Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE IMPACT OF ARTIFICIAL INTELLIGENCE ALGORITHMS ON AUTOMATIC EDITING AND PROOFREADING OF TEXTS

OLEKSII SYTNYK[1], NATALIA KNYRIK[2], OKSANA TIUTIUNNYK[3], YEVGENYA BAZINYAN[4], TETIANA TSEPKALO[5]

[1]Taras Shevchenko National University of Kyiv, Department of Editorial and Publishing Technologies and Production, Ukraine

[2]Admiral Makarov National University of Shipbuilding, Department of Information Control Systems and Technologies, Ukraine

[3]Hryhorii Skovoroda University in Pereiaslav, Department of Ukrainian and Foreign Literature and Teaching Methods, Ukraine

[4]Armenian State University of Economics, Armenia

[5]Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Department of Journalism, Advertising and Public Relations, Ukraine

E-mail: [1]o.sytnyk@gmail.com, [2]nadiia.hope.p@gmail.com, [3]oksanatiutunyk@gmail.com, [4]yevgeniabazinyan1@gmail.com, [5]t.tsepkalo@vspu.edu.ua

ABSTRACT

The relevance of the study is determined by the rapid development of artificial intelligence (AI) and the need to assess its impact on the quality of automatic editing of English texts in the global informational technologies (IT) environment. The aim was to identify and assess the impact of various AI algorithms on the efficiency and quality of automatic editing and proofreading of English texts in an international context. The objective of this study is to identify and analyze how various AI algorithms impact the efficiency and quality of automatic editing and proofreading of English texts in an international context. The study employed the following methods: experimental methods, content analysis, comparative analysis, and focus groups. The reliability was confirmed by using qualitative statistical analysis, Student's t-test, chi-square test, and analysis of variance (ANOVA). The assessment was carried out in comparison with human editing. The texts included news, scientific articles, advertising texts, essays, and technical documentation. Readability analysis showed statistically significantly lower indicators of AI compared to editors (Flesch-Kincaid: 69.4 versus 75.2). The users noted the convenience and speed of intelligent technologies, but pointed to insufficient flexibility and stylistic problems. Analysis of variance confirmed the significant advantage of editors in quality ($F=12.56$, $p<0.001$, $\eta^2=0.45$). The results obtained during the study showed a significant advantage of professional editors over modern AI algorithms in editing quality. At the same time, its use demonstrates the potential to accelerate the process, which is important for practical application. Prospects for further research include improving AI algorithms for stylistic editing and studying hybrid methods that combine automation with human expertise.

Keywords: *Natural Language Analysis; Neural Networks; Proofreading Quality; Text Processing; Text Proofreading*

1. INTRODUCTION

The evolution of artificial intelligence (AI) algorithms has transformed the paradigm of editing and proofreading texts. This has made the process more accessible, faster and more cost-effective [1]. Algorithms integrated into platforms such as Grammarly, DeepL, and ProWritingAid are capable of processing significant amounts of text data. They detect a variety of errors and offer corrections in real time [2].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

However, despite rapid progress, the issue of the effectiveness of these systems compared to professional editors remains open. This especially applies to stylistic naturalness, readability, and contextual adaptation of texts [3]. The problem is that algorithms, while demonstrating high accuracy in detecting spelling and grammatical errors, often do not take into account stylistic, genre, and semantic nuances. This can lead to the creation of technically correct but stylistically inappropriate texts [4],[5]. In contrast, professional editors provide higher stylistic harmony and naturalness due to their experience [6].

The focus of this study is on a comparative analysis of the quality of text editing by AI algorithms and professional editors. The study includes an analysis of the quality of error correction, identification of algorithm weaknesses, and study of user experience. The obtained results are intended to deepen the understanding of the capabilities and limitations of AI. This will also contribute to the improvement of technologies and provide valuable insights for the optimal combination of human experience and automation.

The aim of the study is to identify and analyse the impact of different AI algorithms on the efficiency and quality of automatic editing and proofreading of English texts in an international context. The objectives of the study:

- Compare the quality of automatic editing and proofreading of English texts performed by different AI algorithms with the quality of editing performed by professional human editors;

- Identify the types of errors (grammatical, spelling, stylistic, punctuation) that are most effectively detected and corrected by different AI algorithms;

- Study the perception and evaluation of the results of automatic editing and proofreading by users (editors, authors) in terms of usability, accuracy of corrections, and time spent on the process.

This study differentiates itself from prior work by its applied, comparative, and interdisciplinary nature. While previous research has often focused on isolated aspects of AI editing, our motivation is to provide a holistic empirical evaluation that directly benchmarks multiple state-of-the-art AI algorithms against professional human editors across a diverse set of criteria – including not only error correction rates but also readability scores and qualitative user experience. This approach allows us to move beyond the question of 'if' AI can correct errors to 'how well' it performs across the full spectrum of editing tasks compared to the human gold standard, and to understand the practical implications for users.

The conducted research has academic novelty in theoretical and practical terms. In the theoretical aspect, a comprehensive comparative analysis of the effectiveness of AI algorithms and professional editors was carried out for the first time. The analysis took into account different types of text errors, including grammatical, spelling, stylistic, and semantic.

Research hypothesis. AI algorithms demonstrate effectiveness in correcting text errors, but are inferior to professional editors in ensuring stylistic naturalness and readability of texts.

Research Questions:

1. Does AI edit texts in English qualitatively compared to human editors?

2. What errors in English texts are best detected and corrected by different AIs?

3. How do users evaluate the convenience and accuracy of AI editing tools?

4. Does the perception of different AI tools differ between editors and authors?

5. What is the general experience of using AI for automatic text editing?

Scientific Problem. The study addresses the efficacy of AI-driven text editing and proofreading tools compared to human editors in ensuring high-quality, stylistically natural, and contextually appropriate English texts. While AI algorithms (e.g., Grammarly, DeepL, ProWritingAid) excel in detecting grammatical and spelling errors, their ability to handle stylistic, semantic, and contextual nuances remains limited. This creates a gap between technical correctness (achievable by AI) and human-like readability and coherence (typically ensured by professional editors).

2. LITERATURE REVIEW

Basic concepts of artificial intelligence in the context of natural language processing (NLP).

Modern natural language processing relies on transformer models that can analyse wide contexts. Lin et al. [7] emphasize the significant advantage of transformers over previous models in the error correction accuracy. However, their limitation is their focus on English texts only, which complicates their application in multilingual environments. Lin [8] emphasizes the unexpected results of transformers, deepening the understanding of their functionality, but its analysis requires better structuring. These studies touch on fundamental concepts for understanding AI algorithms in text editing.

AI algorithms for automatic editing and proofreading of texts.

Deep neural networks, in particular the T5 model, demonstrate significant accuracy in text editing tasks. However, as Mallinson et al. [9] point out in their study, the hybrid approach to editing may not sufficiently cover complex stylistic errors. At the same time, the study by Jiang [10] deeply analyses the processing of large text arrays, but focuses mainly on formal aspects, bypassing stylistic nuances. The work of Kadim and Lazrek [11] makes a significant contribution to improving the processing of natural Arabic language. The authors present an extended version of the Nemlar corpus, which significantly improves the quality of linguistic analysis because of more accurate morphological parsing and tagging of parts of speech. The resource has the potential to improve the efficiency of text editing algorithms, in particular in the tasks of error correction, syntax normalization, and data preparation for deep learning. Nevertheless, the aforementioned studies are an important basis for understanding the potential of the algorithms to be evaluated in our study.

Automatic error detection and correction principles.

Error correction methods are based on the analysis of syntax, semantics, and style of texts. Wang et al. [12] provides an overview of existing approaches to grammatical correction, including statistical and neural models. Its advantage is the granularity of the analysis, but it does not take into account modern transformer architectures. Raju et al. [13] indicate that Large Language Models (LLMs) demonstrate high efficiency in correcting stylistic errors. At the same time, the authors note that these models can create new errors when editing complex texts. The mentioned studies provide the background to explore in detail the issue of the operation of algorithms with different types of errors, which is important for this study.

Modern tools and platforms for automatic editing.

Tools like Grammarly and ChatGPT have significantly changed the editing experience. Feuerriegel et al. [14] emphasize the productivity of human-AI collaboration. However, empirical evidence on the effectiveness of these platforms is lacking. Hardack [15] focuses on the ethics of using AI in editing, but her analysis is too broad. Both studies are important for outlining the strengths and weaknesses of the tools considered in our paper.

Advantages and limitations of using AI in the editing process.

AI offers fast and accessible editing, but sensitivity to style and context remains its weakness. Chakrabarty et al. [16] revealed the tendency of AI to standardize language, which is critical in creative texts. Gasaymeh et al. [17] emphasize the risk of losing cultural diversity, but suggest ways to minimize it. At the same time, the researchers note the increase in adaptability of texts because of semi-autoregressive models. The mentioned studies define important criteria for the comparative evaluation of AI algorithms and human editors in our study.

General analysis of the literature and unresolved issues of existing research.

Despite the value of the presented studies, the issue of using AI algorithms for editing natural language requires further study and improvement. A critical analysis of the academic literature indicates significant progress in the development of algorithms for automated text editing. At the same time, certain limitations in taking into account stylistic, cultural and ethical nuances are revealed. The shortcomings of a significant part of the research include the predominant concentration on technical aspects, while is strengths are the proposed innovative solutions and the potential for practical application of the obtained results. The considered theoretical principles are the basis for further empirical research on the effectiveness of AI and its synergy with professional editing.

A critical synthesis of existing literature reveals that while previous studies have effectively documented the raw capabilities of AI in grammatical error correction (GEC), they often present a fragmented view. Research has predominantly followed two parallel tracks: one focusing on the technical evolution of algorithms (e.g., transformer models, F1-scores) [7, 9, 12], and another exploring broader user perceptions or ethical considerations [14, 15]. A significant gap exists in empirically bridging these perspectives—specifically, in using rigorous, comparative methodology to quantify how these technical shortcomings manifest in tangible quality metrics (like readability indices) and are perceived by professional end-users in a direct A/B test against human expertise. Most studies evaluate AI tools in isolation rather than in a controlled, comparative setting against the benchmark of professional human editing.

This article analyses issues that require deeper study in the modern academic paradigm. In particular, the quality of correction of various types of errors by AI algorithms in comparison with professional human editors will be empirically assessed. Particular attention will be paid to the ability of algorithms to ensure stylistic perfection and optimal readability of texts. In addition, user experiences regarding the productivity and ergonomics of automated editing tools will be investigated, as well as their impact on the established process of professional editorial activity.

3. METHODOLOGY

3.1. Design

The research is experimental, applied, and interdisciplinary. Type – comparative analysis. Sort – empirical, using quantitative and qualitative methods. The research was conducted in three stages.

1. The first — preparatory — stage. Selection of texts of different genres with typical errors. The texts were edited by three AI algorithms (Grammarly, DeepL, ProWritingAid) and professional editors.

2. The second — empirical — stage. A comparative analysis of the quality of error correction was conducted using quantitative methods, including readability assessments (Flesch-Kincaid, Gunning Fog) and ANOVA. Error classification was also performed.

3. The third — analytical — stage. The data from focus groups, including editors and authors, were collected for a qualitative analysis of the perception of AI algorithms. Thematic analysis identified key aspects of convenience, efficiency, and impact on the workflow.

3.2. Participants

Five groups of English texts were selected for the study, reflecting genre and stylistic diversity. In particular, the texts contained news, scientific articles, advertising texts, essays and technical documentation. The volume of each text was 300–500 words, and the number of errors reached 15 per 100 words, which ensured their adequate complexity for analysis. The errors included grammatical, spelling, stylistic and punctuation aspects to assess the effectiveness of corrections in different conditions.

The group of professional editors included 10 specialists with work experience from 5 to 10 years. The specialists had an education in the field of philology or linguistics and spoke English at a level not lower than C1. All editors had experience working with automated editing systems and texts of varying complexity, which ensured the representativeness of their results. Twenty people were invited to participate in the focus groups: 10 professional editors and 10 copywriters, aged 25 to 45, most of whom had higher education. The participants had between 1 and 5 years of experience working with AI algorithms. This allowed for a wide range of experiences and assessments of usability, accuracy of corrections, and impact on the workflow.

The study employed strict inclusion/exclusion criteria to ensure methodological rigor. Texts were selected based on genre representation (news, scientific articles, advertising, essays, technical documentation), length (300–500 words), and error density (15 errors/100 words). Non-English texts, highly specialized content (e.g., legal/medical jargon), and multimodal materials were excluded. Professional editors were required to have 5–10 years of experience and C1+ English proficiency; AI tools were limited to widely adopted, NLP-driven platforms (Grammarly, DeepL, ProWritingAid). These criteria controlled variables while reflecting real-world editing scenarios.

3.3. Study variables

This study precisely operationalizes variables within a comparative framework, assessing the efficacy of AI-driven versus human editing. The independent variable is the editing method, categorized into four groups: Grammarly, DeepL, and ProWritingAid (representing AI algorithms), and professional human editors (as the control). Dependent variables quantify editing quality through readability metrics (Flesch-Kincaid Grade Level, Gunning Fog Index), error-correction accuracy (F1-score, BLEU score), and error classification rates across grammatical, spelling, stylistic, punctuation, and semantic categories. User experience is qualitatively evaluated via focus groups, capturing perceived usability, speed, and flexibility. To ensure validity, control variables standardize experimental conditions, including text characteristics (genre, length, error density), editor expertise (experience, English proficiency), and technical settings (uniform AI tool configurations, standardized NLP protocols for metric computation). This design is crucial for isolating the impact of editing methods on outcome variables.

3.4. Data collection

1. *The experiment* involved comparing the quality of editing English texts performed by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

different AI algorithms and professional editors. For this purpose, texts of varying complexity with typical grammatical, stylistic and other errors were selected. Each text was sequentially processed by several common AI algorithms and an independent group of professional editors.[18]

2. *Content analysis* aimed to identify patterns in the work of editing algorithms. Errors were classified into the following categories: grammatical, spelling, stylistic, punctuation, and semantic. It was explored which errors were corrected, omitted or created.[19]

3. A *comparative analysis* of the stylistic characteristics (readability and naturalness) of texts edited by AI algorithms and professional editors was conducted. For this purpose, metrics from the field of natural language processing (NLP), in particular the Flesch-Kincaid and Gunning Fog indices, were applied. Each edited text was evaluated according to readability and language complexity indicators.[20]

4. *Focus groups* were used to collect qualitative data on the experiences of users of automatic text editing tools. Editors and authors who use similar solutions were involved. During moderated discussions, the participants discussed the ease of use of algorithms, their effectiveness and impact on the workflow. The obtained data were analysed to identify key themes and trends.

The study used several groups of tools. Popular AI algorithms such as Grammarly, Deeple, and ProWritingAid were used for automatic text editing. These models were chosen because of their popularity among users and their ease of use. The study used the Flesch-Kincaid and Gunning Fog metrics to assess the quality of editing, which determined the level of readability of texts. BLEU was also used to compare automatically corrected texts with reference ones, and F1 score was used to assess the accuracy of error detection and correction. The Flesch-Kincaid and Gunning Fog readability indices were calculated based on the average sentence length, the number of syllables in words, and the proportion of complex words in the text. The BLEU metric was determined by comparing n-grams (up to the 4th order) of automatically edited texts with versions prepared by professional editors. F1 score was calculated as the harmonic mean of accuracy and completeness based on comparing corrected errors with the total number of existing errors. All metrics were applied using automated tools and provided an objective analysis of the quality of editing. The data obtained during the focus groups were recorded using audio recordings and transcribed as transcripts. The data analysis was performed using thematic analysis software such as NVivo and coding schemes for error classification. SPSS and Python packages with appropriate modules were used for statistical analysis. Tableau tools and Python modules such as Matplotlib and Seaborn were used for visualization of the results.

3.5. Statistical techniques and procedures

In the statistical analysis, the sample size was 200 units, of which each group included: Grammarly algorithm (n=50), DeepL algorithm (n=50), ProWritingAid algorithm (n=50), editors (n=50). A set of methods was used for statistical analysis to ensure the reliability of the conclusions. *ANOVA* compared the average editing qualities between AI and editors. *Student's t-test* detailed pairwise differences. *Chi-square test* revealed patterns in the types of errors processed. Descriptive statistics assessed the basic characteristics of the data. *Qualitative thematic analysis* of focus groups revealed key user views. The selected methods combine quantitative and qualitative approaches, which provide a comprehensive picture of the effectiveness and peculiarities of the work of AI algorithms and human editing. Such a methodology ensures the reliability of the results and their practical significance.

All statistical analyses in this study were conducted with a 95% confidence level ($\alpha=0.05$), aligning with conventional standards in linguistic and computational research. This threshold established a critical p-value of 0.05 as the boundary for statistical significance. Accordingly, outcomes with $p<0.05$ were deemed statistically significant, indicating a $\leq 5\%$ probability that observed differences arose by chance. For pairwise comparisons following ANOVA, the Bonferroni correction was applied to control family-wise error rates, tightening the critical p-value to 0.0083 (0.05/6 comparisons). Significant results ($p<0.05$) were further contextualized via effect sizes ($\eta_2 \geq 0.01$ = small, ≥ 0.06 = medium, ≥ 0.14 = large), ensuring findings reflected both statistical and practical relevance. This rigorous framework safeguarded against Type I errors while maintaining analytical precision in evaluating AI-human editing disparities.

3.6. Instruments

The study used several groups of tools. Popular AI algorithms such as Grammarly, Deeple, and ProWritingAid were used for automatic text editing. These models were chosen because of their popularity among users and their ease of use. The study used the Flesch-Kincaid and Gunning Fog metrics to assess the quality of editing, which determined the level of readability of texts. BLEU was also used to compare automatically corrected texts with reference ones, and F1 score was used to assess the accuracy of error detection and correction. The Flesch-Kincaid and Gunning Fog readability indices were calculated based on the average sentence length, the number of syllables in words, and the proportion of complex words in the text. The BLEU metric was determined by comparing n-grams (up to the 4^{th} order) of automatically edited texts with versions prepared by professional editors. F1 score was calculated as the harmonic mean of accuracy and completeness based on comparing corrected errors with the total number of existing errors. All metrics were applied using automated tools and provided an objective analysis of the quality of editing. The data obtained during the focus groups were recorded using audio recordings and transcribed as transcripts. The data analysis was performed using thematic analysis software such as NVivo and coding schemes for error classification. SPSS and Python packages with appropriate modules were used for statistical analysis. Tableau tools and Python modules such as Matplotlib and Seaborn were used for visualization of the results.

4. RESULTS

It is necessary to compare the technical characteristics of different AI models to fully understand the predicted differences between the results. Table 1 provides a comparison of the algorithms of the models used in the study. The table demonstrates the strengths of each tool depending on their main purpose, which can help in choosing the optimal solution for specific tasks.

Table 1: Technical Comparison Of Algorithms And Models Of Grammarly, Deepl, And ProWritingAid

Characteristics	Grammarly	DeepL	ProWritingAid
Model type	GPT (with NLP optimization)	Transformer (advanced translation)	GPT-3 with stylistic modifications
Main specialization	Automatic editing: grammar, style, tone	Machine Translation (MT) and Text Editing	Stylistics editing, text analysis
Grammar algorithms	Rule-based + context-aware neural networks	Syntactic analysis with translation focus	Linguistic analysis with NLP metrics
Stylistics module	Tone hints, rewriting recommendations	Limited (mainly semantics)	Deep style analysis, text tone selection
Types of errors it corrects	Grammar, spelling, style, punctuation	Grammatical, semantic (for translation)	Stylistic, spelling, semantic
Quality metrics (score)	F1-score: 0.92, BLEU: 0.87	BLEU: 0.94 (in translation), F1: 0.84	F1-score: 0.88, BLEU: 0.83
Context support	Advanced (target audience consideration)	Localization for different Languages	Medium (focus on authorial styles)
User tools	Multi-platform, tone tips	Translation localization, simple interface	Multi-component style analysis
Integration complexity	High (deep API integration)	Average (localized APIs)	Medium (settings for editors)
Processing speed	High	Very high	High
User-centricity	Multi-aspect editing, recommendations	Focus on translation accuracy	Suggestive editing, multi-style approach
Typical audience	Content writers, editors, students	Translators, business users	Authors who focus on style

Source: created by the authors of the research

So, *Grammarly* stands out with its comprehensive approach to text editing, offering grammatical, stylistic, and tonal cues. *DeepL* specializes in translation with high localization accuracy, but has limited stylistic analysis capabilities. ProWritingAid focuses on style and readability, offering tools for deep text analysis. Each tool is optimized for different tasks and audiences, which determines their effectiveness in specific scenarios. The next step was to evaluate the readability of texts edited by three AI algorithms and human editors (Figure 1).

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

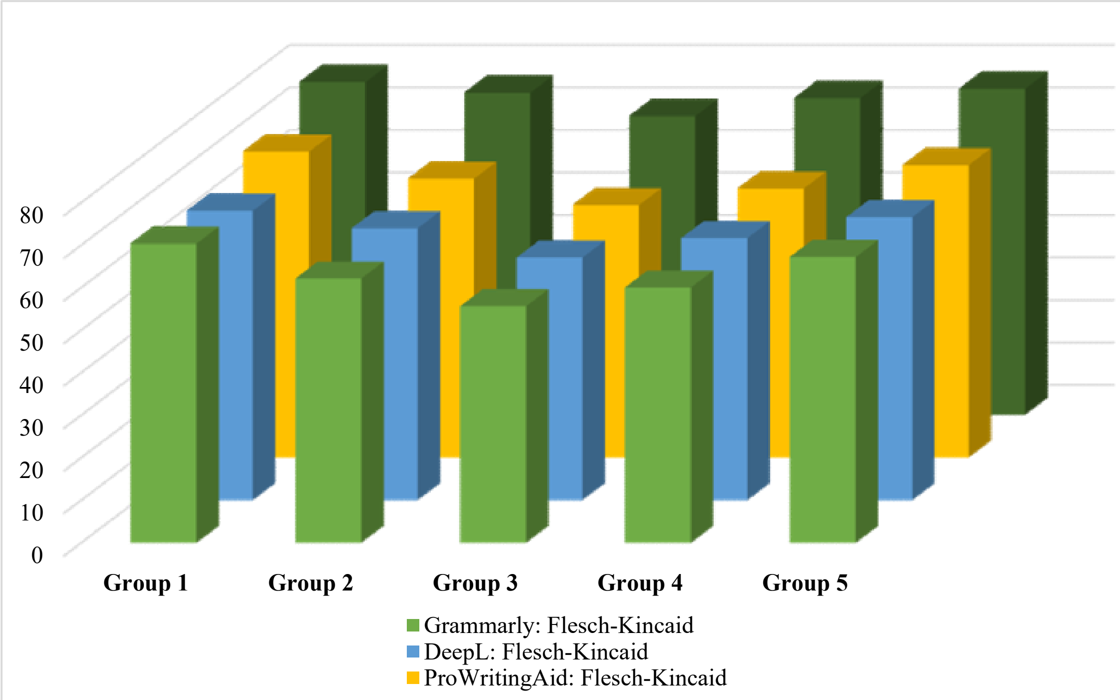

Figure 1:. Readability Assessment Of Texts Edited By Three AI Algorithms And Editors
Source: created by the authors of the research

The data analysis revealed the highest Flesch-Kincaid readability scores (70.4–78.4) for human-edited texts. The editors also showed the best Gunning Fog scores (8.7–10.5), indicating lower complexity of the texts they edited. AI algorithms had similar but higher Gunning Fog scores (9.5–12.5), indicating the need to optimize the elimination of complex syntactic structures. Overall, human editing provides a more natural and easy-to-understand language that is not yet achieved by AI algorithms. Table 2 presents the results obtained from the focus group discussions.

Table 2: Results Of Thematic Analysis Of Readability Of AI-Edited Texts By Focus Groups

Topic	Frequency of mentions	Examples of quotes
Convenience	15	"The algorithm is easy to use, does not require special knowledge."
		"The interface is clear, but some functions need to be configured."
Efficiency	20	"Grammar editing is at a high level, but the style is poor."
		"The algorithms correct spelling well, but ignore context."
Flexibility	10	"I would like to have more options for adjusting the tone of the text."
		"The system does not always adapt to the specifics of the text genre."
Impact on the Workflow	12	"The editing speed has increased significantly, but quality control is needed."
		"The time to complete the project has decreased, but additional refinement is needed."

Source: created by the authors of the research

Analysis of tabular data showed that focus group participants most often noted the effectiveness of algorithms (20 mentions). They emphasized their advantages in correcting grammatical and spelling errors, but noted problems with style and contextual understanding. Ease of use also received a positive assessment (15 mentions), but some functionality requires additional configuration. Flexibility and impact on the workflow were discussed less often (10 and 12 mentions, respectively), but were identified as important aspects for optimizing algorithms. The results of the ANOVA are presented in Table 3.

Journal of Theoretical and Applied Information Technology
30th November 2025. Vol.103. No.22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Results Of The (ANOVA)

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F-value	p-value	Eta-squared (η^2)
Between groups (Grammarly, DeepL and ProWritingAid)	325.4	3	108.47	12.56	< 0.001	0.45
Within groups	392.6	36	10.91	-	-	-
Total	718.0	39	-	-	-	-

Source: created by the authors of the research

The results of the analysis of variance (ANOVA) showed statistically significant differences between groups (Algorithms and Editors) in the quality of text editing, $F(3, 36) = 12.56$, $p < 0.001$. The Eta-squared value ($\eta^2 = 0.45$) indicates a large effect, indicating that group affiliation (algorithm or editors) explains 45% of the total variance in editing quality. The high sum of squares between groups (SS = 325.4) compared to the within-group (SS = 392.6) confirms significant differences between editing methods. The results indicate that the quality of work of editors and algorithms has significant differences, which may be important for choosing an appropriate editing tool.

A classification of errors that most often remain uncorrected or are created by modern text editing algorithms has been developed on the basis of the results and tables. The classification includes five main categories of errors (Figure 2).

Grammatical
- Inconsistency in the agreement of sentence members (up to 7%)
- Incorrect use of verb forms (up to 11%)

Spelling
- Homophones (up to 18%)
- Poor identification of proper nouns (up to 22%)

Styistic
- Oversimplification of sentences (up to 19%)
- Stereotyped phrases (up to 20%)
- Unnatural style (up to 17%)

Punctuation
- Contextual inconsistency (up to 23%)
- Ignoring punctuation in quotes (up to 27%)

Figure 2: Classification Of Typical Errors In The Operation Of Modern Automated Editing Algorithms
Source: created by the authors of the research

AI algorithms demonstrate high efficiency in detecting and correcting elementary grammatical and spelling errors. However, their capabilities are significantly limited when processing contextually dependent stylistic and semantic aberrations. The developed taxonomy of linguistic errors provides detailed information about the weaknesses of modern algorithms, which determines the directions of future research and development.

The results of the study confirmed the hypothesis of higher stylistic naturalness and readability of texts edited by humans. AI algorithms effectively corrected grammatical and spelling errors. However, their work with stylistics turned out to be less precise, as evidenced by qualitative focus group analysis. ANOVA revealed statistically significant differences between editing methods, emphasizing the advantages of editors in creating natural texts.

5. DISCUSSION

The findings of this study enter a growing academic conversation concerning the efficacy of AI-driven text editing. Our results are largely consistent with the body of prior work that acknowledges the proficiency of AI in grammatical and spelling correction while highlighting its shortcomings in stylistic and semantic tasks [2, 3, 16]. For instance, our data robustly supports the conclusions of Li et al. [21] regarding the superiority of human expertise in handling nuanced aberrations, and aligns with Heintz et al. [2] on the high accuracy of tools like Grammarly for grammatical errors.

The asserted superiority of human editors in stylistic and readability aspects [3, 6, 21] is empirically validated, moving beyond qualitative claims with quantitative evidence (e.g., Flesch-Kincaid: 75.2 vs. 69.4; η^2=0.45). While tools like Grammarly demonstrate grammatical proficiency [2, 22], their relative performance nuances against other AI tools and human editors are revealed by the comparative design, which also precisely categorizes their persistent weaknesses through a novel error taxonomy (Figure 2). This work thus consolidates fragmented prior knowledge into a unified, evidence-based framework, offering a benchmark for future development.

The obtained empirical data correlate with the conclusions of Li et al. [21], which emphasize the superiority of human expertise in correcting stylistic and semantic aberrations. At the same time, the use of tools such as Grammarly and DeepL has demonstrated high accuracy in identifying grammatical errors. This is consistent with the results of the research of Gupta et al. [22]. However, the obtained results are somewhat contradictory to the findings of Ara and Sik-Lanyi [23], which express skepticism about the ability of AI to correct complex multi-level errors.

A comparative analysis of the quality of text editing revealed a statistically significantly higher average Flesch-Kincaid index in professional editors (75.2) compared to AI algorithms (69.4). The obtained data are consistent with the findings of Erslev [24], which emphasize the readability advantage of texts edited by humans over automated systems. At the same time, the study by Shanthi et al. [25] indicates the ability of algorithms to achieve comparable results when processing standardized texts. The data demonstrated in the study of the team of researchers partially correlates with the results obtained within standardized syntactic structures. AI algorithms are better at editing texts with standard word order and simple grammatical rules.

The classification of detected linguistic aberrations showed that AI algorithms demonstrate the highest efficiency in correcting grammatical and spelling errors. However, correcting stylistic and semantic inaccuracies remains a significant problem for them. The obtained results are consistent with the data of Ma et al. [26], which show that modern generative models, in particular GPT-like architecture. AI models are mainly optimized for formal-grammatical correction, but often make errors in context-dependent stylistic decisions. At the same time, the research of Fiorillo and Mehta [27] expresses justified optimism about the potential of future iterations of algorithms to overcome these limitations by implementing more complex models.

User feedback showed a high assessment of the speed and ease of use of AI by editors and authors. However, they indicate the limitations of algorithms in complex editorial tasks. Similar conclusions are presented in the study of Algahtani [28], which emphasizes the irreplaceability of human expertise in the final stages of text processing. At the same time, Balluff et al. [29] suggest the potential superiority of AI over humans in certain scenarios. In particular, this concerns the editing of large arrays of text data containing typical and predictable errors.

Our study confirmed the conclusions of earlier studies and outlined key vectors for further improvement. The results of the thematic analysis of focus groups demonstrated the need for contextual and stylistic adaptation of algorithms. This need is also emphasized in the study of Sani [30]. It should be noted that the obtained empirical data emphasize the importance of synergistic AI integration with the human editing process. Such a symbiotic approach can be the foundation for further developments in this field.

A detailed classification of errors has been developed identifying the weaknesses of modern editing algorithms. Furthermore, the study deepens the understanding of the interaction of AI algorithms with the stylistic context of text materials. The practical significance is the possibility of using the results to optimize existing automatic editing tools, such as Grammarly, DeepL, and ProWritingAid. The results confirmed the research hypothesis, showing that AI algorithms are inferior to editors in ensuring stylistic naturalness and readability of texts. At the same time, the algorithms demonstrated high efficiency in correcting grammatical, spelling, and punctuation errors. The results of the study can be used by translation agencies to automate translation quality control. Editing platforms can integrate algorithms to reduce text processing time and increase efficiency. In the field of education (EdTech), the results will be useful for the development of teaching tools to improve written literacy. However, achieving stylistic naturalness of texts will remain a task that requires further research and improvement of algorithms.

5.1. Limitations

The study has some limitations that need to be considered when interpreting the results. The analysis was reduced to three AI algorithms, which

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

makes it difficult to extrapolate the findings to a wider range of tools. The study was conducted on English texts only, which does not allow for assessing their effectiveness for other linguistic systems. The sample of texts covered five genres, which potentially did not take into account the specifics of rarely used styles and formats. Fiction texts, legal documents, medical reports, user instructions, and texts on social networks could be included for further analysis. The subjectivity of focus group participants' assessments could have affected the interpretation of qualitative data. The size of the sample of texts and the editors involved creates some limitations for scaling the obtained findings.

Despite its contributions, this study is not without limitations, which should be considered when interpreting the results and present avenues for future research. Firstly, the selection was limited to three AI algorithms (Grammarly, DeepL, ProWritingAid), which, while popular, does not encompass the entire ecosystem of NLP tools. Emerging or specialized models might yield different results. Secondly, the study focused exclusively on English texts. The performance of these tools, particularly in stylistic adaptation, may vary significantly in languages with different syntactic structures or morphological complexity, a noted challenge in NLP. Thirdly, while the text genres were diverse, they did not include highly specialized domains like legal or medical writing, where terminology and phrasing are rigidly defined and AI performance could potentially differ. Lastly, the subjective element in the focus group assessments, though valuable, introduces a qualitative variable that is difficult to fully standardize. These limitations underscore that our findings are a robust snapshot of the current state of AI editing within a specific context, rather than a definitive, universal conclusion. They highlight the need for replication across languages, a broader array of AI tools, and more specialized genres.

5.2. Recommendations

The following recommendations are offered for improving the editing and proofreading of texts using AI tools:

1. Personalization of editing – adaptation of AI to the individual style of the author by training on his or her texts to preserve stylistic features.

2. Interactive editing – creation of interfaces for joint work of a person and AI, with the possibility of controlling changes and training the model based on feedback.

3. Analysis of complex errors – improvement of models for detecting errors in a broader context, taking into account semantic and stylistic connections; development of algorithms capable of an integrated analysis of the relationship between different types of errors at the sentence and text levels.

4. Implementation of personalized editing involves fine-tuning of modern NLP models (for example, GPT, BERT) on the corpus of texts of a particular author to preserve idiostyle. The use of vector models (Word2Vec, Doc2Vec) is also recommended to identify individual lexical and syntactic features. The user may be given the option to select style templates or initiate custom model training.

5. The development of interactive interfaces involves the creation of an API with WebSockets support for real-time editing, the implementation of a versioning system (like git) with diff algorithms for tracking changes, as well as the implementation of Active Learning — the ability to mark AI edits as acceptable/unacceptable for further training of the model.

6. The mechanisms should be developed to detect and correct complex multi-component errors. From a technical perspective, transformer models with contextual attention mechanisms can be used for this purpose. In particular, such models as T5 or BART are able to analyse semantic relationships between words within a paragraph. The implementation of multimodal checking, combining the analysis of grammar, style and logic, can be implemented using ensemble models. It is also important to carry out contextual and stylistic analysis in order to check the correspondence of the tone of the text to a given style (formal, scientific, journalistic).

6. CONCLUSIONS

The relevance of the presented results is determined by their reflection of the current state of development of AI algorithms in the field of automated editing of English texts. In the academic discourse, the obtained data contribute to a deeper understanding of the potential of technologies to replace or integrate with human expertise in ensuring high-quality editing. Identification of strong and weak attributes of various algorithmic solutions provides the possibility of targeted optimization of existing systems and increasing their applied efficiency. The readability analysis of edited texts demonstrated a statistically significant lower performance for AI algorithms compared to human editors. The average value of the Flesch-Kincaid index was 75.2 for human editing, while this indicator reached only 69.4for the most effective algorithm (ProWritingAid). The results of the thematic analysis of focus groups revealed a positive perception of algorithms in terms of convenience (15 mentions) and processing speed (20 mentions). At the same time, users noted the lack of flexibility of automated systems and the problems with stylistic correctness. Analysis of variance confirmed statistically significant differences in the quality of editing between the studied groups ($F=12.56$, $p<0.001$) with a significant effect size ($\eta^2=0.45$). This indicates a significant advantage of professional editors over the three considered AI algorithms. The algorithms Grammarly, DeepL, and ProWritingAid showed similar results, but Grammarly and ProWritingAid slightly outperformed DeepL, especially for complex texts. The results of the study can be used to improve AI algorithms, develop educational programmes for editors, and optimize automatic text editing tools. The work will also be interesting for teachers of information technology in linguistics and specialists in automatic translation when creating educational programmes. Further research should focus on developing algorithms that better account for stylistic context and the specifics of different text genres. It may also be promising to study the impact of hybrid approaches that combine AI and human editing on the NLP efficiency.

REFERENCES:

[1] Y. Zhang, "Natural scene text editing based on AI", arXiv preprint arXiv:2111.15475, 2021. doi: 10.48550/arXiv.2111.15475.

[2] K. Heintz, Y. Roh, and J. Lee, "Comparing the accuracy and effectiveness of Wordvice AI Proofreader to two automated editing tools and human editors", *Science Editing*, Vol. 9, No. 1, 2022, pp. 37-45. doi: 10.6087/kcse.261.

[3] O. Bulavina, D. Kyslenko, V. Hmil-Chuprina, S. Artemenko, and L. Cherednyk, "Efficiency of virtual reality technologies in the development of strategic thinking of future professionals", *Journal of Theoretical and Applied Information Technology,* Vol. 102, No. 18, 2024, pp. 6750-6760. Available in: https://www.jatit.org/volumes/Vol102No18/15Vol102No18.pdf (15.06.2025).

[4] M. Abinaya, and G. Vadivu, "AI tools for efficient writing and editing in academic research", in: *Utilizing AI Tools in Academic Research Writing*. Pennsylvania: IGI Global. 2024, pp. 141-157. doi: 10.4018/979-8-3693-1798-3.ch009

[5] N. Curtis, "To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing", *The Pediatric Infectious Disease Journal*, Vol. 42, No. 4, 2023, pp 275. doi: 10.1097/INF.0000000000003852

[6] M. Hosseini, L. M. Rasmussen, and D. B. Resnik, "Using AI to write scholarly publications", *Accountability in Research*, Vol. 31, No. 7, 2024, pp. 715-723. doi: 10.1080/08989621.2023.2168535.

[7] N. Lin, M. Zeng, W. Huang, S. Jiang, L. Xiao, and A. Yang, "A simple yet effective corpus construction framework for Indonesian grammatical error correction", *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 24, No. 1, 2025, pp. 1-15.

[8] S. Lin, "Evaluating LLMs' grammatical error correction performance in learner Chinese", *PloS One*, Vol. 19, No. 10, 2024, art. e0312881. doi: 10.1371/journal.pone.0312881.

[9] J. Mallinson, J. Adamek, E. Malmi, and A. Severyn, "EdiT5: Semi-autoregressive text editing with T5 warm-start", in: Y. Goldberg, Z. Kozareva, and Y. Zhang (eds.). *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022.* Mumbai: Association for Computational Linguistics, 2022, pp. 2126-2138. doi: 10.18653/v1/2022.findings-emnlp.156.

[10] H. Jiang, "Developing an artificial intelligence model for English grammar correction: A computational linguistics approach", *Journal of Computational Methods in Sciences and Engineering*, Vol. 24, No. 4, 2025. doi: 10.1177/14727978251318799.

[11] A. Kadim, and A. Lazrek, "Nemlar corpus improvement for Arabic natural language processing", *Journal of Theoretical and Applied Information Technology,* Vol. 103, No. 4, 2025, pp. 1320-1327. Available in: https://www.jatit.org/volumes/Vol103No4/14Vol103No4.pdf (15.06.2025).

[12] Y. Wang, Y. Wang, K. Dang, and J. Liu, "A comprehensive survey of grammatical error correction", *ACM Transactions on Intelligent Systems and Technology*, Vol. 12, No. 5, 2021, pp. 1-51. doi: 10.1145/3474840.

[13] R. Raju, P. B. Pati, S. A. Gandheesh, G. S. Sannala, and K. S. Suriya, "Grammatical versus spelling error correction: An investigation into

the responsiveness of transformer-based language models using BART and MarianMT", *Journal of Information & Knowledge Management*, Vol. 23, No. 03, 2024, art. 2450037.

[14] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI", *Business & Information Systems Engineering*, Vol. 66, No. 1, 2024, pp. 111-126. doi: 10.1007/s12599-023-00834-7.

[15] R. Hardack, "It's always already too late for white men: Personified nature and corporate personhood, from moby-dick to avatar", *Angelaki*, Vol. 29, No. 6, 2024, pp. 152-190. doi: 10.1080/0969725X.2024.2430908.

[16] T. Chakrabarty, P. Laban, and C.-S. Wu, "Can AI writing be salvaged? Mitigating idiosyncrasies and improving human-AI alignment in the writing process through edits", arXiv, 2024. doi: 10.48550/arXiv.2409.14509.

[17] A.-M. M. Gasaymeh, M. A. Beirat, and A. A. Abu Qbeita, "University students' insights of generative artificial intelligence (AI) writing tools", *Education Sciences*, Vol. 14, No. 10, 2024, art. 1062. doi: 10.3390/educsci14101062.

[18] S. Mathôt, and J. March, "Conducting linguistic experiments online with OpenSesame and OSWeb", *Language Learning*, Vol. 72, No. 4, 2022, pp. 1017-1048. doi: 10.1111/lang.12509.

[19] R. Kircher, and L. Zipp, (eds.). "Research methods in language attitudes". Cambridge: Cambridge University Press, 2022. Available in: https://surl.li/cidklo (15.06.2025).

[20] L. Litosseliti, (ed.). "Research methods in linguistics", London: Bloomsbury Publishing, 2024. Available in: https://surl.li/fyexay (15.06.2025).

[21] M. Li, J. Sun, and X. Tan, "Evaluating the effectiveness of large language models in abstract screening: A comparative analysis", *Systematic Reviews*, Vol. 13, No. 1, 2024, art. 219. doi: 10.1186/s13643-024-02609-x.

[22] A. Gupta, D. Sajnani, and G. Anumanchipalli, "A unified framework for model editing", arXiv preprint arXiv:2403.14236, 2024. doi: 10.48550/arXiv.2403.14236.

[23] J. Ara, and C. Sik-Lanyi, "Automated evaluation of accessibility issues of webpage content: Tool and evaluation", *Scientific Reports*, Vol. 15, No. 1, 2025, art. 9516. doi: 10.1038/s41598-025-92192-5

[24] M. S. Erslev, "A poetics of misrepresentation: The mimesis of machine learning in ReRites", in: *The Routledge Companion to Literary Media*. London: Routledge, 2023, pp. 197-208. Available in: https://surl.li/lhtacd (15.06.2025).

[25] R. Shanthi, M. Mythili, R. D. Gomathi, S. Maheswaran, S. Sathesh, G. Murugesan, P. Duraisamy, S. Pavithra, and A. S. Haryni, "AI-driven education: An empirical study on evaluating LSRW acquisition through active involvement and learning", in: *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Kamand: IEEE, 2024, pp. 1-7. doi: 10.1109/ICCCNT61001.2024.10726048.

[26] Z. Ma, G. Jia, and B. Zhou, "Adapedit: Spatio-temporal guided adaptive editing algorithm for text-based continuity-sensitive image editing", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, No. 5, 2024, pp. 4154-4161. doi: 10.1609/aaai.v38i5.28210.

[27] L. Fiorillo, and V. Mehta, "Accelerating editorial processes in scientific journals: Leveraging AI for rapid manuscript review", *Oral Oncology Reports*, Vol. 10, 2024, art. 100511. doi: 10.1016/j.oor.2024.100511.

[28] A. M. I. R. A. H. Algahtani, "A comparative study of ai-based educational tools: Evaluating user interface experience and educational impact", *Journal of Theoretical and Applied Information Technology*, Vol. 102, No. 5, 2024, pp. 1746-1758. Available in: https://www.jatit.org/volumes/Vol102No5/7Vol102No5.pdf (15.06.2025).

[29] P. Balluff, H. G. Boomgaarden, and A. Waldherr, "Automatically finding actors in texts: A performance review of multilingual named entity recognition tools", *Communication Methods and Measures*, Vol. 18, No. 4, 2024, pp. 371-389. doi: 10.1080/19312458.2024.2324789.

[30] M. A. Sani, "Contextual understanding and generation: The role of large language models in modern AI", *Journal of Innovative Technologies*, Vol. 7, No. 1, 2024, pp. 1-7. Available in: https://acadexpinnara.com/index.php/JIT/article/view/223 (15.06.2025).