Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
45(4) 2026, pp. 85-108
https://doi.org/10.22099/tesl.2026.56334.3510
Online ISSN: 2981-1546
Research Paper

Interactional Metadiscourse Markers in the Era of Artificial Intelligence: An Analysis of Human and GPT-Generated Academic Writing

Ali Roohani [1*] Ahmadreza Nikbakht [2]

Abstract

As the boundaries between human and machine-authored academic texts continue to blur, there is a compelling need to evaluate how artificial intelligence (AI)-generated content navigates the landscape of interpersonal engagement and persuasion in research articles. This investigation examined the functional role of interactional metadiscourse markers in academic texts authored by humans versus those generated by ChatGPT. Relying on Hyland's (2005) metadiscourse framework, this research study compared the frequency, rhetorical function, and contextual use of five key interactional markers in the conclusion and discussion sections of 40 applied linguistics research articles. Results from a mixed-methods corpus-based discourse analysis revealed significant differences in rhetorical style and communicative intent. The human-authored texts employed more hedges, engagement markers, and attitude markers, indicating a more cautious, evaluative, and reader-aware stance. In contrast, the AI-generated texts overused self-mentions and lacked contextual subtlety, often resulting in a more expository and impersonal tone. These findings highlight AI's current limitations in replicating the rhetorical sensitivity required in academic writing and suggest the continued need for critical human oversight in AI-assisted scholarly communication.

Keywords: AI-generated text, human-authored texts, ChatGPT, Metadiscourse

[*] Review History:
Received: 06/06/2026 Revised: 24/08/2026 Accepted: 07/09/2026

1. Associate Professor, English Department, Faculty of Letters and Humanities, Shahrekord University, Shahrekord, Iran. (Corresponding Author) roohani.ali@gmail.com
2. Ph.D. Student, , English Department, Faculty of Letters and Humanities, Shahrekord University, Shahrekord, Iran.. ahmadreza13771385@gmail.com

How to cite this article:
Roohani, A. & Nikbakht. A. (2026). Interactional metadiscourse markers in the era of artificial intelligence: An analysis of human and GPT-Generated academic writing. *Teaching English as a Second Language Quarterly*, *45*(4), 85-108.
https://doi.org/10.22099/tesl.2026.56334.3510

The role of metadiscourse, particularly for academic and professional success, cannot be overstated. For second/foreign language (L2) writers, mastering metadiscourse is arguably the critical step in transitioning from being a competent language user to an effective and authoritative communicator. It unlocks the nuanced rhetorical expectations of L2 discourse, allowing an L2 writer to do more than present punctuation and grammar; it lets them manage how their writing is received. With rapid breakthroughs in AI, particularly in natural language processing (NLP), AI-generated academic texts have gained significant attention. Tools such as ChatGPT are increasingly being used for academic writing, raising questions about their rhetorical effectiveness and alignment with human-authored texts (Biber & Conrad, 2021; Pham & Wang, 2023).

Academic writing is not simply a means of conveying information; rather, it involves a complex interplay of linguistic and rhetorical strategies that contribute to meaning-making (Ädel & Mauranen, 2020; Benelhadj, 2025). A key aspect of academic discourse is the role of interactional metadiscourse markers, including attitude markers, hedges, boosters, engagement markers, and self-mentions (Hyland, 2005, 2019). These markers allow writers to frame their own contribution relative to the claims they present, signal their degree of certainty, acknowledge alternative perspectives, and engage readers in academic discussions (Dahl, 2021; Gillaerts & Van de Velde, 2010). While human authors strategically employ these markers to navigate scholarly debates, the extent to which AI-generated texts incorporate them effectively remains underexplored. AI-generated texts have shown considerable potential to assist academic writers, particularly L2 writers, in improving coherence and fluency (Li, 2023). AI tools can serve as scaffolding mechanisms, providing suggestions for sentence structuring, lexical choices, and discourse organization (Sudajit-apa, 2025; Zhang & Liao, 2023). However, whether these tools enhance or undermine the rhetorical sophistication of academic writing remains an open question.

As AI-generated texts become increasingly integrated into academia, serious concerns about their rhetorical adequacy, coherence, and persuasiveness emerge. (Floridi & Chiriatti, 2020; Khajavi & Ezhdehakosh, 2025). Several studies (e.g., AbdAlgane et al.,2026; Pham & Wang, 2023) suggest that AI-generated academic texts may show grammatical and syntactic accuracy but may lack the intricate rhetorical strategies human writers use to engage readers and construct persuasive arguments. Several studies (e.g., Dreyfus et al., 2021; Kuteeva & Mauranen, 2022) also suggest that AI-generated texts often exhibit inconsistent rhetorical strategies, fail to align with disciplinary norms, and raise concerns about their suitability for high-stakes academic writing, such as journal publications and doctoral dissertations. This deficit is further clarified by recent works (e.g., Li, 2023; Tang, 2023; Yu & Chen, 2024; Zhang & Liao, 2023), indicating that while AI can mimic surface-level conventions and assist writers, particularly L2 writers, with coherence, fluency, and structural suggestions, it largely lacks the deeper rhetorical competence required for persuasive scholarly work.

Moreover, while AI models can generate coherent texts, their reliance on pre-existing linguistic patterns and probabilistic predictions sometimes leads to formulaic writing, a lack of critical voice, and insufficient engagement with disciplinary debates (Bhatia, 2020). This issue is particularly salient in the discussion and conclusion sections, where scholars are expected to evaluate findings, position their work within existing research, and articulate broader implications (Swales & Feak, 2021). If AI-generated texts fail to employ interactional metadiscourse markers effectively, they may lack the persuasive force and argumentative depth characteristic of human-originated discourse (Bruce, 2022).

In light of these issues, the present investigation aimed to compare interactional metadiscourse markers in human-authored and AI-generated academic texts, focusing on the discussion and conclusion sections of articles. Through a systematic analysis of metadiscourse use in AI- and human-produced texts, this work advances the literature on AI's role in academic writing. The analysis highlights patterns of interactional metadiscourse in AI-generated versus human-authored texts and considers the implications for scholarly communication.

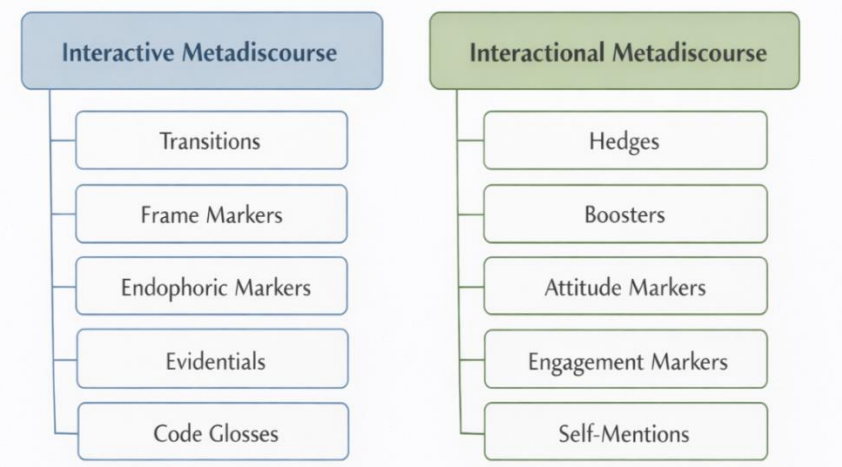
Literature Review

Metadiscourse

Metadiscourse refers to the linguistic resources through which writers organize a text, express a stance toward its propositions, and guide readers' interpretation of the unfolding argument (Hyland, 2019). It is therefore not an ornamental layer added to content; rather, it is part of the interpersonal work through which academic knowledge is negotiated. Hyland's (2005) influential model distinguishes between interactive and interactional metadiscourse. Interactive resources help readers follow a text's propositional organization through features such as transitions, frame markers, endophoric markers, evidentials, and code glosses. Interactional resources, in contrast, reveal the writer's position and construct a relationship with readers through hedges, boosters, attitude markers, engagement markers, and self-mentions. Figure 1 displays the main categories and subcategories of these markers based on Hyland (2005).

Figure 1

The Main Categories and Subcategories of Metadiscourse Markers According to Hyland (2005)



The value of metadiscourse becomes especially visible when academic writing is understood as a form of disciplinary interaction. Writers use these resources to calibrate commitment, acknowledge uncertainty, signal evaluation, and anticipate a scholarly audience's needs and expectations (Ädel & Mauranen, 2020; Khedri & Kritsis, 2018). Hedges such as *might*, *possibly*, and *could* allow authors to limit a claim's scope and leave room for alternative explanations, whereas boosters such as clearly, definitely, and in fact signal a stronger commitment to a proposition (Hyland, 2019). Attitude markers encode evaluation, engagement markers explicitly orient readers toward an interpretation, and self-mentions make the author's presence visible. The effectiveness of these devices depends not simply on their frequency but on how their rhetorical markers contribute to academic credibility when their strength, position, and function align with the evidence and the conventions of the discipline. Empirical research also supports this context-sensitive view. For instance, Baofang and Shamsudin (2023) found that transition markers were highly frequent in the English abstracts of chemistry master's theses, whereas frame markers were comparatively underused. Their findings suggest that writers may recognize the need to connect propositions without fully exploiting the resources that signal stages of an argument. This imbalance is pedagogically important because it shows why instruction based only on lists of forms or frequency counts is insufficient.

Recent work has further demonstrated that metadiscourse marker use can be taught through approaches that make its rhetorical functions visible. Esfandiari and Allaf-Akbary (2024) reported that both hands-on and hands-off data-driven learning activities improved advanced Iranian L2 English learners' ability to identify and use interactional metadiscourse markers, with the group working directly with Microsoft Copilot showing stronger gains. At the same time, Pearson and Abdollahzadeh's (2023) systematic review of 370 empirical studies cautions against treating metadiscourse markers as a decontextualized inventory. More than 80% of the reviewed studies used cross-sectional descriptive corpus methods, and the authors called for

greater attention to diachronic development and contextually grounded interpretation. Taken together, metadiscourse markers should be examined as a functional resource in discourse, and quantitative distributions should be interpreted alongside the local rhetorical purposes they serve.

Interactional Metadiscourse in Human and AI-Generated Academic Texts

The rapid incorporation of generative AI into academic writing has sparked debates about its effectiveness. AI systems can produce grammatically accurate and apparently coherent prose, and they may assist L2 writers with sentence construction, lexical choice, fluency, and initial organization (Floridi & Chiriatti, 2020; Li, 2023; Zhang & Liao, 2023). Surface-level fluency, however, does not necessarily indicate rhetorical competence. Academic writers must make situated decisions about how strongly to state a claim, when to acknowledge uncertainty, which perspective to foreground, and how to position a contribution in relation to earlier scholarship. Genre, discipline, evidence, and an anticipated readership shape these decisions. Consequently, the central issue is not whether AI can generate individual hedges or boosters, but whether it can deploy interactional resources as a coordinated system of stance and reader engagement. The available comparative evidence points to a gap between formal well-formedness and rhetorical adaptability.

Biber et al. (2023) found that AI-generated texts used hedges and boosters inconsistently, while Shalevska (2024) reported frequent reliance on a limited hedge, such as *may*, alongside a less varied repertoire of boosters. Such patterns may produce writing that sounds cautious or academic at the sentence level without showing a consistent relationship between epistemic strength and evidential support. Jiang and Hyland (2025) similarly observed that ChatGPT-generated argumentative essays were structurally cohesive, but they contained fewer interactional resources than essays written by British university students. The student texts displayed more intricate stance-taking and more personalized reader orientation, whereas the AI texts tended toward an impersonal, expository presentation. These findings suggest that cohesion and interaction should be distinguished; a text may be easy to follow while still offering readers limited guidance about how to evaluate its claims.

Evidence from other genres reinforces the importance of genre-sensitive analysis. Xu (2025) found differences in the frequency and sequencing of rhetorical moves in scholar-written and AI-generated abstracts, with AI texts often departing from conventional genre structures. Pham and Wang (2023) reported limited variation in engagement markers in AI-generated academic essays, a pattern that can reduce persuasive force when readers are not actively guided through the argument. Conversely, Zeng et al. (2024) identified an overuse of self-mentions, such as *I argue* and *we propose*, in AI-generated research writing. The contrast between underdeveloped reader engagement and excessive authorial visibility is revealing; AI-generated texts may reproduce recognizable academic signals without coordinating them with the social purposes of scholarly discourse. In this sense, the problem is not the presence or absence of a particular

lexical item but the lack of a stable rhetorical relationship among stance, audience, evidence, and genre.

The literature therefore supports a more differentiated account of the human–AI comparison. Human-authored academic writing is not uniformly superior, nor is AI-generated writing uniformly deficient; both vary across disciplines, genres, prompts, and levels of human intervention. Nevertheless, existing studies converge on three relevant observations. First, AI systems can produce fluent, structurally organized academic prose. Second, their use of interactional metadiscourse appears less diverse, less context-sensitive, or less aligned with evidential and genre-specific demands than in human-authored texts (Kuteeva & Mauranen, 2022; Tang, 2023; Yu & Chen, 2024). Third, frequency alone cannot explain the communicative consequences of these differences.

A marker's contribution to coherence and persuasiveness depends on its placement within an argument and on its relation to the writer's stance and the reader's interpretive role. The empirical gap is particularly visible in the discussion and conclusion sections of research articles. These sections do more than summarize results; they interpret findings, compare them with previous research, acknowledge limitations, formulate implications, and establish the study's significance (Dahl, 2021; Swales & Feak, 2021). These are likely the areas where hedging, boosting, attitude, engagement, and self-mention play their crucial interpersonal roles. Although previous studies have compared human and AI writing in essays, abstracts, and general academic prose, fewer studies have examined how some interactional categories operate in the culminating sections of applied linguistics research articles. A comparison that combines frequency with contextual and functional interpretation can therefore clarify whether AI-generated texts merely imitate isolated academic expressions or use them effectively to construct reader-centered scholarly arguments.

Against this background, the current investigation compared the use of several interactional metadiscourse markers in the discussion and conclusion sections of human-authored and ChatGPT-generated applied linguistics research articles. Drawing on Hyland's (2005) framework, this study examined the distribution of five interactional categories: hedges, boosters, attitude markers, engagement markers, and self-mention. It also analyzed the rhetorical functions these categories performed in the context. This mixed-methods design aimed to connect the quantitative question of how often markers occurred with the qualitative questions of how they were embedded in claims, evidence, evaluations, and reader guidance. This approach allowed assessment not only of whether the two text types differed in frequency or category, but also of whether those differences could affect the coherence, clarity, credibility, and persuasive orientation of academic discourse. The study addressed the following research questions:

1. Do the frequencies and types of interactional metadiscourse markers significantly differ between human-authored and AI-generated discussion and conclusion sections?

2. In what ways do metadiscourse markers in AI-produced texts influence coherence and clarity of discourse in the academic discussion and conclusion sections when compared with the human-authored texts?

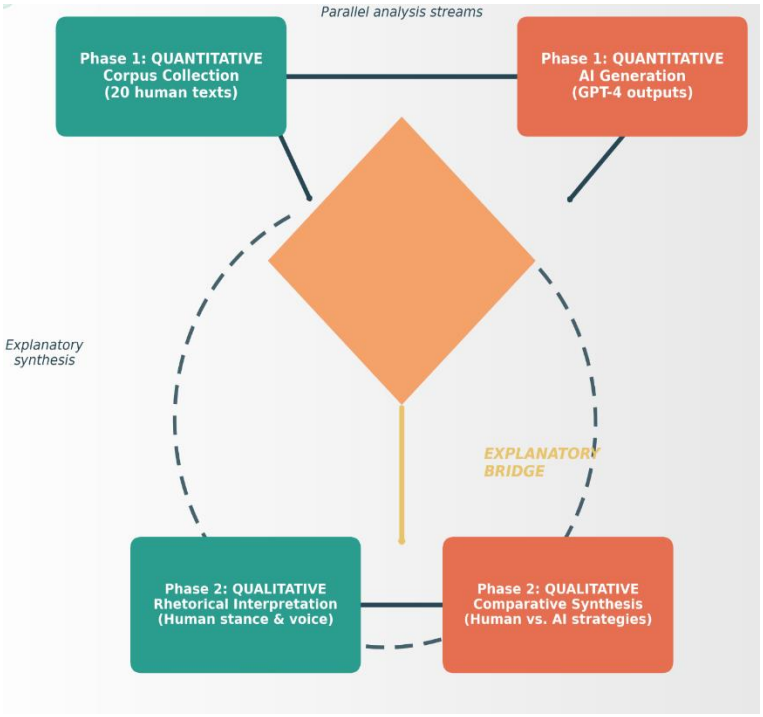
Method

Research Design

An explanatory sequential mixed-methods approach was used to address the research questions. According to Creswell and Plano Clark (2017), this two-phase design begins with collecting and analyzing quantitative data, followed by a qualitative phase intended to further explore and elaborate on the initial quantitative findings. This design was selected for its utility in investigating the uncharted territory of interactional metadiscourse in AI-produced vs. human-authored academic texts. The initial quantitative phase provided a broad, systematic comparison of the distribution and frequency of metadiscourse markers, identifying what patterns and significant differences exist. The subsequent qualitative phase then delved into the contextual use, rhetorical function, and nuanced appropriacy of these markers to explain *how* or *why* the observed quantitative patterns might have emerged. This design enabled a comprehensive analysis: the quantitative data revealed generalizable trends, while the qualitative data provided explanatory depth. Figure 2 illustrates the design's procedural flow.

Figure 2

Procedural Flow of the Explanatory Sequential Mixed-Methods Design



Corpus

Discussion and Conclusion of the Published Articles (human-generated corpora)

The data were gathered from the final parts of the studies, namely "discussion" and "conclusion, from the *Journal of Research in Applied Linguistics*. These sections most clearly communicate the authors' findings and conclusions, along with overarching generalizations, in which metadiscourse markers are particularly prominent. (e.g., Hunston, 1994). The combined use of both terms "discussion" and "conclusion" reflects the fact that the naming of this section is often subject to authorial preference and disciplinary practice, in line with the submission guidelines and conventions of the target journal. In both cases, the reference is to the concluding sections in which the authors connect the results more closely to real-world phenomena and then generalize them to increase general applicability. In one instance in the dataset, the authors used both headings and separated the discussion of the results from the study's conclusion; in such cases, both sections were included in the corpus. Another rationale for this decision is that the model (ChatGPT) produces the most effective response when the prompt contains both headings.

The *Journal of Research in Applied Linguistics* is an international, double-blind peer-reviewed, open-access journal in applied linguistics, and is published by Shahid Chamran University of Ahvaz, Iran. This journal is indexed in major databases, including Web of Science and Scopus. The journal is issued twice yearly in both print and electronic formats. It provides a scholarly forum for examining language use as a component of social life. The journal publishes studies that investigate the relationship between language and context in instructed and uninstructed L2 learning settings, drawing on a wide range of methodological approaches within the humanities and social sciences.

ChatGPT

ChatGPT (GPT-4 version), used in this study, is an AI language model developed by OpenAI that performs complex natural language tasks in response to carefully tailored prompts (OpenAI, 2023). A structured prompt instructed GPT-4 to generate academic-style text in applied linguistics, creating the primary AI-authored corpus for analysis. The GPT-4 model underlying the training, which incorporates reinforcement learning from human feedback (OpenAI, 2023), ensures that it can produce coherent, contextually appropriate text in a specified genre. This capability allowed the research to systematically generate a comparable body of AI-generated academic writing, which was then examined alongside human-authored texts to investigate patterns in interactional metadiscourse.

ChatGPT (GPT-4 version) could produce texts exceeding 500 words. This issue required an additional instruction to meet the approximate length requirement. The final dataset was created using the prompt that produced the optimal output for this study's parameters.

Prompt: *Based on the following abstract, compose a combined discussion and conclusion section (maximum 700 words) for a research paper in the field of applied linguistics, specifically focusing on English language teaching and learning.*

In summary, the first corpus resource came from the *Journal of Research in Applied Linguistics*. It comprised "discussion" and "conclusion" sections of articles published between 2021 to 2024. The corpus consisted of discussions and conclusions from 20 different research studies. The mean length of the discussions and conclusions was about 700 words. Therefore, the data harvested from ChatGPT (GPT-4 version) included an instruction to make the discussions and conclusions approximately 700 words. Subsequently, we built the ChatGPT corpus using the prompt described above and occasionally used alternative completions, which produced a new and different response based on the same prompt. This was done only when the language model produced a discussion or conclusion that was too short.

Corpus Size and Comparability

A crucial methodological consideration in corpus-based metadiscourse studies is corpus-size comparability, as unequal corpora can yield misleading frequency comparisons without appropriate normalization procedures (Hyland, 2005). To address this issue, we carefully ensured that both corpora were closely matched in size. The human-authored corpus comprised 20 discussion and conclusion sections totaling 14,023 words (mean = 701.15 words, standard deviation = 48.32, range = 623-782). The AI-generated corpus comprised 20 discussion and conclusion sections totaling 13,877 words (mean = 693.85 words, standard deviation = 52.17, range = 618-795). A Levene's test for equality of variances confirmed that the two corpora did not differ significantly in their variance *[F* (1,38) = 0.672, *p* = .418], and an independent-samples *t*-test confirmed no significant difference in mean word count between the two corpora (*t*(38) = 0.467, *p* = .643). These statistics confirm that the two corpora are closely matched in size, with the human-authored corpus being only 146 words larger than the AI-generated corpus (a difference of approximately 1%). Because the corpus sizes were nearly identical, we also conducted direct frequency comparisons, following the methodology of leading scholars in metadiscourse research, including Hyland (2005, 2019) and Ädel & Mauranen (2020). They argue that similar corpus sizes allow for a stable comparison.

Data Collection and Analysis

The data collection procedure for the present investigation included several steps: First, articles were selected from the *Journal of Research in Applied Linguistics*. The discussion and conclusion sections were selected, as these parts typically feature a high density of interactional metadiscourse markers that reflect authors' interpretations, findings, and generalizations.

Second, 20 articles were selected to ensure topic variety while maintaining comparability in style and structure. We extracted and compiled the discussion and conclusion sections into a single corpus. Third, the word count for each article's section was recorded to calculate the

mean word length (m =700 words). Fourth, we developed a prompt to ensure consistency and comparability. Each abstract from the 20 selected articles in the human-authored corpus was input into ChatGPT (GPT-4 version). ChatGPT generated a discussion and conclusion section for each abstract. If the generated text was too short or low quality, we requested an alternative completion using the same prompt. This ensured that all texts met the approximate 700-word target, aligned with the mean word count of the human-authored corpus. Fifth, both corpora were pre-processed to ensure consistency. The texts were standardized to a uniform format, with clear sections. The texts were tokenized to count total words and accurately identify metadiscourse markers. At this stage, two experts (professors of applied linguistics) reviewed the texts to confirm that the discussion and conclusion sections adhered to academic conventions.

Sixth, the researchers carried out quantitative and qualitative analyses of the interactional metadiscourse markers identified in the corpus. The researchers conducted content analysis of the corpora to identify the types and frequencies of each metadiscourse marker in the discussion and conclusion sections of both corpora. The analysis drew on Hyland's (2005) metadiscourse framework. The focus was on the interactional markers, such as hedges (e.g., might), boosters (e.g., clearly), attitude markers (e.g., interestingly), engagement markers (e.g., let us), and self-mentions (e.g., we suggest), for identifying and categorizing these markers according to their rhetorical functions within the texts. To establish the reliability of the coding procedure, two coders, a PhD student of teaching English and an associate professor of applied linguistics with extensive experience in discourse analysis, independently coded a 10% stratified random sample of the corpus (n = 4 texts, selected proportionally from both corpora). Inter-rater/coder reliability was assessed using Cohen's kappa coefficient. The kappa values for each marker category were as follows: hedges (κ = .87), boosters (κ = .82), attitude markers (κ = .84), engagement markers (κ = .88), and self-mentions (κ = .91). Table 1 below summarizes the inter-coder reliability results**.**

Table 1

Inter-Coder Reliability for Interactional Metadiscourse Marker Coding

Marker Type	Cohen's Kappa (κ)	Agreement Level
Hedges	.87	Almost Perfect
Boosters	.82	Almost Perfect
Attitude Markers	.84	Almost Perfect
Engagement Markers	.88	Almost Perfect
Self-Mentions	.91	Almost Perfect

These observed values indicated excellent agreement for self-mentions and substantial to excellent agreement for all other categories, according to Landis and Koch's (1977) benchmarks (where κ > .80 indicates almost perfect agreement). After calculating inter-coder reliability, the two primary researchers resolved discrepancies through discussion and consensus. To

safeguard coding consistency, a third expert (an external senior professor) was consulted when necessary to arbitrate any persistent disagreements.

After this initial reliability and accuracy check, the primary researchers coded the remaining corpus. The established coding scheme guided the coding process. It included ongoing consultation with the third independent coder to maintain consistency throughout the analysis, which resulted in high inter-coder reliability for the established categories: hedges ($\kappa = .97$), boosters ($\kappa = .98$), attitude markers ($\kappa = .98$), engagement markers ($\kappa = .98$), and self-mentions ($\kappa = .97$). Any minor coder disagreements between the two primary coders were also resolved by the third coder/expert to finalize the coding of interactional metadiscourse markers.

Moreover, Chi-square tests were run in the data software to examine whether there were statistically significant differences in the use of metadiscourse markers in human-authored and AI-generated corpora. For the qualitative phase of this study, excerpts containing interactional markers were extracted from both corpora and segmented into meaningful units, such as sentences or clauses, where these markers appeared, following Hyland's (2005) interactional metadiscourse framework. Deductive-inductive content analysis was used; the segmented data were coded and thematically explored using Hyland's (2005) interactional metadiscourse markers. That is, each interactional marker was treated as a code, tagged by category (theme), and accompanied by contextual notes to identify and explain the markers' rhetorical functions and pattern distinctions. That is, discourse markers (codes) were grouped into broader discourse marker types (themes), with annotations reflecting functional differences, rhetorical clarity, and pattern distinctions.

To ensure data accuracy, the analysis underwent iterative refinement, with discrepancies discussed and resolved in collaboration with a second and third coder/analyst. To ensure credibility and dependability, we calculated inter-coder reliability using Cohen's kappa. We conducted peer debriefing with another subject matter expert (an external senior professor of applied linguistics) to verify interpretations. The final discourse marker categories were synthesized into a narrative comparing rhetorical strategies across human and AI-authored texts, highlighting key strengths, weaknesses, and distinctive patterns, thereby ensuring a systematic and rigorous qualitative analysis.

Results

Quantitative Analysis of Interactional Metadiscourse Markers

Table 2 reports the total number of interactional metadiscourse markers observed in both corpora, presented in both raw frequencies and normalized frequencies per 1,000 words. The human-authored corpus contained 1,057 markers (75.38 per 1,000 words), while the AI-generated corpus included 886 markers (63.84 per 1,000 words).

Table 2

Distribution of Interactional Metadiscourse Markers across the Two Corpora (Raw Frequencies and Normalized Frequencies per 1,000 Words)

Marker Type	Human-Authored		AI-Generated		Percentage (for the type)
	Raw (n)	Per 1000	Raw(n)	Per 1000	
Hedges	421	30.02	292	21.04	26.56
Boosters	276	19.68	248	17.87	24.67
Attitude Markers	173	12.34	112	8.07	21.72
Engagement Markers	108	7.70	62	4.47	5.92
Self-Mentions	79	5.63	172	12.39	21.13
Total	1057	75.38	886	63.84	100

The analytical results in Table 2 indicate that the human-authored texts showed a higher overall frequency of interactional metadiscourse markers in both raw counts and per 1,000 words (normalized rates per 1,000 words). The consistency between raw and normalized frequencies confirms that the observed differences were not due to corpus size but reflected meaningful differences in metadiscourse use. The human-authored corpus contained 171 more markers in total (a difference of 16.2%), reflected in a normalized frequency of 75.38 markers per 1,000 words compared to 63.84 per 1,000 words in the AI-generated corpus.

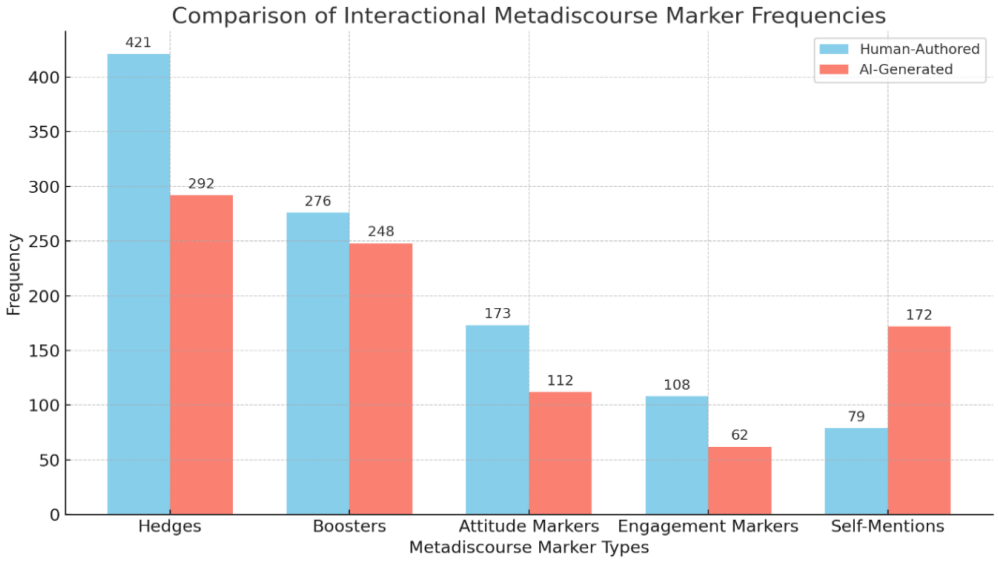
The results demonstrate that hedges were the most highly recurring marker type in both corpora, consistent with Hyland's (2011) observation that humanities writing often relies on cautious or tentative claims. However, the human-authored texts employed significantly more hedges (421 vs. 292; 30.02 vs. 21.04 per 1,000 words), suggesting greater rhetorical sensitivity to academic uncertainty and a higher degree of epistemic caution. AI-generated texts showed underuse of hedging language, which could lead to overconfident generalizations or less acknowledgment of alternative perspectives. Boosters appeared at similar rates in both corpora (276 vs. 248; 19.68 vs. 17.87 per 1,000 words). Both human authors and AI systems recognize the rhetorical value of asserting confidence and certainty in key claims.

Engagement markers were noticeably more prevalent in the human corpus (108 vs. 62; 7.70 vs. 4.47 per 1,000 words), reflecting more explicit attempts to involve the reader. AI texts used them less frequently, likely due to their focus on monologic, information-driven exposition. Engagement markers, which explicitly involve the reader, appeared more often in the human corpus.

Interestingly, AI-generated texts included significantly more self-mentions than human-authored texts did. This finding implies that ChatGPT, when prompted to mimic an academic tone, may overuse first-person references. Figure 3 displays a visual comparison of interactional metadiscourse markers in academic texts produced by humans and AI.

Figure 3

Frequency of Interactional Metadiscourse Markers in Human-Authored vs. AI-generated Academic Texts



As Figure 3 shows, human-authored texts used more hedges, attitude markers, and engagement markers. Boosters occurred at comparable frequencies across both groups, whereas self-mentions were more frequent in AI-generated texts. To test whether the differences across metadiscourse marker types, presented in Table 2 and Figure 3, were statistically significant, We Ran Chi-Square tests of independence. Table 3 reports the results.

Table 3

Chi-Square Test Results on Interactional Metadiscourse marker Types

Marker Type	**χ^2**	***p*-value**
Hedges	15.37	< 0.001
Boosters	1.22	0.269
Attitude Markers	10.45	0.001
Engagement Markers	9.83	0.002
Self-Mentions	26.14	< 0.001

Table 3 shows statistically significant differences in the frequency of most interactional metadiscourse markers ($p < .05$), except boosters. Significant differences were found for attitude markers, hedges, engagement markers, and self-mentions. However, the corpora did not differ significantly in booster frequency. Overall, the quantitative analysis shows that AI-

generated academic texts used interactional metadiscourse markers less frequently than human-authored texts. Human writers tend to adopt greater use of hedging, attitude, and engagement markers. Moreover, the ChatGPT-produced texts exhibited the overuse of self-mentions and underuse of nuanced rhetorical markers.

Qualitative Analysis of Metadiscourse Markers

To answer the second research question, I conducted a qualitative analysis to examine how and *why* the observed quantitative patterns might have emerged, offering a richer understanding of the textual and pragmatic dimensions. In other words, a qualitative approach was crucial to uncover how these markers function rhetorically within context and how they shape clarity, coherence, and author–reader interaction. The analysis focused on five interactional metadiscourse categories: boosters, hedges, attitude markers, engagement markers, and self-mentions, drawing on Hyland's (2005) interpersonal model. Rather than foregrounding tabulated results, the sections proceed thematically. Each marker type is first introduced conceptually, followed by a comparative discussion of its use in human-authored vs. AI-generated texts. Contextualized excerpts clearly labeled as human- or AI-generated are then provided to illustrate rhetorical effects. Each set of examples is accompanied by interpretation to highlight indications of academic credibility, coherence, and reader engagement. The study then synthesizes the qualitative patterns.

Hedges

Hedges play a central role in academic discourse by allowing writers to express caution, acknowledge complexity, and open space for alternative interpretations (Esfandiari & Allaf-Akbary, 2024). In the human-authored texts, hedging devices, such as *might suggest*, *could indicate*, and *is likely to*, were employed strategically to moderate claims and signal epistemic humility (see Excerpt 1). Qualitative data analysis showed that these markers were typically embedded within methodologically or theoretically sensitive contexts, particularly when discussing implications, limitations, or tentative interpretations of results.

Excerpt 1 (Human-authored):

"These findings *might suggest* that learner autonomy plays a greater role in long-term retention than previously assumed, particularly in contexts where instructional scaffolding is limited."

In the above excerpt, the hedge *may serve to soften the claim and* frame it as an interpretation rather than a definitive conclusion. The surrounding context further reinforces caution by delimiting the scope of applicability, thereby enhancing credibility and scholarly distance.

In contrast, AI-generated texts used hedging less frequently and less context-sensitively. Although forms such as *may suggest* or *possibly* appeared sporadically, they were often inserted formulaically and without sufficient alignment to methodological constraints or contextual variables. Excerpt 2, generated by GPT-4, shows that the phrase *"may be linked to"* is inserted

at the conclusion of a descriptive results paragraph, despite no prior discussion of research design limitations or alternative explanations. The hedge is not supported by methodological reflection such as small sample size or correlational data, nor does it connect to the specific variables under investigation. Instead, it appears as a routine addition, likely prompted by the instruction to adopt an academic register. Unlike the human-authored example, which integrates hedging into a carefully qualified interpretive frame, the AI version deploys the hedge as an isolated lexical token, detached from the epistemic logic that would justify its use in context.

Excerpt 2 (AI-generated):

"This clearly proves that the strategy is universally effective across educational settings. The results demonstrate that the intervention improves student engagement in all contexts."

Despite the complexity typically associated with educational interventions, the AI-generated excerpt lacks appropriate hedging and instead relies on categorical assertions. Without contextual qualifiers (e.g., participant variation, instructional setting, or limitations), it overgeneralizes. The minimal and superficial use of hedging in AI texts thus risks overstating findings and undermining academic credibility.

Overall, the contrast between Excerpts 1 and 2 illustrates that human-authored hedging promotes interpretive depth and rhetorical balance, whereas AI-generated texts tend toward assertiveness that compromises scholarly caution.

Engagement Markers

Engagement markers actively engage the reader in the discourse, guiding interpretation and establishing a dialogic relationship. Human-authored texts used engagement markers consistently and purposefully, such as *consider*, *we can observe*, and *it is important to note*. These devices were particularly salient at argument transitions or when addressing implications (see Excerpt 3).

Excerpt 3 (Human-authored):

"*Consider* the role of cultural context when interpreting these outcomes, as local norms may shape learner expectations and classroom interaction."

In Excerpt 3, the engagement marker *consider* directly addresses the reader and invites active cognitive participation. This strategy facilitates coherence by signaling how the argument should be processed and interpreted.

AI-generated texts, however, employed engagement markers infrequently and often without rhetorical development. When present, they tended to appear as generic phrases repeated across texts without effective rhetorical development (see Excerpt 4).

Excerpt 4 (AI-generated):

"You can see that the method is effective. You can see that the results support previous research."

The repetitive and undeveloped use of *you can see* in Excerpt 4 creates a mechanical tone and possibly fails to guide the reader through the reasoning process meaningfully. Generally,

the AI texts often appeared monologic, reducing rhetorical accessibility and reader involvement.

Attitude Markers

Attitude markers signal the writer's evaluative stance toward propositions and findings, contributing to argumentative coherence and interpretive clarity (Esfandiari & Allaf-Akbary, 2024). The human-authored texts used attitude markers such as interestingly, notably, and unfortunately to convey critical engagement with the data and structure evaluative judgments (see Excerpt 5).

Excerpt 5 (Human-authored):

"*Interestingly*, novice learners outperformed intermediate learners in spontaneous speaking tasks, contradicting our initial hypothesis and warranting further investigation."

In Excerpt 5, the marker *interestingly* highlights the finding's evaluative significance and prepares the reader for an interpretive shift. This rhetorical move enhances coherence by explicitly signaling the author's stance. However, in the AI-generated texts like Excerpt 6, attitude markers were either absent or were used in vague, formulaic statements lacking analytical depth.

Excerpt 6 (AI-generated):

"Interestingly, the results were positive and support previous studies."

In Excerpt 6, the absence of specificity or elaboration renders the marker rhetorically weak. Rather than guiding interpretation, it functions as superficial filler, resulting in a flatter, less engaging argumentative style.

Self-Mentions

Self-mentions are a key resource for establishing authorial presence in academic writing (Esfandiari & Allaf-Akbary, 2024). In the human-authored corpus, self-mentions, such as *we argue* or *we interpret* (see Excerpt 7), were used selectively to assert stance without dominating the discourse.

Excerpt 7 (Human-authored):

"We interpret this pattern as indicative of increased learner autonomy under conditions of reduced instructional control."

The concise use of *we interpret* in Excerpt 7 foregrounds authorial responsibility while maintaining rhetorical balance. By contrast, the AI-generated texts frequently overuse self-mentions and often rely on repetitive sentence openings *(*see Excerpt 8).

Excerpt 8 (AI-generated):

"We believe that our findings are significant. We believe that they support previous research. Our study demonstrates important implications."

This repetitive structure, as in Excerpt 8, creates redundancy and disrupts textual flow, suggesting limited rhetorical flexibility. Excessive self-reference in the AI texts thus weakens coherence and diminishes stylistic variation.

Boosters

Boosters signal certainty and emphasize strong assertions. Both corpora employed boosters, such as *clearly* and *undoubtedly* (see Excerpt 9). However, their rhetorical impact differed markedly. The human-authored texts used boosters sparingly and primarily in contexts supported by strong evidence.

Excerpt 9 (Human-authored):

"This *clearly* indicates a shift in learner preferences, particularly in digitally mediated learning environments."

In the AI-generated texts, boosters were more uniformly distributed and occasionally applied to claims lacking sufficient justification (see Excerpt 10).

Excerpt 10 (AI-generated):

"This undoubtedly proves the success of the method."

As Excerpt 10 shows, such overassertion risks rhetorical overreach and weakens argumentative rigor. While boosters enhanced clarity in the human texts, their indiscriminate use in the AI texts often distorted evidential balance. In closing, Table 4 presents a brief synthesis of the comparative qualitative data analysis.

Table 4

A Summary of Qualitative Comparison of Metadiscourse Marker Use

Metadiscourse Marker Type	Human-Authored Texts	AI-Generated Texts	Effect on Clarity & Coherence
Hedges	Frequently used to express uncertainty, acknowledge complexity, and soften claims.	Used less often and often inserted without context, leading to overgeneralization.	The human texts displayed nuanced arguments; The AI texts appeared overly confident or simplistic.
Boosters	Selectively used to emphasize well-supported claims.	Used at similar frequency, but sometimes applied to weak or unsupported claims.	Overuse in the AI texts might undermine argumentative rigor and distort balance.
Attitude Markers	Reflect critical stance and evaluative voice; strategically placed to guide reader interpretation.	Used formulaically, basically lacking evaluative depth or specificity.	The human texts appeared more engaged; the AI texts seemed impersonal.
Engagement Markers	Actively guide the reader (e.g., "consider," "note	Sparse usage; argument flows are more linear and reader-detached.	The human texts facilitated reader navigation; the AI texts felt more monologic.

Metadiscourse Marker Type	Human-Authored Texts	AI-Generated Texts	Effect on Clarity & Coherence
	that"); establish dialogic interaction.		
Self-Mentions	Used selectively to assert author stance without overwhelming presence.	Overused, with repeated phrases like "we argue" or "our study," creating redundancy.	Overuse in the AI texts disrupted flow and reduced rhetorical variation.

Discussion

The quantitative results showed that the AI-generated texts mostly underutilized critical metadiscourse categories, particularly hedges, attitude markers, and engagement markers. These markers are essential in constructing a dialogic, reader-aware academic voice (Hyland, 2019). Based on the results, the human-authored texts demonstrated significantly higher frequencies of such markers, suggesting greater rhetorical adaptability and genre awareness. Notably, hedges were the most prevalent marker type in both corpora, consistent with Hyland's (2011) assertion that cautious language is a hallmark of academic discourse. However, the AI-generated texts employed hedges with less strategic variation, often resulting in overgeneralized or overly assertive conclusions. Thus, drawing on Hyland's (2005) model and supported by recent scholarship (e.g., Jiang & Hyland, 2025; Xu, 2025), the above findings of this study confirm that AI-generated texts can demonstrate syntactic coherence and surface-level academic formality, yet they may fall short in replicating the nuanced rhetorical strategies characteristic of human writing. This issue suggests that human writers, when crafting academic conclusions and discussions, tend to engage more actively with their readers, adopt more cautious stances, and explicitly position themselves within the discourse. In contrast, the AI-generated texts, though competent in structure and academic tone, tend to be more neutral and mechanical in their rhetorical presentation.

Additionally, the AI corpus displayed a statistically significant overuse of self-mentions, a finding echoed by Zeng et al. (2024), who observed similar tendencies in AI-generated research articles. This over-personalization may stem from AI misinterpreting academic conventions or over-relying on training data where first-person constructions are commonplace. While some authorial presence is acceptable in academic writing, its excessive and repetitive use in AI texts often disrupts rhetorical flow and detracts from coherence. Jiang and Hyland's (2025) study supports these findings, showing that AI-generated texts exhibit limited strategic variation in hedging despite frequent surface-level usage, and Xu's (2025) research, which identified statistically significant overuse of self-mentions in AI-produced academic prose. Zeng et al. (2024) further corroborate this; their corpus analysis revealed similar patterns of rhetorical rigidity and misapplied metadiscourse in AI-generated research articles.

Qualitative analysis further highlights how AI-generated texts, though grammatically competent, largely lack the pragmatic depth required for effective academic argumentation. In particular, the AI-authored conclusions and discussions in the corpus exhibited a monologic

tone, with limited reader engagement and evaluative commentary. This rhetorical shortcoming is evident in the AI texts' minimal use of questions and direct address to the reader. For instance, where human-authored excerpts often include phrases such as *"what remains unclear, however, is…"* or *"a surprising result was…"* to guide reader attention and acknowledge gaps, the AI-generated conclusions tended to state outcomes without marking their significance or limitations. Additionally, the AI texts rarely positioned findings within broader scholarly conversations; they presented results as self-evident rather than as contributions to ongoing debates. This absence of evaluative commentary and metadiscursive framing suggests that while AI can replicate academic register at the sentence level, it struggles to construct the dialogic, reader-oriented stance that characterizes expert academic writing. This issue corroborates Jiang and Hyland (2025), who have maintained that AI-generated essays often lack the interactive stance-building typical of skilled academic writing.

The quantitative analysis revealed that engagement markers and attitude markers were significantly more frequent in the human-authored corpus than in the AI-generated corpus, with Chi-square tests confirming statistically significant differences for both marker types. Human authors used engagement markers nearly twice as often as the AI model, reflecting more explicit attempts to involve readers in dialogue. Attitude markers appeared considerably more often in the human texts, indicating a stronger evaluative stance and critical engagement with the findings. These patterns are consistent with Hyland's (2005) model of interpersonal metadiscourse, which positions such markers as central to constructing a reader-aware, dialogic academic voice.

The qualitative analysis further illuminated why these quantitative disparities matter. In the AI-generated texts, engagement and attitude markers were often deployed formulaically and without meaningful integration into the surrounding argument. Engagement markers such as *you can see* were repeated mechanically across texts, and attitude markers like *interestingly* were attached to vague, unsupported claims. By contrast, the human-authored texts embedded these markers within context-sensitive rhetorical moves, using them to guide reader interpretation, signal interpretive significance, and maintain dialogic engagement throughout the discourse. This disjuncture between surface-level presence and pragmatic function suggests that while large language models can approximate the lexical features of academic writing, they may not acquire the rhetorical competence to deploy metadiscourse strategically in service of argumentation and reader relations.

The functional underuse of engagement and attitude markers in the AI-generated conclusions and discussion sections points to a broader issue about the limits of large language models in replicating the interpersonal dimension of academic writing. While these texts demonstrate syntactic control and discipline-specific lexis, they may lack the pragmatic competence required to position claims within a shared intellectual space with the reader. Engagement markers such as *consider* or *note that* are not merely stylistic additions but rhetorical tools for guiding audience interpretation and constructing solidarity. Similarly, attitude markers

like *notably* or *unfortunately* do not simply convey emotions; they signal what merits attention, what challenges expectations, and how findings should be evaluated. Their absence or unnecessary insertion in AI-authored prose reflects an inability to perform the social function of academic argumentation. This issue suggests that current AI writing systems may operate within a monologic paradigm, treating texts as information containers rather than as sites of scholarly exchange. AI had a limited ability to replicate discourse-level intentions such as inviting reflection, indicating stance, or guiding reader interpretation. As s qualitative data analysis showed, human writers employed expressions like *consider* or *interestingly* to signal interpretive shifts, contextualize findings, or invite dialogue.

In contrast, the AI outputs relied on generic formulations (e.g., *interestingly, the results were positive*) that lacked rhetorical precision and contextual sensitivity. This issue highlights that while AI can identify which lexical items are associated with academic writing, it cannot yet grasp the pragmatic conditions under which such items are rhetorically meaningful. The result is a form of register approximation that mimics scholarly style at the sentence level. Still, it fails to enact the reader-oriented, dialogic stance that distinguishes expert academic prose.

ChatGPT (GPT-4 version), as used in the current study, generated coherent drafts with appropriate syntactic structure and discipline-specific vocabulary, but it failed to consistently observe the interpersonal and epistemic conventions required for effective academic argumentation. The results of the present investigation underscore the challenges in relying on AI for high-stakes academic writing. While tools like ChatGPT can generate coherent drafts, they often fail to embody the genre-specific conventions required for persuasive and credible scholarship. This point reinforces the call from Floridi and Chiriatti (2020) and Amirjalili et al. (2024) for critical human oversight in AI-mediated academic production. Furthermore, the disparity in metadiscourse use observed in this study, particularly the underuse of engagement and attitude markers and the overuse of self-mentions, suggests a need for more refined training data and model fine-tuning. However, the present analysis was limited to a single AI model (GPT-4 version) and a specific genre: conclusions and discussion sections of academic articles in applied linguistics. Whether these patterns persist across other AI models, disciplines, or writing tasks remains an open question. Within these boundaries, the findings indicate that incorporating discipline-specific rhetorical patterns and genre-sensitive prompts could enhance AI's ability to simulate human-like argumentation and stance-taking. In closing, perhaps until such advancements occur and are validated across diverse contexts, L2 educators and researchers must remain cautious about AI's role in academic authorship, particularly in evaluative and interpretive genres where rhetorical nuance and reader awareness are paramount.

Conclusion

Interactional metadiscourse markers, such as attitude markers, hedges, self-mentions, and engagement markers, are widely recognized as potential resources for effective academic writing in applied linguistics and related fields, enabling writers to negotiate claims, engage

readers, and position themselves within disciplinary communities (Hyland, 2019). The present research examined the rhetorical use of interactional metadiscourse markers in AI-generated versus human-authored academic discussion and conclusion sections. The findings revealed notable differences in how these markers were employed, reflecting broader contrasts in rhetorical awareness, communicative intention, and audience engagement between the two types of data, namely, two corpora: one consisting of human-authored conclusion and discussion sections drawn from published applied linguistics articles, and another comprising AI-generated texts by ChatGPT (GPT-4 version) in response to prompts designed to replicate comparable writing tasks.

Based on the results, hedges, attitude, and engagement markers appeared significantly more frequently in the human-authored corpus, while self-mentions appeared more often in the AI-generated corpus, with boosters showing no significant difference between the two. This quantitative disparity in interactional metadiscourse use contributes to the coherence patterns observed in the qualitative analysis. The AI-generated academic writing in the corpus exhibited a high degree of surface-level fluency and structural coherence, often conforming to grammatical norms and organizational patterns typical of academic discourse. However, ChatGPT (GPT-4 version) tended to underutilize key interactional metadiscourse elements, particularly hedges, attitude markers, and engagement markers, that play a crucial role in signaling stance, building writer-reader relationships, and managing the dialogic nature of academic argumentation. As a result, the AI texts often appeared rigid, overly assertive, or impersonal, lacking the rhetorical nuance expected in scholarly communication.

In contrast, human-authored texts demonstrated a more strategic and context-sensitive use of metadiscourse, reflecting a deeper understanding of academic conventions and audience expectations. Human writers were more likely to use hedging to express caution and epistemic humility, attitude markers to signal evaluation, and direct reference or rhetorical questioning to engage the reader. These rhetorical moves serve not only to position the writer within a broader scholarly conversation but also to anticipate and address potential reader responses, an aspect of discourse that AI has not yet been able to replicate effectively.

The findings have implications for both writing pedagogy and the development of AI-assisted writing tools. For L2 educators, the findings highlight the importance of teaching the interpersonal and rhetorical dimensions of academic writing. For researchers in natural language processing (NLP), the results suggest a need to refine AI models further to account for rhetorical context, audience awareness, and genre-specific conventions. This point may involve training models on more discipline-specific corpora and integrating pragmatic and discourse-level features into AI systems. Additionally, the findings add to the expanding body of literature about AI and academic discourse by focusing on underexplored textual zones, discussion and conclusion sections, where rhetorical nuance and authorial voice are especially salient. Ultimately, the results of the current investigation suggest that while AI can serve as a valuable aid in drafting and idea generation, it cannot yet replace the rhetorical sensitivity and

communicative intentionality of a skilled human writer, especially in evaluative and interpretive sections such as discussions and conclusions. Human oversight remains essential to ensure that AI-generated texts in applied linguistics meet the clarity, credibility, and scholarly engagement expected in academic discourse.

The present study has certain limitations. The analysis was restricted to a single AI model (ChatGPT, GPT-4 version) and a relatively small, discipline-specific corpus of conclusion and discussion sections. This point limits the generalizability of the findings. Additionally, this study used a version of ChatGPT (GPT-4), which may have limited prompt-engineering capabilities due to constraints on AI in Iran. This issue might have restricted the range and sophistication of the AI-generated outputs. Future research could validate and expand this analysis across disciplines by employing more AI models and sophisticated interfaces to investigate whether metadiscourse conventions vary across fields.

Additionally, longitudinal studies tracking improvements in newer AI models may help determine whether rhetorical performance develops over time. Moreover, because the corpus sizes were nearly identical, we conducted direct frequency comparisons. Raw frequencies and normalized frequencies per 1,000 words (a standard sample of the data or a unit) were reported. Nonetheless, normalization of the raw data would provide a more robust basis for comparison. We recommend that future studies apply normalization techniques to the entire dataset.

Acknowledgments

We thank the editorial team of TESL Quarterly for the opportunity to submit and publish this synthesis. We also thank the anonymous reviewers for their careful, detailed reading of our manuscript and their insightful comments and suggestions. We also acknowledge all the participants who took part in this study.

Declaration of conflicting interests

The authors declare no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for this article's research, authorship, and/or publication.

References

AbdAlgane, M., Ali, R., Othman, K., Ibrahim, I. Z. A., Alhaj, M. K. M., MT, E., & Ali, F. S. A. (2026). Exploring AI-generated texts vs. human-written texts in EFL academic writing: A case study of Qassim University in Saudi Arabia. *World Journal of English Language*, *16*(2), 114-131.

Ädel, A., & Mauranen, A. (2020). *Metadiscourse: Exploring interaction in writing*. John Benjamins Publishing Company.

Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education, 9*, Article 1347421. https://doi.org/10.3389/feduc.2024.1347421

Anthony, L. (2022). *AntConc* (Version 4.2.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Baofang, L., & Shamsudin, S. (2023). A corpus-based analysis of interactive metadiscourse in chemistry theses. *Journal of English for Academic Purposes, 61*, Article 101222. https://doi.org/10.1016/j.jeap.2023.101222

Benelhadj, F. (2025). How do master's students link their ideas? A comparative lexico-grammatical analysis of conjunctions in professional master's dissertations and their cited research articles. *Journal of Research in Applied Linguistics*, *16*(2), 63-79. https://doi.org/10.22055/rals.2025.50020.3571

Bhatia, V. K. (2020). *Academic discourse: Genre, rhetoric and identity*. Routledge.

Biber, D., & Conrad, S. (2021). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., Egbert, J., & Gray, B. (2023). AI writing tools and academic discourse: Challenges and opportunities. *Written Communication, 40*(1), 7–36. https://doi.org/10.1177/07410883221101923

Bruce, I. (2022). *Constructing critical stance in academic writing: Authority and voice*. Bloomsbury Academic.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Dahl, T. (2021). Metadiscourse in research article conclusions across disciplines. *Linguistics and Education, 64*, Article 100942. https://doi.org/10.1016/j.linged.2021.100942

Dreyfus, S., Humphrey, S., Mahboob, A., & Martin, J. R. (2021). *Genre pedagogy in higher education: The SLATE project*. Palgrave Macmillan.

Esfandiari, R., & Allaf-Akbary, M. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia, 14*, 51. https://doi.org/10.1186/s40468-024-00326-9

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*, 681–694. https://doi.org/10.1007/s11023-020-09548-1

Gillaerts, P., & Van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes, 9*(2), 128–139. https://doi.org/10.1016/j.jeap.2010.02.004

Hunston, S. (1994). Evaluation and organization in a sample of written academic discourse. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 191–218). Routledge.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.

Hyland, K. (2011). Disciplines and discourses: Social interactions in the construction of knowledge. In D. Starke-Meyerring et al. (Eds.), *Writing in Knowledge Societies* (pp. 171–192). The WAC Clearinghouse; Parlor Press.

Hyland, K. (2019). *Second language writing*. Cambridge University Press.

Jiang, F., & Hyland, K. (2025). Metadiscourse in AI-generated versus student academic writing. *Journal of English for Academic Purposes, 62*, Article 101273. https://doi.org/10.1016/j.jeap.2024.101273

Khajavi, Y., & Ezhdehakosh, M. (2025). Enhancing language teacher education: exploring the integration of ai-powered tools in preservice teacher education. *Journal of Research in Applied Linguistics*, *16*(1), 174-191. https://doi.org/10.22055/rals.2024.48226.3405

Khedri, M., & Kritsis, K. (2018). Metadiscourse in applied linguistics and chemistry research article introductions. *Journal of Research in Applied Linguistics*, *9*(2), 47-73. https://doi.org/10.22055/rals.2018.13793

Kuteeva, M., & Mauranen, A. (2022). Genre-specific use of metadiscourse in AI-generated and human-authored texts. *English for Specific Purposes, 65*, 112–127. https://doi.org/10.1016/j.esp.2022.01.003

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Li, Y. (2023). ChatGPT as a writing aid for non-native speakers: Pedagogical perspectives. *Language Learning & Technology, 27*(1), 45–62. https://doi.org/10.1016/j.langlt.2023.01.005

OpenAI. (2023). Introducing ChatGPT. https://openai.com/chatgpt

Pearson, L., & Abdollahzadeh, E. (2023). Metadiscourse in academic writing: A systematic review. *Journal of English for Academic Purposes, 62*, Article 101262. https://doi.org/10.1016/j.jeap.2024.101262

Pham, T., & Wang, Y. (2023). AI in academic writing: A corpus-based comparison of metadiscourse. *Language Teaching Research, 27*(2), 189–211. https://doi.org/10.1177/13621688221112345

Shalevska, A. (2024). Hedges and boosters in AI vs. human-written essays. *Applied Linguistics Review, 15*(1), 51–68. https://doi.org/10.1515/applirev-2023-0021

Sudajit-apa, M. (2025). Dismantling the discursive representation of women in AI-generated life-changing narratives: A critical discourse analysis. *Journal of Research in Applied Linguistics*, *16*(1), 38-54. https://doi.org/10.22055/rals.2024.46100.3232

Swales, J. M., & Feak, C. B. (2021). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). University of Michigan Press.

Tang, M. (2023). Exploring rhetorical strategies in AI-generated scholarly writing. *Discourse Studies, 25*(1), 104–122. https://doi.org/10.1177/14614456221112345

Xu, L. (2025). AI and genre awareness: A comparative analysis of research abstracts. *ESP Today*, *13*(1), 33–52. https://doi.org/10.18485/esptoday.2025.13.1.2

Yu, J., & Chen, L. (2024). The rhetorical shortcomings of AI-generated academic texts. *Computers and Composition, 67*, 102788. https://doi.org/10.1016/j.compcom.2023.102788

Zeng, Q., Luo, Y., & Hu, G. (2024). Analyzing self-mention use in AI-generated academic writing. *Journal of Second Language Writing, 63*, Article 101847. https://doi.org/10.1016/j.jslw.2023.101847

Zhang, W., & Liao, X. (2023). Enhancing writing fluency through AI assistance: Insights from EFL classrooms. *System, 114*, 102967. https://doi.org/10.1016/j.system.2023.102967