Lexical diversity, syntactic complexity, and readability: a corpus-based analysis of ChatGPT and L2 student essays

Daniel R. Fredrick* and Laurence Craven*

Department of English, American University of Sharjah, Sharjah, United Arab Emirates

This study compares AI-generated texts (via ChatGPT) and student-written essays in terms of lexical diversity, syntactic complexity, and readability. Grounded in Communication Theory—especially Grice's Cooperative Principle and Relevance Theory—the research investigates how well AI-generated content aligns with human norms of cooperative communication. Using a corpus of 50 student essays and 50 AI-generated texts, the study applies measures such as Type-Token Ratio (TTR), Mean Length of T-Unit (MLT), and readability indices like Flesch–Kincaid and Gunning-Fog. Results indicate that while ChatGPT produces texts with greater lexical diversity and syntactic complexity, its output tends to be less readable and often falls short in communicative appropriateness. These findings carry important implications for educators seeking to integrate AI tools into writing instruction, particularly for second-language (L2) learners. The study concludes by calling for improvements to AI systems that would better balance linguistic complexity with clarity and accessibility.

KEYWORDS

AI-generated writing, linguistic complexity, L2, Grice's cooperative principle, readability analysis

1 Introduction

Tools like ChatGPT have significantly reshaped how students draft and complete essays. Powered by large language models (LLMs), ChatGPT has transformed how students write by giving them unprecedented access to advanced language support (De Angelis et al., 2023). While these tools provide substantial practical benefits, their growing use raises important concerns about the quality, authenticity, and communicative effectiveness of student writing. A key question in this study is whether AI-generated texts can achieve the standards of human communication. This study aims to address that question by comparing the linguistic complexity of student-written essays with texts generated by ChatGPT. Drawing on Communication Theory particularly Grice's Cooperative Principle and Relevance Theory (Grice, 1975, 1989; Sperber and Wilson, 1986, 1995)—the analysis evaluates how well AI-produced texts conform to the core principles of effective communication. Although AI-generated content often displays greater lexical variety and syntactic complexity, it frequently falls short in terms of readability and overall communicative effectiveness (Fredrick et al., 2024). This mismatch highlights the importance of better understanding the extent to which AI tools align with the norms of academic communication. To address this issue, this study uses linguistic complexity measures, including Type-Token Ratio (TTR), Mean Length of T-Unit (MLT), and readability indices like the Flesch–Kincaid Grade Level and Gunning-Fog Index. By comparing these features in student-written and AI-generated essays, the study aims to clarify both the strengths and limitations of AI tools in fostering meaningful communication

skills, particularly among second-language (L2) learners. The findings contribute to ongoing discussions about AI in education and offer practical guidance for integrating these technologies into writing instruction. Grounded in Communication Theory, the research goes beyond technical evaluation to assess how well AI-generated texts align with the broader standards of effective human communication. Grice's Cooperative Principle and Relevance Theory (Grice, 1975, 1989; Sperber and Wilson, 1986, 1995) form the primary theoretical foundation for this study. Grice's framework stresses the importance of clarity, relevance, and appropriateness in communication, suggesting that effective writing must align with the needs and expectations of its intended audience. Likewise, Relevance Theory (Sperber and Wilson, 1995) emphasizes the balance between linguistic sophistication and cognitive effort, focusing on how readers process information efficiently. Using these frameworks, the study evaluates whether AI-generated texts conform to these principles or impose unnecessary cognitive demands on the reader. Although such texts often feature advanced vocabulary and complex syntax, these elements may hinder readability—making them less suitable for instructional settings where comprehension and engagement are critical. This study also considers its broader implications for educators and policymakers. Teachers play a vital role in helping students understand the importance of balancing complexity with communicative purpose, ensuring that clarity and accessibility are not sacrificed for stylistic sophistication. Additionally, the study underscores the value of interdisciplinary collaboration in refining AI systems to meet the communicative needs of diverse populations, including L2 learners, individuals with disabilities, and non-native speakers. By addressing the gap between AI-generated language and human communication standards, this research supports the design of tools that do more than mimic human output—they foster meaningful communication in learning environments.

2 Literature review

Nothing connects the current debate about AI in education more clearly to language pedagogy than the question of how machine-generated text compares to student writing. As generative AI becomes more integrated into academic settings, scholars have begun to systematically examine these differences, particularly in the context of English as a Foreign Language (EFL) learning. Of all recent studies, the most striking is Mizumoto et al. (2024) investigation of Japanese university students, which reveals the stark contrast between human and AI-authored texts. Using linguistic fingerprinting and machine learning to analyze 263 essays, the authors found that ChatGPT-generated texts exhibited significantly greater lexical diversity, higher syntactic complexity, more nominalizations, and fewer grammatical errors than student essays. At the same time, students relied more on modals, epistemic markers, and discourse signals—linguistic elements that reflect personal voice and uncertainty, often central to the development of critical thinking in L2 writing. Indeed, this reflects another layer of complexity in their findings: while AI performs with grammatical and lexical sophistication, it lacks the rhetorical tentativeness that characterizes learner engagement.

Herbold et al. (2023) offer valuable insight into this same issue, extending the inquiry to large-scale comparisons between German high school students and multiple versions of ChatGPT. Their

research reinforces the view that AI models consistently outperform human writers in language mastery and structural clarity. Herbold et al. (2023) suggests reinventing homework, especially considering that AI-generated essays routinely outscore human ones across standardized rubrics. Yet, this raises a serious pedagogical dilemma: if students can rely on AI for grammatically superior output, educators must reconsider whether surface correctness should remain a primary learning outcome. Based on these examples, it is possible to suggest that writing instruction must shift from product-oriented assessment to process-oriented guidance that emphasizes idea development, voice, and rhetorical awareness.

Yildiz Durak et al. (2025) provide one of the most methodologically detailed examinations of this issue. Their comparative study of student writing versus texts generated by ChatGPT, Gemini, and BingAI revealed that human-written texts featured greater lexical diversity, longer sentences, and more conceptual range. By contrast, AI-generated texts—particularly from Gemini—tended to overuse stopwords and repeated phrases, producing formulaic patterns that diminished their rhetorical depth. While AI texts could sometimes mimic surface-level fluency, the authors found that syntactic and semantic markers such as unique word count, conjunction usage, and topical variety helped distinguish machine output from student writing. Interestingly, ChatGPT-generated texts proved the hardest to distinguish from student work, highlighting the increasing sophistication of AI models but also complicating assessment practices. Their findings echo Mizumoto et al. (2024) in showing that high fluency does not necessarily translate into critical engagement or authentic voice. Indeed, both studies suggest that while AI can imitate the outward form of academic prose, it often lacks the internal richness and variation that characterize student writing rooted in lived experience and individual cognition. This distinction between form and depth is also explored in Fredrick et al. (2024), who found that experienced instructors identified AI writing not by surface correctness, but by its lack of subjective detail, contextual anchoring, and rhetorical spontaneity.

Barrot (2023) brings attention to both the possibilities and risks of using ChatGPT in writing instruction. While ChatGPT provides adaptive feedback, support for brainstorming, and writing assistance that may benefit L2 learners in real-time, Barrot warns that overreliance on AI risks diminishing students' originality, critical thinking, and writing voice. Barrot is right to claim that AI can support the writing process. He argues instead for clear guidelines and scaffolding so that AI becomes a tool—not a replacement—for responsible composition. Understanding Barrot's framework is key to grasping the larger significance of AI in second language writing: it is not the presence of AI that threatens pedagogy, but the absence of reflective practices around it. Still, Barrot does not fully address the unusual case of teacher misidentification of AI-generated essays. This issue is sharply illuminated in recent findings from Fleckenstein et al. (2024), who conducted two experimental studies involving novice and experienced teachers. Surprisingly, both groups struggled to distinguish AI-generated texts from student-authored ones and were frequently overconfident in their (often inaccurate) judgments. The implications are clear: essay evaluation cannot depend solely on intuition or superficial traits. As Fleckenstein et al. argue, the educational landscape must now reckon with the blurred boundaries

between genuine and generated writing. Their study underscores the need for systemic strategies—beyond teacher judgment alone—to assess authenticity.

Despite its grammatical polish and surface-level fluency, ChatGPT's language output often masks a deeper failure to meet the expectations of human communication. Recent cognitive-pragmatic research reveals that beneath its well-formed sentences lies a consistent inability to adhere to Grice's Cooperative Principle—particularly the conversational maxims of quantity, relation, and manner. Attanasio et al. (2024) argue that ChatGPT frequently produces responses that are verbose, tangential, or ambiguous. This overelaboration, they suggest, results in a discourse style that mirrors the speech patterns of individuals with high-functioning autism, where linguistic accuracy is present but pragmatic nuance is lacking. Barattieri di San Pietro et al. (2023) confirm this trend, noting that ChatGPT often supplies excessive, unsolicited information, creating exchanges that compromise conversational relevance and efficiency. Rather than promoting clarity, such verbosity appears to reflect algorithmic overcompensation, undermining the concise and targeted nature of effective human dialog.

Adding a cognitive dimension, Cong (2024) demonstrates that GPT models regularly fail to interpret manner implicatures—subtle communicative cues that hinge on indirectness, tone, and social context. These limitations are not only linguistic but also functional. Wölfel et al. (2024) argue that for AI to serve as trustworthy educational agents, it must adhere to Gricean norms—not just to mimic human conversation, but to build clarity, trust, and pedagogical coherence. Their comparative study of knowledge-based and generative AI systems finds that large language models often violate the maxims of quantity and quality, generating output that is either too verbose or insufficiently reliable. Notably, the authors propose a "Maxim of Trust" to supplement Grice's original framework, highlighting how user confidence in AI responses depends on more than just correctness—it hinges on the perceived honesty, relevance, and efficiency of the exchange. These observations are further underscored by findings from the PUB benchmark study (Settaluri et al., 2024), which explicitly frames these deficiencies within the legacy of Grice's work. Although large language models excel in syntax and semantics, they consistently struggle with implicature, presupposition, reference, and deixis—domains at the core of Gricean pragmatics.

Taken together, these studies converge on a crucial insight: ChatGPT's rhetorical output, while linguistically refined, lacks the contextual awareness and inferential economy that define human communicative competence. Its violations of Gricean norms are not occasional glitches, but structural consequences of being trained to simulate form without fully grasping intention. This persistent shortfall marks the boundary between language generation and genuine understanding. AI's strengths in grammar and fluency often obscure its lack of argumentative depth, rhetorical strategies, and emotional appeals. Sympathetic critics such as Mizumoto et al. (2024), Barrot (2023), Fleckenstein et al. (2024), Yildiz Durak et al. (2025), and Attanasio et al. (2024) have all argued that clarity around the ethical and pedagogical use of AI is urgently needed—not to prohibit its use, but to ensure it supports rather than erodes educational goals. Thus, educators must not only prepare students to recognize linguistic excellence but to assert their own ideas,

interpretive positions, and ethical voices in an age of machine fluency.

Although there are many studies of AI compared to student produced texts, there is a gap in the literature regarding the measures used to measure writing complexity used in the present study. As far as the authors are aware no study has looked at these key markers of complexity. The study thus attempts to answer the following research questions:

1. How does the lexical diversity and syntactic complexity of AI-generated texts (ChatGPT) compare to student-authored essays in terms of linguistic richness and structural sophistication?
2. To what extent do AI-generated texts align with principles of effective communication, particularly in terms of readability, relevance, and accessibility, as outlined by Grice's Cooperative Principle and Relevance Theory?
3. What are the pedagogical and ethical implications of integrating AI tools like ChatGPT into academic writing instruction, particularly for second-language (L2) learners and diverse student populations?

3 Methodology

This study used a corpus-based approach to examine the linguistic complexity of student essays compared to content generated by ChatGPT. Corpus-based methods are widely recognized for their empirical rigor and systematic analysis, allowing researchers to explore large collections of texts with detailed linguistic tools (McEnery and Hardie, 2011; Granger, 2021). The dataset consisted of two matched sets of 50 full-length essays—one written by students and the other produced by ChatGPT. To maintain consistency and control for variation, all student essays were written in class in a monitored environment to mitigate the chance of the students using AI to write their essays. The students were given 10 min planning time and 30 min writing time. Their response was to identical prompts, ensuring similar topic coverage, structural expectations, and overall length (Biber et al., 1998). ChatGPT was given the same essay prompt as the students and instructed to produce responses with a word count matching the student average word count. The student essays were written by university students enrolled in an academic writing program in the United Arab Emirates and were written by non-native English speakers, mainly with Arabic and Urdu as their first languages. Ethical guidelines were strictly followed by anonymizing all submissions and obtaining informed consent from participants. The AI-generated texts were produced using OpenAI's GPT-4 Turbo, the version of GPT-4 provided to ChatGPT Plus subscribers via the web interface at the time of data collection. All texts were generated in December 2024, with identical prompts to those assigned to students. Each prompt was submitted five times, and average scores were calculated across all linguistic complexity measures to reduce variability and capture representative model performance.

To analyze differences between student and AI-generated writing, a range of linguistic complexity metrics were applied, focusing on lexical diversity, syntactic complexity, readability, and lexical sophistication. These measures were chosen based on their relevance to

academic writing and their established use in previous corpus-based studies.

3.1 Lexical diversity

- Type-Token Ratio (TTR): Assesses the ratio of unique words to total words, indicating vocabulary richness.
- Measure of Textual Lexical Diversity (MTLD): A more robust measure that accounts for text length in evaluating lexical variety (Malvern et al., 2004).
- Vocabulary Diversity (Voc-D): A metric assessing the breadth of vocabulary used in a text.

3.2 Syntactic complexity

- Mean Length of T-Unit (MLT): Measures the average length of independent clauses, providing insight into sentence complexity.
- Dependent Clauses per T-Unit (DC/T): Evaluates the frequency of subordinate clauses, which contribute to syntactic elaboration (Lu, 2010).

3.3 Readability

- Flesch–Kincaid Grade Level: Determines how easy or difficult a text is to read based on sentence and word length.
- Gunning-Fog Index: Calculates the number of years of formal education required to understand the text (Benjamin, 2012).

3.4 Lexical sophistication

- Lexical Sophistication (LS): Measures the proportion of less common, advanced vocabulary used in a text.

3.5 Data processing and statistical analysis

All texts were analyzed using computational linguistic tools such as Coh-Metrix and Text Inspector, and Lu's (2010) Syntactic Complexity Analyzer which provide precise calculations for the selected complexity measures.

By employing this methodology, the study systematically assesses linguistic complexity differences between AI-generated and student-authored texts, shedding light on the strengths and weaknesses of AI as a writing tool in academic contexts.

4 Results

This chapter is an in-depth analysis of linguistic complexity measures applied to 50 essays written by students and 50 essays generated by the ChatGPT. The research spotlights some key measures, including Type-Token Ratio (TTR), Mean Length of T-Unit (MLT), Dependent Clauses per T-Unit (DC/T), Flesch–Kincaid Grade Level, Gunning-Fog Index, Measure of Textual Lexical Diversity (MTLD), Vocabulary Diversity (Voc-D), and Lexical Sophistication (LS). These measures provide us with information on lexical diversity, syntactic complexity, readability, and vocabulary sophistication in human-written and machine-written texts.

4.1 Lexical diversity

One notable finding relates to lexical diversity, which refers to the range and richness of vocabulary used in a text, is that ChatGPT-composed texts further boasted a higher Type-Token Ratio (TTR) of 0.69 compared to student essays, which scored 0.61. This indicates that AI-composed texts boast more lexical diversity relative to overall word count, an indication of its ability to employ vast datasets and form novel combinations of words. ChatGPT further boasted a much higher Measure of Textual Lexical Diversity (MTLD) score of 118 compared to 66.56 from student essays. In terms of vocabulary diversity (Voc-D), ChatGPT scored significantly higher at 105, compared to 69.73 for student-written texts. These findings point to ChatGPT's ability to generate text that is lexically dense, with a very broad repertoire of vocabulary that tends to exceed human levels. Such conspicuous strength is purchased at significant trade-offs, however. Enhanced lexical diversity in AI-generated text raises issues of contextual appropriateness and communicative ease. For instance, while ChatGPT shows aptitude in exposing varied vocabulary, it might lack a more sophisticated understanding required for tailoring language choices to specific audiences or purposes, potentially compromising the effectiveness of communication overall. The analysis of syntactic complexity reveals distinct patterns in AI-written and human-written texts. Student essays also manifested greater complexity in sentence structures, as determined by an MLT of 23.61, whereas ChatGPT had an MLT of 15.91. This result suggests that students tend to write longer, potentially more complex sentences, reflecting an effort to elaborate their ideas more thoroughly. ChatGPT, however, had a more favorable DC/T ratio of 0.75 versus 0.57 for student essays. This implies that AI-generated texts have a higher number of dependent clauses, with syntactically complicated sentences with subordinate clauses as a usual result. Although this feature enhances the perceived complexity of ChatGPT's responses, it also costs readers greater cognitive efforts, rendering the text more difficult to process. The differences in syntactic structure indicate the special strategies that humans and AI use in the structuring of written communication.

4.2 Readability scores

Readability measures show dramatic contrasts between AI and student writing, indicating the trade-off between linguistic complexity and readability. ChatGPT responses were more difficult to read by a wide margin, with a Flesch–Kincaid Grade Level of 16.61 compared to 12.12 for student essays. Similarly, the Gunning-Fog Index also

TABLE 1 Summary of the results.

Measure	ChatGPT essays	Student essays	Interpretation
Lexical diversity			
Type-token ratio (TTR)	0.69	0.61	AI exhibits greater lexical diversity.
Measure of textual lexical diversity (MTLD)	118	66.56	AI-generated texts have a more varied vocabulary.
Vocabulary diversity (Voc-D)	105	69.73	AI uses a broader range of words than students.
Syntactic complexity			
Mean length of T-Unit (MLT)	15.91	23.61	Students write longer and more elaborate sentences.
Dependent clauses per T-unit (DC/T)	0.75	0.57	AI tends to use more complex subordinate clauses.
Readability scores			
Flesch–Kincaid grade level	16.61	12.12	AI-generated texts are more difficult to read.
Gunning-Fog index	20.10	14.18	AI responses require higher literacy to comprehend.
Lexical sophistication			
Lexical sophistication (LS)	0.41	0.35	AI uses more sophisticated and uncommon words.

found ChatGPT responses to be more complex, with a score of 20.10 compared to 14.18 for student essays.

The results demonstrate that AI-generated texts impose a higher cognitive burden on readers, necessitating higher literacy and understanding of more complex linguistic constructions. The heightened complexity level of texts generated by ChatGPT presents obstacles in the pedagogical environment, where accessibility and clarity are paramount. For example, students who depend on AI instruments for writing may end up being unable to revise and polish the resulting texts because the initial production might be too heavy or cumbersome for everyday use.

4.3 Lexical sophistication

The lexical sophistication analysis highlights both the weaknesses and strengths of texts generated by artificial intelligence. ChatGPT scored 0.41 for Lexical Sophistication (LS), which is higher than the 0.35 score found in essays written by students. This reveals ChatGPT's preference for using more sophisticated or less common words, reflecting its extensive training data and more developed linguistic capacity. However, this increased sophistication may come at the expense of decreased accessibility, particularly for readers not familiar with sophisticated terms.

For example, L2 students or readers with limited experience with academic language may struggle to comprehend AI-generated texts, undermining their usefulness in classroom settings. The emphasis on lexical complexity demands innovation in AI design, such that models can adapt to varying levels of reader proficiency without sacrificing inclusivity (Table 1).

5 Discussion

The results highlight some of the main differences between ChatGPT and student writing. ChatGPT outperforms student essays on measures of lexical diversity (TTR, MTLD, Voc-D) and syntactic complexity (DC/T), pointing to its ability for linguistically dense output. These strengths are counterbalanced by lower readability scores and reduced communicative relevance, as indexed by higher Flesch–Kincaid and Gunning-Fog scores. At first glance, the syntactic results seem to point in opposite directions. Student essays showed a higher Mean Length of T-unit (MLT: 23.61 vs. 15.91), while ChatGPT essays scored higher on Dependent Clauses per T-unit (DC/T: 0.75 vs. 0.57). Yet these metrics capture different aspects of syntactic complexity. Students' longer T-units often reflected features typical of developing L2 writing: run-on sentences, weak punctuation control, and heavy reliance on coordination with markers like *and*, *but*, or *so*. These patterns inflate sentence length without adding structural depth. ChatGPT, by contrast, produced shorter T-units but used subordination more consistently, yielding denser hierarchical structures. The higher DC/T ratio indicates that AI-generated sentences more often contained embedded relative, adverbial, or complement clauses—hallmarks of academic prose. In comparison, many student texts stretched length through coordination and linear chaining rather than true layering. The broader lesson here is that syntactic complexity is multidimensional: length-driven elaboration differs fundamentally from clause-embedding, and neither greater length nor greater subordination alone ensures readability or sophistication.

Additionally, while ChatGPT demonstrates greater lexical sophistication, this may reduce accessibility to some readers. The results of this study have significant implications for the understanding of the contrast between AI-generated texts and human writing in terms of linguistic sophistication and communicative efficacy in academic contexts. By alleviating the trade-offs found, educators and designers can develop AI tools that reconcile sophistication with simplicity and ease of use, and hence enhance the support of learning goals. These differences cut across all dimensions, from lexical richness to syntactic complexity, readability, and communicative adequacy. Even though ChatGPT performs better in lexical diversity and syntactic complexity, student essays outshine AI-texts consistently in readability and communicative success. These results have significant implications for our current understanding of how AI applications align with— or diverge from—conventions of human communication, particularly in the classroom. ChatGPT is capable of producing lexically diverse and syntactically intricate texts.

For example, the results indicate increased Type-Token Ratios (TTR) and more complex sentence structures, typically including dependent clauses and higher-level vocabulary (Kalyan et al., 2021). These features characterize the model's capacity to take huge datasets and generate new combinations of words, thus demonstrating its technical proficiency as a tool for creating language. However, there are some downsides to these features. The excessive utilization of complicated vocabulary and complicated sentence structures tends to detract from the readability of texts composed with the aid of ChatGPT, rendering them less appropriate for academic environments where readability and relevance are paramount. This inclination is in line with Grice's Cooperative Principle, which calls for a maxim of clarity, relevance, and appropriateness in communication (Grice, 1975). From a theoretical point of view, the verbosity of ChatGPT and its inclination toward complicatedness often violate Grice's Maxims of Quantity and Relevance. For instance, the use of highly complicated terminology has the potential to confuse readers, thereby undermining the communicative intention of the text (Brennan, 1991). Likewise, in the context of Relevance Theory, the heightened cognitive effort exacted by AI-authored texts can undermine their utility, especially for students learning basic writing competencies (Sperber and Wilson, 1986). By contrast, essays written by students exhibit better readability and communicative relevance. These writings prioritize clarity and coherence, enabling the communication of ideas efficiently without imposing unnecessary cognitive loads on the readers. For instance, while the writing of students may not match the syntactic level of AI-generated texts, it often demonstrates a better sense of audience need and rhetorical intention. This strain between simplicity and complication underscores the value of human-authored texts in academic settings, where the cultivation of critical thinking and clear communication is paramount.

Furthermore, student writing will also exhibit a sense of purpose, logical flow, and argumentative appeal—skills that are frequently less developed in AI-generated writings (Shin et al., 2024). These findings only point to the need for educators to train students to see the trade-offs between linguistic complexity and communicative purpose so they are equipped to succeed in an AI-permeated world.

6 Implications for educators

These findings have important implications for instructors interested in incorporating AI tools like ChatGPT into writing instruction. On one hand, ChatGPT offers valuable opportunities for students to explore new possibilities in vocabulary, syntax, and style. For example, L2 learners can benefit from studying AI-generated texts as models for improving their writing. By analyzing these outputs, students can identify trends in vocabulary usage, sentence structure, and organizational strategies, helping them develop their linguistic skills and refine their writing. AI tools can also support the brainstorming and early drafting stages, enabling students to overcome creative blocks and experiment with different writing approaches. However, the integration of AI tools in education also presents significant challenges that need attention. A major concern is that excessive reliance on AI-generated content could hinder students' development of essential writing skills such as originality, critical thinking, and rhetorical awareness. Relying too much on AI support could prevent students from engaging deeply with their own

ideas and reasoning, ultimately stunting their growth as independent thinkers and communicators.

To address these concerns, instructors should adopt a balanced approach that encourages responsible technology use. For example, assignments that ask students to compare and contrast their own writing with AI-generated content can help them understand the trade-offs between complexity and clarity. Collaborative projects that combine human creativity with AI assistance can encourage students to view technology as a tool that complements, rather than replaces, their original thinking. Furthermore, educators have a crucial role in teaching students to critically assess AI-generated content. Although ChatGPT is an advanced tool for language generation, its outputs should not be accepted uncritically. Instead, students must be taught to evaluate the strengths and weaknesses of AI-generated text to ensure it aligns with learning goals and communication standards. For instance, instructors can show students how to identify where AI writing may lack factual accuracy, context, or rhetorical finesse. By fostering critical thinking, educators can help students benefit from AI-assisted writing while maintaining authenticity and creativity in their work.

7 Ethical considerations

Overreliance on AI-generated content can blur the lines between authorship and automation, raising concerns about plagiarism, accountability, and copyright. For example, students who submit AI-generated essays without proper acknowledgment violate academic policies, undermining the value of original content. To address these issues, institutions should establish clear guidelines and policies for AI use in education, grounded in transparency, integrity, and equitable access. These policies should ensure that students and staff use AI tools responsibly and ethically. One approach is requiring students to disclose when and how they have used AI in their writing. This would promote transparency, allowing educators to assess how students are engaging with AI tools and guide them toward more responsible use.

Additionally, integrating academic integrity training into the curriculum would help students understand the ethical considerations of AI-generated work and emphasize the importance of originality. By fostering a culture of accountability, educators can support students in navigating the ethical challenges of AI tools while upholding standards of transparency and integrity. Another important ethical concern is ensuring equal access to AI tools. While AI technologies like ChatGPT offer exciting possibilities for learning and creativity, their use can exacerbate existing inequalities in education. For example, students from disadvantaged backgrounds may lack the technology or resources needed to fully benefit from AI tools, creating a divide between privileged and underprivileged students. To address this, policymakers must ensure that AI technologies are accessible to all students, regardless of socioeconomic status. This could include providing subsidies for students in need, offering professional development for teachers, and designing AI tools that are accessible to diverse users, such as L2 learners, individuals with disabilities, and non-native speakers.

Beyond identifying these concerns, more concrete strategies are needed to reduce bias and promote equitable access. Algorithmic auditing, along with the inclusion of more diverse linguistic and cultural data in model training, can help mitigate systemic bias in AI output. At the classroom level, instructors can build AI literacy by

asking students to critically evaluate AI-generated texts for biased or misleading language, turning potential risks into teachable moments. Transparency policies—requiring students to disclose how AI tools were used—further balance innovation with accountability. Together, these strategies move the conversation from broad ethical concerns to actionable practices for responsible AI integration.

As AI tools continue to advance, it is essential to find a balance between innovation and responsibility in their application to education. While tools like ChatGPT hold great promise for enhancing writing pedagogy, their use must be guided by a commitment to inclusive and ethical practices. Collaboration among teachers, software developers, and policymakers is crucial to ensuring that AI models align with educational goals. For instance, developers can prioritize creating tools that adjust to different levels of reader proficiency, such as incorporating readability sliders or language modes for specific audiences. Educators can also partner with technologists to design AI programs that prioritize clarity and simplicity without compromising on sophistication. Interdisciplinary collaboration will be key in shaping the future of AI in writing. Linguists, educators, ethicists, and technologists must work together to address the complex issues surrounding AI integration. Linguists can offer insights into the nuances of human language that current models struggle to capture, such as rhetorical structure, argumentative depth, and emotional impact. Educators can provide feedback on how AI tools perform in the classroom, identifying both their strengths and limitations. Ethicists can raise concerns about the broader societal implications of AI, including issues of bias, fairness, and privacy. By fostering interdisciplinary collaboration, we can develop AI tools that not only replicate human language but also promote effective communication in both academic and professional contexts.

The findings of this research suggest that integrating AI tools into writing instruction should be approached cautiously. While ChatGPT excels in lexical diversity and syntactic complexity, its texts often fall short in terms of readability and relevance. These findings underscore the importance of teaching students to recognize the trade-offs between linguistic sophistication and clarity, preparing them to navigate an AI-driven world. Moreover, the ethical implications of using AI in education must be carefully considered, and institutions should establish clear guidelines for its responsible use. By grounding this project in strong theoretical frameworks and encouraging interdisciplinary collaboration, we can leverage AI's capabilities to enhance meaningful communication in scholarly settings. As AI technology continues to evolve, a careful, thoughtful approach is necessary for its integration into education. Only by fully understanding the complexities of human communication can we ensure that AI tools are used as catalysts for growth and transformation, empowering students to become confident, independent, and critically engaged writers in an increasingly AI-driven world.

8 Conclusion

This study highlights the linguistic depth of AI-generated text and student essays through the lens of Communication Theory. By analyzing factors like lexical variation, syntactic complexity, and readability, it demonstrates ChatGPT's ability to create sophisticated content, while also pointing out its challenges with communicative simplicity and relevance. These findings emphasize the importance of

collaboration between educators and developers to improve AI tools for better alignment with learning goals. Future research should explore ways to adjust AI models to prioritize clarity and readability without losing complexity. This work lays the foundation for designing AI tools that enhance communication in educational settings. To fully capitalize on AI's benefits while minimizing its drawbacks, educators must adopt a balanced approach, focusing on building writing skills and ensuring the safe use of technology.

Practical recommendations include activities that compare student and AI texts, encouraging students to weigh the trade-offs between complexity and clarity. Group projects combining human creativity and AI technology could help students use AI responsibly. Developers should focus on creating models that adapt to varying reader abilities, such as incorporating readability sliders or audience-specific language options. Policymakers should ensure AI tools are accessible, with attention to bias and fairness in their design and implementation. Finally, this study's limitations should be addressed in future research. While the analysis offers valuable insights into AI's linguistic sophistication, the relatively small sample size (50 student essays and 50 AI texts) limits the generalizability of the results, particularly across different genres, fields of study, or levels of writing proficiency. At the same time, the controlled corpus provides a meaningful pilot baseline for identifying contrasts in syntactic and rhetorical patterns. Future research could expand the dataset to include a wider range of genres, fields, and proficiency levels to provide a clearer understanding of AI's capabilities and limitations.

Another limitation concerns the absence of a qualitative layer. While rhetorical analysis and reader-response surveys would add important nuance, they were intentionally excluded here to preserve the study's scope and focus. The present project was designed as investigation to establish replicable, quantitative benchmarks and given the modest corpus size, a qualitative component risked being anecdotal rather than systematic. Future research, drawing on larger and more diverse datasets, will be better positioned to integrate qualitative methods in a way that complements and deepens the quantitative findings.

A further limitation lies in the study's narrow AI scope. The analysis focused solely on ChatGPT (GPT-4 Turbo), which was the default model available to ChatGPT Plus subscribers in December 2024. While this provides a clear benchmark, it does not account for differences across other large language models such as Gemini, Grok, or DeepSeek, nor newer releases of GPT. Future research should compare multiple AI systems to capture variation in linguistic output and communicative effectiveness across platforms.

In conclusion, this study lays the groundwork for developing AI tools that not only replicate human language but also foster effective communication in education. By bridging the gap between AI's potential and human communication needs, we can unlock AI's power to support learning, creativity, and innovation. However, achieving this vision will require continued collaboration among educators, developers, and policymakers, with a focus on ethical and inclusive practices.

As AI technology advances, it is vital to proceed cautiously, with curiosity and an understanding of human communication's complexities. Only then can we ensure that AI becomes a catalyst for growth and transformation, empowering students to be confident, independent, and critically engaged writers in an increasingly AI-driven world.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

DF: Writing – review & editing, Writing – original draft. LC: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2025.1616935/full#supplementary-material

References

Attanasio, M., Mazza, M., Le Donne, I., Masedu, F., Greco, M. P., and Valenti, M. (2024). Does ChatGPT have a typical or atypical theory of mind? *Front. Psychol.* 15:8172. doi: 10.3389/fpsyg.2024.1488172

Barattieri di San Pietro, C., Frau, F., Mangiaterra, V., and Bambini, V. (2023). The pragmatic profile of ChatGPT: assessing the communicative skills of a conversational agent. *Sistemi Intelligenti* 35, 379–399. doi: 10.1422/108136

Barrot, J. S. (2023). Using ChatGPT for second language writing: pitfalls and potentials. *Assess. Writ.* 57:100745. doi: 10.1016/j.asw.2023.100745

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty *Educational Psychology Review*, 24, 63–88. doi: 10.1007/s10648-011-9181-8

Biber, D., Conrad, S., and Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. New York: Cambridge University Press.

Brennan, S. E. (1991). "Grounding in communication" in Perspectives on socially shared cognition. eds. L. Resnick, J. M. Levine and S. D. Teasley (Washington, DC: American Psychological Association), 127–149.

Cong, Y. (2024). Manner implicatures in large language models. *Sci. Rep.* 14:Article 29113. doi: 10.1038/s41598-024-80571-3

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., et al. (2023). ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front. Public Health* 11:1166120. doi: 10.3389/fpubh.2023.1166120

Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., and Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Comp. Educ.* 6:100209. doi: 10.1016/j.caeai.2024.100209

Fredrick, D. R., Craven, L., Brodtkorb, T., and Eleftheriou, M. (2024). The role of faculty expertise and intuition in distinguishing between AI-generated text and student writing. *English Scholars Beyond Borders* 10, 126–150.

Granger, S. (2021). "Phraseology, corpora and L2 research" in Perspectives on the L2 Phrasicon: The view from learner corpora. ed. S. Granger (Bristol: Multilingual Matters), 3–21.

Grice, H. P. (1975). "Logic and Conversation" in Syntax and semantics: Vol. 3. Speech acts. eds. P. Cole and J. L. Morgan (New York: Academic Press), 41–58.

Grice, H. P. (1989). "Logic and conversation" in Studies in the way of words (Cambridge, MA: Harvard University Press), 41–58.

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., and Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Sci. Rep.* 13:Article 18617. doi: 10.1038/s41598-023-45632-3

Kalyan, K. S., Rajasekharan, A., and Sangeetha, S. (2021). AMMUS: A survey of transformer-based pretrained models in natural language processing. arXiv preprint arXiv:2108.05542. Available online at: https://arxiv.org/abs/2108.05542

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing *International Journal of Corpus Linguistics*, 15, 479–496. doi: 10.1075/ijcl.15.4.02lu

Malvern, D. D., Richards, B. J., Chipere, N., and Durán, P. (2004). *Lexical Diversity and Language Development: Quantification and Assessment.* Palgrave Macmillan.

McEnery, T., and Hardie, A. (2011). Corpus linguistics: Method, theory and practice. Cambridge: Cambridge University Press.

Mizumoto, A., Yasuda, S., and Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: through comparative analysis of linguistic fingerprints. *Appl. Corpus Linguist.* 4:106. doi: 10.1016/j.acorp.2024.100106

Settaluri, V. S., Subramani, R., Krishna, R., and Sachan, M. (2024). PUB: a benchmark for probing the understanding of pragmatics in large language models. *Proceed. AAAI Conf. Artif. Intellig.* 38, 12075–12084. Washington, DC: AAAI Press. doi: 10.1609/aaai.v38i13.29300

Shin, D., Koerber, A., and Lim, J. S. (2024). Impact of misinformation from generative AI on user information processing: how people understand misinformation from generative AI. *New Media Soc.* 27, 4017–4047. doi: 10.1177/14614448241234040

Sperber, D., and Wilson, D. (1986). Relevance: Communication and cognition. Cambridge, MA: Harvard University Press.

Sperber, D., and Wilson, D. (1995). Relevance: Communication and Cognition (2nd ed.). Oxford, UK: Blackwell Publishers.

Wölfel, M., Shirzad, M. B., Reich, A., and Anderer, K. (2024). Knowledge-based and generative-AI-driven pedagogical conversational agents: a comparative study of Grice's cooperative principles and trust. *Big Data Cogn. Comput.* 8:2. doi: 10.3390/bdcc8010002

Yildiz Durak, H., Eğin, F., and Onan, A. (2025). A comparison of human-written versus AI-generated text in discussions at educational settings: investigating features for ChatGPT, Gemini and BingAI. *Eur. J. Educ.* 60:e70014. doi: 10.1111/ejed.70014