features to the text. Thus, the human-in-the-loop approach acts not only as a technical stage, but as a cognitive mechanism for ensuring the scientific quality of academic writing in the context of the widespread use of generative artificial intelligence.

5. Conclusions

The study revealed statistically significant diachronic changes in the linguistic characteristics of academic annotations for the period 2015–2025. In particular, an increase in lexical density and frequency of cohesive markers was recorded, while measures of syntactic complexity and traditional lexical diversity remained relatively stable. The identified trends are consistent with individual characteristics described in AI-assisted academic writing studies, but the data obtained do not allow for a direct causal relationship between these changes and the use of large language models.

The scientific novelty of the study lies in the quantitative fixation of diachronic changes in the genre of scientific annotations – the most standardized and conservative genre of academic discourse – during the period of active implementation of generative AI models. Unlike previous papers, which mostly compared "human vs AI" texts under experimental conditions, this study offers a real diachronic cross-section of academic writing, allowing us to distinguish between long-term genre trends and the impact caused by LLM.

The results obtained allow us to draw an important theoretical conclusion: the influence of generative artificial intelligence on academic discourse is manifested mainly at the superficial stylistic level (increase in formal coherence and information density), while deep syntactic mechanisms and basic lexical variability remain relatively stable. This suggests that modern LLMs have more influence on form than structure of academic thinking.

The practical significance of the work lies in substantiating the need for a human-in-the-loop model as an obligatory component of academic writing-creation in the era of AI. Human editing acts not

only as a stage of correction, but as a key cognitive and ethical mechanism that compensates for excessive stylistic standardization, restores the individual author's voice, and ensures the factual and argumentative reliability of the text.

Thus, in modern conditions, academic writing is increasingly acquiring the features of hybrid cognitive activity, in which artificial intelligence performs the function of a powerful generative tool, and a person performs the function of an intellectual editor and guarantor of scientific quality. Promising areas of further research are a comparative analysis of the impact of LLMs on different academic genres, as well as the development of linguistic markers that will allow for more accurate identification of the degree of human intervention in AI-assisted texts.

At the same time, the design of the study does not allow directly linking the detected changes with the use of large language models, since the corpus does not contain information about the way individual texts are created. Therefore, the observed trends should be seen as diachronic changes in academic discourse that are consistent with the results of AI-assisted writing research, but cannot be considered as direct evidence of the impact of LLMs.

Author Contributions

Conceptualization, T.N. and H.I.; methodology, T.N. and M.P.; software, K.D.; validation, H.I., K.D. and M.P.; formal analysis, M.P. and A.Y.; investigation, K.D. and A.Y.; resources, S.L.; data curation, K.D. and M.P.; writing—original draft preparation, T.N. and H.I.; writing—review and editing, M.P., A.Y. and S.L.; visualization, K.D.; supervision, A.Y. and S.L.; project administration, H.I.; funding acquisition, T.N. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (NeurIPS 2017). https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., & Amodei, D. (2020). Language models are few-shot learners. *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

[3] OpenAI. (2024). GPT-4 technical report. *arXiv*. https://arxiv.org/abs/2303.08774

[4] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT'21) (pp. 610-623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

[5] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2022). On the opportunities and risks of foundation models. *arXiv*. https://arxiv.org/abs/2108.07258

[6] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv*. https://arxiv.org/abs/2303.12712

[7] Tudino, G., & Qin, Y. (2024). A corpus-driven comparative analysis of AI in academic discourse: Investigating ChatGPT-generated academic texts in social sciences. *Lingua, 312*, 103838. https://doi.org/10.1016/j.lingua.2024.103838

[8] Liu, D., & Demberg, V. (2023). ChatGPT vs human-authored text: Insights into controllable text summarization and sentence style transfer. In V. Padmakumar, G. Vallejo, & Y. Fu (Eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Volume 4: Student Research Workshop) (pp. 1–18). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-srw.1

[9] Lin, D., Zhao, N., Tian, D., & Li, J. (2025). ChatGPT as linguistic equalizer? Quantifying LLM-driven lexical shifts in academic writing. *arXiv*. https://doi.org/10.48550/arXiv.2504.12317

[10] Martínez, G., Hernández, J. A., Conde, J., Reviriego, P., & Merino-Gomez, E. (2024). Beware of words: Evaluating the lexical diversity of conversational LLMs using ChatGPT as case study. *ACM Transactions on Intelligent Systems and Technology, 16*(6), 138, 1–15. https://doi.org/10.1145/3696459

[11] Liu, W., & Liu, X. (2025). A comparative analysis of syntactic complexity in argumentative essays from a rhetorical perspective: ChatGPT vs. English native speakers. *PLOS ONE, 20*(8), e0329410. https://doi.org/10.1371/journal.pone.0329410

[12] Nasseri, M. (2021). Is postgraduate English academic writing more clausal or phrasal? Syntactic complexification at the crossroads of genre, proficiency, and statistical modelling. *Journal of English for Academic Purposes, 49*, 100940. https://doi.org/10.1016/j.jeap.2020.100940

[13] Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

[14] Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

[15] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). *Survey of hallucination in natural language generation. ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

[16] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Stadler, M. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[17] Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Vodrahalli, K., He, S., Smith, D. S., Yin, Y., McFarland, D. A., & Zou, J. (2024). Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis. *NEJM AI*, *1*(8). https://doi.org/10.1056/aioa2400196

[18] Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature, 614*(7947), 214–216. https://doi.org/10.1038/d41586-023-00340-6

[19] Gero, K. I., Long, T., & Chilton, L. B. (2023). Social dynamics of AI support in creative writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23) (Article 245, pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3544548.3580782

[20] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., et al. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, 26. https://doi.org/10.1007/s40979-023-00146-z

[21] Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2025). Can AI-generated text be reliably detected? *arXiv*. https://arxiv.org/abs/2303.11156

[22] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv*. https://arxiv.org/abs/2203.02155

[23] Zhu, H., Wang, T., & Pang, N. (2024a). Diachronic changes in lexical density of research article abstracts: A corpus-based study. *Lingua, 312*, 103837. https://doi.org/10.1016/j.lingua.2024.103837

[24] Zhu, H., Wang, T., & Pang, N. (2024b). Investigating diachronic changes in lexical density of academic texts: A corpus-based study. *SSRN*. https://doi.org/10.2139/ssrn.4774270

[25] McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

[26] Fredrick, D. R., & Craven, L. (2025). Lexical diversity, syntactic complexity, and readability: A corpus-based analysis of ChatGPT and L2 student essays. *Frontiers in Education, 10*, 1616935. https://doi.org/10.3389/feduc.2025.1616935