



Research on Frontiers

ARTICLE

A Systemic Functional Linguistic Analysis of ChatGPT's Academic Writing Responses: Metafunctional Patterns and Implications for Writing Pedagogy

Taimur Khan, Feng Lei

DOI: <https://doi.org/10.62978/2601tl100115>



North American Chinese Research Press

A Systemic Functional Linguistic Analysis of ChatGPT's Academic Writing Responses: Metafunctional Patterns and Implications for Writing Pedagogy

Taimur Khan, Feng Lei*

School of Languages and Communication Studies, Beijing Jiaotong University, Beijing 100044, P. R. China.

lfeng@bjtu.edu.cn

Abstract: Generative artificial intelligence (AI), particularly ChatGPT, is having a significant impact on how academic writing is produced, taught and assessed in higher education. This study uses Systemic Functional Linguistics (SFL) to analyze how ChatGPT constructs meaning when it responds to academic writing prompts. Drawing on Halliday's three metafunctions (ideational, interpersonal and textual), it examines ten ChatGPT-generated texts representing five academic part-genres: the research introduction, the argumentative essay, the literature review, the methodology section and the discussion section. The analysis shows that ChatGPT produces highly formal, register-appropriate academic discourse characterized by dense nominalization, median- and low-value modality and formulaic evaluation. It makes sophisticated use of transitivity, mood and thematic progression, yet it constructs an "epistemic space of safe neutrality" in which alternative views are acknowledged but, where argument is most expected, left unresolved. Cross-genre analysis indicates that discussion sections have the highest modal density and nominalization and the most heavily hedged interpretive claims, argumentative essays leave their concessions unresolved, and methodology sections show the most procedural sequencing. These findings have implications for writing pedagogy and assessment at a time when machines can reproduce the form of scholarship with great fidelity. The study argues that SFL gives teachers an explicit metalanguage for educating students as critical consumers of language, able not only to produce texts but also to understand how meaning is constructed, negotiated and contested in and through language.

Keywords: Systemic Functional Linguistics; ChatGPT; Academic Writing; Metafunctions; Writing Pedagogy

1. Introduction

Many academics will recognize the scene: a student types a carefully worded prompt into ChatGPT and, within seconds, receives several paragraphs of coherent prose. The text is grammatically correct, structurally conventional and relevant to its topic; it cites no sources, yet it reads as though its author has read everything. The moment is both impressive and unsettling. It raises the familiar question of whether students will submit such text as their own, but also a less obvious one: what does it mean that a machine can reproduce the form of academic discourse so convincingly while remaining silent on the substance of scholarly inquiry?

Generative artificial intelligence (AI) tools based on large language models (LLMs) have not only given writers a new resource but have also challenged established notions of authorship, originality and scholarly work.^{1,2} Students now use them for brainstorming, drafting and many other academic tasks, and it is often difficult to distinguish what the student has contributed from what the machine has produced.³ The central question is therefore not whether AI should be part of academic writing, but how we can understand what it produces and teach students to think critically about it.

*Corresponding Author

This study adopts Systemic Functional Linguistics (SFL), developed by Halliday and extended by Martin and White in their work on evaluative language⁴ and by Halliday and Matthiessen in their account of functional grammar.⁵ SFL treats language as a resource for making meaning in social contexts rather than as a set of rules. It holds that all language use simultaneously produces ideational meaning, which construes experience; interpersonal meaning, which enacts social relationships; and textual meaning, which organizes language into coherent discourse. These metafunctions are strands of meaning woven together in every clause. As Banks has argued, the value of SFL in the digital era lies in its ability to relate linguistic form to social function.⁶

Discussion of AI-generated text is often narrowed to whether a reader can tell the difference between human and machine writing, a question evaluated through classification accuracy, stylometric features or rater judgment.⁷ For writing pedagogy, SFL reformulates the question more productively: instead of asking whether ChatGPT's text can be distinguished from a human's, it asks what kind of meaning the text makes, even when it is indistinguishable from human writing. The pedagogical problem raised by generative AI is not, after all, a forensic one: a text that is indistinguishable from human writing may still have a fundamentally different relationship to knowledge, evidence and readership from the kind of text it resembles. It may be competent academic prose and yet perform none of the interpersonal work that academic prose is supposed to do: staking a position, risking being wrong and engaging a specific reader. This gap motivates the present study's move from the detection paradigm to a metafunctional paradigm. Accordingly, the study analyzes ten ChatGPT responses to academic writing prompts across five part-genres through the lens of the three metafunctions, compares these patterns across genres, and consider what they reveal about "safe neutrality" and its implications for writing pedagogy.

This study moves beyond the polarized debate over whether AI should be banned or embraced, offering instead a linguistically informed account of what academic discourse looks like when it is produced by AI. It is relevant to teachers who must design assessments when students can use AI tools, to students who need to use these tools critically, and to an academic community that is redefining legitimate research practice. It also extends SFL to an emerging digital discourse.

The study is timely. Malik et al. found that 88% of 245 Indonesian undergraduates had used ChatGPT in their academic writing, and although they reported gains in efficiency and quality, many were concerned about originality and critical thinking.¹ Delcker et al. found, in a study of 634 first-year students, that attitudes toward AI appeared to predict its use for learning.² The concern extends to publishing: Gulumbe reports AI-generated anomalies in peer-reviewed journal articles, suggesting that recognizing them is not yet part of editors' and reviewers' linguistic literacy.³ Like students, editors and reviewers can benefit from an explicit metalanguage for identifying the metafunctional characteristics of machine-generated discourse. The pedagogical tools developed in this study therefore have implications not only for the writing classroom but also for the wider academic community that validates scholarly knowledge.

2. Literature Review

SFL has long been used to study academic discourse across educational levels and disciplines because it links linguistic choice to social meaning. It has guided students' academic communication in multicultural second language (L2) contexts⁸ and traced students' movement from everyday to academic ways of meaning.⁹ Systematic reviews show a strong focus on the ideational metafunction, especially transitivity, and call for more studies of the interpersonal and textual metafunctions.¹⁰

2.1 Systemic Functional Linguistics: Register and the Three Metafunctions

SFL approaches language as a social semiotic whose organization reflects the social purposes it serves.^{5,11} Grammar is modeled as networks of options, and every text is the outcome of choices made in a particular context. Register theory describes that context through three variables: field, the social activity and subject matter; tenor, the relationship between participants; and mode, the role and channel of language.⁵ Each variable is realized by one metafunction. Field is realized ideationally, mainly through transitivity, that is, process types (material, mental, relational, verbal, behavioral and existential) with their participants and circumstances. Tenor is realized interpersonally through mood, modality and, in Martin and White's extension, the appraisal systems of attitude, engagement and graduation.⁴ Mode is realized textually through Theme–Rheme structure, thematic progression¹² and the cohesive resources of reference, conjunction and lexical cohesion.¹³ Because context and language correspond in this way, SFL can explain, not merely describe, why a text sounds as it does, which makes it well suited to text produced by a system that has no social context of its own.

2.2 The Metafunctions in Academic Writing

Ideationally, academic discourse is distinguished by grammatical metaphor, particularly nominalization, whereby processes and qualities are realized as nouns: *students revise their drafts after they receive feedback* becomes *post-feedback revision*.^{14,15} Nominalization packs information into nominal groups, raises lexical density and allows processes to be classified and evaluated as “things”, typically through relational clauses; academic writing therefore favors relational processes and agentless or nominalized material processes.¹⁶ Control of grammatical metaphor is widely regarded as a key developmental threshold in academic literacy.¹⁷

Interpersonally, academic writing is overwhelmingly declarative but far from neutral. Writers grade commitment through modality and position themselves among other voices through engagement, which either expands the dialogic space (entertain, attribute) or contracts it (deny, counter, concur, pronounce, endorse).⁴ Hood shows that expert writers acknowledge alternative views and then resolve them in favor of their own position.¹⁸ Hyland's corpus studies reach a compatible conclusion: hedging is a central resource for negotiating the status of claims, and expert writers hedge selectively, more around contestable and central claims than around established or peripheral ones.^{19,20} Argumentation is thus an interpersonal achievement, the calibrated interplay of commitment and concession.

Textually, academic writing favors unmarked topical Themes, textual Themes that signal logical relations (*however, therefore*) and thematic progression that is constant (the same Theme is kept), linear (the Rheme of one clause becomes the next Theme) or derived from a hyper-Theme.^{12,21} At text level, periodicity describes how macro- and hyper-Themes predict what is to come and hyper-News consolidate what has been argued.²¹ Argument becomes textually visible when new information accumulates into a position that the writer carries forward.

2.3 Linguistic Characteristics of AI-Generated Text: A Metafunctional Reading

Research on AI-generated writing has relied mainly on corpus-based, multidimensional or stylometric methods and has seldom interpreted its findings functionally. Read through the three metafunctions, however, it points to a consistent profile. Herbold et al. found that ChatGPT's argumentative essays were rated higher in quality than student essays and contained more nominalizations and more complex sentences, whereas the student essays used more modal and epistemic constructions.⁷ In SFL terms, this indicates heavier ideational grammatical metaphor, the hallmark of written technical registers,¹⁴ and weaker interpersonal modality.

Ali compared dissertation introductions by Saudi doctoral students with introductions produced by ChatGPT, using multidimensional analysis of more than 150 linguistic features, and found the ChatGPT texts more informational and explicit but less abstract than the human ones.²² Informational density is ideational, while explicitness corresponds to textual conjunction and reference.¹³ Ali's finding of lower abstraction, however, sits uneasily with Herbold et al.'s finding of heavier nominalization, raising a question that SFL is well placed to answer: does the nominal density of AI text construe field-specific knowledge, or is it a generic nominal style that merely sounds academic?

Mizumoto et al. compared essays by Japanese learners of English as a foreign language (EFL) with ChatGPT essays written to the same prompt and found significant differences on all nine measures examined, including nominalizations, modals, epistemic markers and discourse markers; a classifier based on these measures separated the groups almost perfectly.²³ Several of these measures realize interpersonal (modality, engagement) and textual (conjunction) meanings, suggesting that machine and human writing diverge most in how stance is enacted and discourse organized. Because a single prompt was used, however, it remains unclear whether the differences hold across genres.

These studies reveal a discontinuity between statistical distinguishability and perceived quality: AI text differs measurably from human writing, yet evaluators may rate it more highly.⁷ This discontinuity is pivotal for pedagogy: when impressionistic reading is an unreliable guide to machine authorship, a metafunctional analysis offers a principled alternative that focuses on how meaning is built rather than on how a text happens to read. Moreover, register theory predicts that ChatGPT's linguistic profile should vary with the genre it is asked to produce, so its discourse should be described genre by genre, and patterns such as nominalization density, modal distribution and thematic progression treated as genre-shaped tendencies. Above all, SFL replaces the question of whether a human or a machine wrote a text with a more pedagogically useful one: what kind of meaning the text makes, and what it models for a student writer.

2.4 SFL, Metadiscourse and ChatGPT's Academic Discourse

A small body of work now examines AI-generated academic writing functionally. Comparing ChatGPT and student argumentative essays, Jiang and Hyland found that ChatGPT used fewer lexical bundles, relying on noun- and preposition-based bundles for abstract description and signposting, whereas students showed more epistemic stance and authorial presence;²⁴ that ChatGPT used significantly fewer interactional metadiscourse markers such as hedges, boosters and attitude markers, relying instead on transitions;²⁵ and that it used fewer engagement markers, particularly questions and personal asides.²⁶ Although these studies use Hyland's metadiscourse model, their categories correspond closely to SFL systems: hedges and boosters realize modality and graduation, and engagement markers realize engagement. The present study therefore reads their findings through SFL. Within SFL itself, Zein et al. found that ChatGPT's academic responses relied heavily on declarative mood, whereas its casual responses contained more appraisal resources, suggesting sensitivity to tenor.²⁷

Other research has examined ChatGPT's effects on learners from a pedagogical perspective. Mahapatra found that English as a second language (ESL) students' writing improved across three task types when ChatGPT provided feedback rather than drafted texts,²⁸ and Lai and Trinh found positive effects on English majors' vocabulary, accuracy, idea generation and motivation, while cautioning against over-reliance.²⁹ These studies, however, examine what ChatGPT does for learners rather than the discourse it produces. The functional studies, for their part, show that AI and human writing differ less in whether stance resources are present than in how they are deployed, but they leave three gaps: they count markers without relating them to ideational and textual choices; they focus on student essays rather than

research part-genres; and they offer no theoretical account of the apparent neutrality and lack of argumentation that teachers observe in AI prose.

2.5 Toward a Metafunctional Account of AI Academic Discourse

This study proposes that “safe neutrality” and “lack of argumentation” can be described precisely as configurations of metafunctional choice (Table 1). “Safe neutrality” is primarily interpersonal: (a) modality clusters at median and low values; (b) hedging is applied similarly across claims rather than concentrated on contestable ones;¹⁹ (c) engagement entertains alternatives and disclaims only extreme positions, while pronouncement is absent; and (d) attitude is realized mainly through generic appreciation. It is reinforced ideationally when nominalization removes the human agents who could be held to a position. “Lack of argumentation” is interpersonal and textual: concessions are not followed by a counter or pronouncement, so alternative views are acknowledged but never resolved,¹⁸ and thematic progression links information without accumulating it, so that conclusions restate rather than advance. These definitions determine the categories of the analytical framework (Section 3.3).

Table 1 Metafunctional Indicators of AI-Generated Academic Discourse

Property or Feature	Metafunction	SFL Resources	Indicators Analyzed
Informational density ^{7,22}	Ideational	Grammatical metaphor; transitivity	Nominalizations per 100 words; process-type distribution
Explicitness and strong coherence ^{22,25}	Textual	Conjunction; reference; lexical cohesion	Conjunctions by type; lexical chains
Reduced modal and stance marking ^{7,23,25}	Interpersonal	Modality (type and value); attitude	Modal tokens by type and value; attitude by sub-type
Reduced reader engagement ²⁶	Interpersonal	Mood; engagement	Mood choice; expansive versus contractive engagement
Safe neutrality	Interpersonal (and ideational)	Modality value; engagement; appreciation; agentless nominalization	Median/low modality; concession without pronouncement; generic appreciation
Lack of argumentation	Interpersonal and textual	Concession-counter patterns; thematic progression; periodicity	Unresolved concessions; progression types; conclusions that restate

2.6 Research Gap and Research Questions

Existing research shows that ChatGPT’s academic prose differs from human writing in nominal density, stance and engagement, but mainly through detection-oriented or metadiscourse-based methods and within single genres. No study has yet explained how the three metafunctions jointly produce the qualities teachers observe in AI text, or how these vary across academic part-genres. The study therefore asks the following research questions (RQs):

RQ1. How does ChatGPT realize the interpersonal and metafunction in its academic writing responses?

RQ2. How do these metafunctional vary across five academic part genres: Introduction, argumentative essay, Literature review, methodology and discussion.

RQ3. To what extent do these patterns construct “safe neutrality” and lack of argumentation, and what are the implications for academic writing pedagogy and assessment?

3. Methodology

3.1 Research Design

The study employs qualitative discourse analysis grounded in SFL. Following established principles of SFL-informed research, it examines texts in their social and cultural contexts, exploring how linguistic choices both shape and are shaped by register and genre. The clause-by-clause analysis is supported by descriptive frequency counts normalized per 100 words.

3.2 Data Collection

The corpus consists of ten ChatGPT responses to five academic writing prompts. Each prompt was submitted twice in separate chat sessions to ensure independent generation, yielding two responses (R1 and R2) per prompt. Each prompt asked for a specific academic part-genre of a stated length on a topic related to AI in education, and every prompt ended with the same instructions (“Use formal academic register. Do not include citations.”), so that field, tenor and mode were held constant and differences between texts could be related to genre. The introduction prompt also specified the three moves of the Create a Research Space (CARS) model.³⁰ The responses were generated with ChatGPT version 4.0. The corpus totals 1,934 words, 208 clauses and 95 T-units (Table 2). Texts are identified by prompt and response; for example, P2-R1 is the first response to the argumentative essay prompt.

Table 2 Composition of the Corpus

Prompt	Genre and Task	Requested Length	Words (R1/R2)	Clauses (R1/R2)	T-Units (R1/R2)
P1	Research introduction on AI and academic writing (CARS model)	150	149/153	16/13	6/7
P2	Balanced argumentative essay on student use of AI tools	200	208/207	28/29	11/11
P3	Literature review of research on AI in language education (2020–2025)	200	207/218	15/18	10/10
P4	Mixed-methods methodology section on student use of AI writing tools	200	194/193	21/22	10/10
P5	Discussion of findings on AI feedback and independent revision	200	203/202	24/22	10/10
Total	10 texts	1,900	1,934	208	95

3.3 Analytical Framework

The framework operationalizes the model in Section 2.5: each metafunction is analyzed through the systems identified in Section 2.2, and each system is linked to a research question (Table 3). Two units of analysis were used. For transitivity, the unit is the clause with its own verbal group, comprising all finite clauses and the non-finite clauses that function adverbially or as verb complements; non-finite postmodifiers of nouns were treated as part of the nominal group.⁵ For Theme and thematic progression, the unit is the T-unit, that is, an independent clause together with its dependent clauses.

Table 3 Analytical Framework

Metafunction	System	Categories Coded	Key Source	RQ
Ideational	Transitivity	Material, mental, relational, verbal, behavioral and existential processes	Halliday and Matthiessen ⁵	RQ1
Ideational	Grammatical Metaphor	Nouns agnate with a verb or adjective (derived or converted forms, noun uses only)	Halliday and Martin ¹⁴	RQ1
Interpersonal	Mood	Declarative, interrogative, imperative	Halliday and Matthiessen ⁵	RQ1
Interpersonal	Modality	Epistemic, deontic and dynamic; value: high, median, low	Halliday and Matthiessen ⁵	RQ1
Interpersonal	Attitude	Affect, judgment, appreciation	Martin and White ⁴	RQ1
Interpersonal	Engagement	Entertain; disclaim (deny, counter); pronounce; concession resolved or unresolved	Martin and White ⁴	RQ1
Textual	Theme	Textual and topical Theme; marked versus unmarked topical Theme	Halliday and Matthiessen ⁵	RQ3
Textual	Thematic Progression	Constant, linear, derived, new	Daneš ¹²	RQ3
Textual	Cohesion	Adversative, additive, causal and sequential conjunctions; lexical chains	Halliday and Hasan ¹³	RQ1
All	Genre Staging	CARS moves (P1); functional stages of P2–P5	Swales ³⁰ ; Martin and Rose ²¹	RQ2

(1) Ideational Analysis

Each clause was coded for process type, participants and circumstances.⁵ The six process types were distinguished by their core meaning: material (doing and happening), mental (sensing and thinking), relational (being and having), verbal (saying, including reporting by non-human sources such as *studies report*), behavioral and existential (*there is*; *gaps remain*). Examples from the corpus include material *AI tools can enhance learning* (P2-R1), mental *It aims to evaluate both the opportunities and challenges...* (P1-R1), relational *A dominant theme across the literature is the use of AI-powered tools...* (P3-R1), verbal *Studies consistently report improvements...* (P3-R1) and existential *There is also a lack of longitudinal research...* (P3-R1). Example 1 shows how a single clause was analyzed for all three metafunctions (Table 4).

Example 1: *On the other hand, excessive dependence on AI may weaken students' critical thinking, creativity, and independent writing abilities.* (P2-R1)

Table 4 Metafunctional Coding of Example 1

Layer	On the Other Hand	Excessive Dependence on AI	May	Weaken	Students' Critical Thinking, Creativity...
Transitivity	–	Actor	–	Process: Material	Goal
Grammatical Metaphor	–	Nominalization ← <i>students depend on AI excessively</i>	–	–	Nominalizations ← <i>students think critically, create...</i>
Mood and Modality	–	Subject	Finite: epistemic, low value	Predicator	Complement
Theme	Textual Theme	Topical Theme (unmarked)	Rheme	Rheme	Rheme

Unpacked congruently, the clause reads *if students depend on AI too much, they may not learn to think critically*, and the human participants reappear. Nominalizations were identified by an agnation test: a noun was counted when it had a congruent verbal or adjectival agnate (e.g., *dependence* ← *depend*; *integrity* ← *be honest*). Verb uses of ambiguous forms (e.g., *tools support*) and lexicalized nouns without a process meaning (e.g., *university*, *intelligence*) were excluded.

(2) Interpersonal Analysis

Mood was coded for each T-unit. Modal operators were coded by type, as epistemic (probability), deontic (obligation) or dynamic (ability), corresponding in SFL to modalization and modulation,⁵ and by value: high (e.g., *must*), median (e.g., *should*, *would*) or low (e.g., *may*). Attitude was coded as affect, judgment or appreciation, and engagement as entertain (e.g., *may*), disclaim (deny or counter, e.g., *rather than*), or pronounce (e.g., *indeed*).⁴ In Example 1, *may* is low-value epistemic modality coded as *entertain*. Example 2 illustrates deontic modality and disclaiming:

Example 2: *Therefore, students should be permitted to use AI tools under appropriate academic guidelines rather than without restrictions.* (P2-R1)

Here, *should* is median-value deontic modality, the passive *be permitted* leaves the obliging agent implicit, and *rather than without restrictions* disclaims an extreme position. Each adversative or concessive marker was coded by function: as a resolved concession if the conceded view was overridden by a counter, as an unresolved juxtaposition if opposing views were set side by side without resolution, as a contrast without concession, or, for *while* meaning “at the same time”, as non-adversative. This coding operationalizes “lack of argumentation” (Section 2.5) and thus provide evidences for RQ3.

(3) Textual Analysis

The Theme was identified as everything up to and including the first experiential element of the T-unit.⁵ Textual Themes (e.g., *however*, *on the other hand*) were distinguished from the topical Theme; a topical Theme conflated with the Subject was coded as unmarked, and a fronted circumstance or dependent clause as marked. Textual Themes do not make the topical Theme marked. In Example 1, *on the other hand* is a textual Theme and *excessive dependence on AI* an unmarked topical Theme.

Rheme (P1-R1)

Thematic progression between successive T-units was coded as constant (the Theme repeats the previous Theme), linear (the Theme takes up an element of the previous Rheme), derived (the Theme is a sub-topic of a hyper-Theme or of the text topic) or new (no thematic link).¹² In P1-R1, for example, the Themes run *AI* → *AI-powered tools* (derived) → *their growing adoption* (constant) → *existing discussions* (new) → *this study* (linear, taking up a need for a more balanced examination) → *it* (constant). Conjunctions were coded as adversative, additive, causal or sequential,¹³ and lexical chains were traced through repeated and related terms.

(4) Genre Staging

For RQ2, introductions (P1) were segmented into Swales's CARS moves,³⁰ and the other texts into their functional stages, such as thesis, arguments, counter-arguments and conclusion in P2.

3.4 Coding Procedure

All ten texts were coded manually, clause by clause and T-unit by T-unit, using the categories in Table 3, and every frequency reported in Section 4 was calculated from this coding of the complete corpus. Given the size of the corpus, the frequencies are treated as descriptive indications of tendency that support the qualitative analysis, not as grounds for statistical generalization.

4. Results

Table 5 summarizes the corpus-wide coding results by genre; the following sections interpret them by metafunction.

Table 5 Corpus-Wide Coding Results by Genre

Measure	P1 Intro	P2 Essay	P3 Lit. Review	P4 Method	P5 Discussion	Total
Words	302	415	425	387	405	1,934
Clauses	29	57	33	43	46	208
Material (%)	65.5	84.2	42.4	76.7	60.9	68.3
Verbal (%)	10.3	3.5	27.3	0.0	23.9	12.0
Relational (%)	10.3	7.0	15.2	11.6	4.3	9.1
Mental (%)	6.9	3.5	3.0	11.6	10.9	7.2
Existential (%)	6.9	1.8	12.1	0.0	0.0	3.4
Nominalizations per 100 Words	20.5	15.4	17.2	12.9	24.4	18.0
Modal Tokens (per 100 Words)	1 (0.33)	15 (3.61)	2 (0.47)	0 (0.00)	16 (3.95)	34 (1.76)
T-Units (Marked Themes)	13 (2)	22 (3)	20 (2)	20 (0)	20 (1)	95 (8)
T-Units with Textual Theme	1	9	1	2	2	15
Progression: Derived/Linear/Constant/New	4/2/3/2	5/2/7/6	11/5/0/2	15/1/2/0	9/4/0/5	44/14/12/15
Concessions: Resolved/Unresolved	2/0	1/3	2/0	0/0	3/0	8/3

4.1 Ideational Meanings

(1) Transitivity

Of the 208 clauses, 126 are finite and 82 non-finite. Material processes dominate (142 clauses, 68.3%), followed by verbal (25, 12.0%), relational (19, 9.1%), mental (15, 7.2%) and existential processes (7, 3.4%); no behavioral processes occur. This profile departs from the expectation that academic prose favors relational, classifying clauses (Section 2.2): ChatGPT construes academic knowledge mainly as actions performed by abstract, often nominalized participants, such as *AI tools can enhance learning* or *excessive reliance may limit the development of critical thinking*. The distribution varies by genre. The argumentative essays are the most material (84.2%), while the literature reviews use the most verbal (27.3%) and existential (12.1%) processes, as they attribute findings to sources (“Studies consistently report improvements...”) and assert that “gaps remain”. Verbal processes are also frequent in the discussions (23.9%). Across the corpus, 20 of the 25 verbal processes have non-human Sayers (e.g., *studies, research, findings, this pattern*), so that claims are attributed to sources rather than to the writer.

The opening sentence of P1-R1 illustrates the pattern: “Artificial intelligence (AI) has become an increasingly influential technology in higher education, transforming the ways students and educators approach academic writing.” The first clause is relational attributive (AI = technology); the non-finite clause *transforming the ways...* is material, with AI as the implicit Actor. Of the 16 clauses in P1-R1, 10 are material, 2 mental, 2 verbal, 1 relational and 1 existential.

(2) Nominalization

The corpus contains 348 nominalizations, or 18.0 per 100 words. They are densest in the discussions (24.4) and introductions (20.5), followed by the literature reviews (17.2), argumentative essays (15.4) and methodology sections (12.9). The most frequent are *writing* (47 occurrences), *learning* (16), *feedback* (15), *education* (11), *development* (10) and *use* (10). Such grammatical metaphor turns processes into things that can be evaluated and acted upon (*the observed decline in independent revision behavior raises concerns*), creating the abstract register of academic texts.¹⁴ The most frequent nominalizations, however, belong to general academic vocabulary rather than to a field-specific taxonomy, which helps to reconcile Herbold et al.’s finding of heavy nominalization⁷ with Ali’s finding of non-abstract discourse.²² ChatGPT’s texts are nominally dense without being technical.

4.2 Interpersonal Meanings

(1) Mood, Modality and the Construction of Stance

All 95 T-units are declarative; no interrogatives or imperatives occur. This suits the academic register but is interpersonally flat, with no mood variation to engage, challenge or invite the reader into dialogue. The corpus contains 34 modal tokens (1.76 per 100 words): *should* (13, 38.2%), *can* (11, 32.4%), *may* (8, 23.5%) and *would* (2, 5.9%). No high-value modal (*must, will*) occurs, and *might* and *could* are absent. Obligation is thus expressed only at median value (*should*), possibility mainly at low value (*may*), and capability through dynamic *can*.

Modal density varies sharply by genre. It is highest in the discussions (P5; 16 tokens, 3.95 per 100 words: 7 *should*, 4 *can*, 3 *may*, 2 *would*) and the argumentative essays (P2; 15 tokens, 3.61: 4 *should*, 6 *can*, 5 *may*), low in the literature reviews (P3; 2 *should*, 0.47) and introductions (P1; 1 *can*, 0.33), and absent from the methodology sections (P4). The resulting stance is prescriptive but not commanding, and cautious but not uncertain. In Example 2, ChatGPT combines median-value *should*, the agentless passive *be permitted* and a disclaimer of the extreme alternative (*rather than without restrictions*): the recommendation appears balanced but conveys no conviction of its own. Epistemic *may* behaves similarly. In “excessive dependence on AI may weaken students’ critical thinking” (Example 1), it does not appear to

mark genuine doubt; one plausible explanation is that the model has been trained to avoid disputable categorical statements. Such hedging appears algorithmic rather than epistemic, producing what may be called “empty hedging”: caution without substance.

(2) Appraisal and Engagement

ChatGPT relies heavily on appreciation, the evaluation of phenomena, processes and products.⁴ Positive appreciation draws on a small set of items: *significant* (4), *important* (3), *comprehensive* (3), *sustainable* (3), *equitable* (2), *meaningful* (2) and *authentic* (1); negative appreciation on another: *challenges* (6), *concerns* (6), *excessive* (4), *risks* (3) and *bias* (2). Judgment is realized almost only through *responsible* and *responsibly* (5 tokens), and affect through *confidence* (2). The evaluations are thus generic: when ChatGPT credits AI with “significant” effects, the evaluation is not anchored in disciplinary argument.

Engagement is expansive but not argumentative. Entertain is realized by *may* and *would* (10 tokens), and the disclaimer *rather than* occurs 9 times, in 8 of which it rejects an extreme position (*rather than replacing human expertise; rather than completely prohibited*). No pronouncement (e.g., *indeed, it is clear that*) occurs. Of the 17 adversative and concessive markers (*while* 9, *despite* 4, *however* 2, *on the other hand* 2), 4 instances of *while* express simultaneity rather than contrast, and 2 mark contrast without concession. Of the 11 concessive moves, 8 are resolved by a counter, but they cluster where genre conventions require them: the four *despite* concessions all establish a niche or research gap (*Despite the rapid growth of this field, several research gaps remain*), and the discussions counter positive findings with concerns. All 3 unresolved juxtapositions occur in the argumentative essays, the genre in which argument is most expected: *While these technologies offer substantial educational benefits, they also present challenges* (P2-R1), followed by *On the one hand... On the other hand* (both essays). The essays state a moderate thesis before the dialectic and restate it unchanged afterward (*students should be permitted to use AI tools under appropriate academic guidelines; AI tools should be viewed as educational assistants rather than substitutes*), so the concessions do not develop the argument.

4.3 Textual Meanings

(1) Theme and Thematic Progression

Unmarked Themes predominate: 87 of the 95 T-units (91.6%) begin with the Subject as topical Theme. All 8 marked Themes are concessive, contrastive or conditional framings (*Despite... 4; While... 2; When used responsibly; Rather than replacing...*), so marked Themes are used almost exclusively to stage concession rather than for emphasis or shifts in perspective. Textual Themes open 15 T-units (15.8%), 9 of them in the argumentative essays (*On the one hand, On the other hand, Therefore, Consequently, Furthermore, In conclusion*). The resulting information flow is smooth and predictable.

Derived progression dominates the 85 transitions between T-units (44, 51.8%), followed by new (15, 17.6%), linear (14, 16.5%) and constant Themes (12, 14.1%). ChatGPT thus organizes text mainly by listing sub-topics of a hyper-Theme, most clearly in the methodology sections (15 of 18 transitions derived: *this study → the quantitative component → participants → the qualitative component → data collection*) and literature reviews (11 of 18: *a dominant theme → studies → another prominent theme → research*). Linear progression, which moves new information forward into a chain of reasoning, accounts for only one in six transitions. The argumentative essays show the most constant (7) and new (6) Themes, alternating between *AI tools* and new topics such as *students* and *clear institutional guidelines* rather than building each point on the last.

(2) Cohesion and Conjunction

Conjunction is explicit: besides the 17 adversative markers analyzed above, the corpus contains 11 additive (*also, furthermore*), 5 causal (*therefore, consequently, as a result, thereby*) and 4 sequential conjunctions (*first, second, and in conclusion* twice). Adversatives are most frequent in the argumentative essays (6), followed by the discussions (5), literature reviews (3), introductions (2) and methodology sections (1). Lexical cohesion is consistent but shallow. Chains of general terms such as *AI tools, writing support, feedback* and *learning outcomes* are maintained throughout, but the texts do not build the theoretically loaded chains of expert writing (e.g., *scaffolding, zone of proximal development, mediated learning*). The cohesion is correct but not resonant.

4.4 Cross-Genre Variation

The five genres show distinct configurations (Table 5). The introductions (P1) follow the CARS model mechanically: establishing a territory (AI has become influential), establishing a niche through a resolved concession (*Despite their growing adoption, important questions remain*) and occupying the niche (*This study addresses this gap*). The stated purpose (“a more balanced examination”) is descriptive rather than argumentative. The argumentative essays (P2) are the most material and among the most modalized texts, signal a thesis–antithesis–conclusion structure through textual Themes, and leave their concessions unresolved. The literature reviews (P3) attribute knowledge to sources through verbal processes, assert gaps through existential processes and progress by derivation from a hyper-Theme. The methodology sections (P4) contain no modals, the lowest nominalization density (12.9 per 100 words) and the most derived progression, which suits a genre in which the study is the fixed point of reference. The discussions (P5) have the highest modal density (3.95) and nominalization (24.4 per 100 words), combining *may, would* and *suggest* as they move from findings to implications and future research; a claim such as “excessive reliance may limit the development of critical thinking” is so hedged that it risks saying nothing. The two responses to each prompt differed in length by 1 to 11 words and kept the same generic structure, so these configurations are characteristic of each genre rather than artifacts of a single generation.

5. Discussion

5.1 Metafunctional Patterns and Algorithmic Hedging

The ideational and textual findings are consistent with previous research: the dense nominalization confirms the heavy nominal style reported by Herbold et al.,⁷ and the reliance on explicit transitions corresponds to Jiang and Hyland’s finding that ChatGPT relies on transitions for coherence.²⁵ The interpersonal findings, however, are the most revealing. Read against Hyland’s account of hedging (Section 2.2),^{19,20} ChatGPT’s modality differs in kind as well as degree. Expert writers hedge according to a claim’s contestability;¹⁹ ChatGPT restricts itself to median and low values, uses no high-value modality at all and never pronounces, so that caution becomes a default rather than a response to the stakes of a particular claim. Similarly, concessive resources are frequent, but they are not paired with engagement markers (reader pronouns, directives, questions) that would position the reader as a participant in the argument,²⁰ a finding consistent with Jiang and Hyland’s results for ChatGPT essays.^{25,26} ChatGPT’s prose is dialogically aware and hedged, but it lacks the rhetorical calculation that ties interpersonal resources to specific claims and readers.

5.2 Genre as a Shaping Factor

The cross-genre results confirm the hypothesis of register theory that ChatGPT’s linguistic profile varies with the genre it is asked to produce. Each part-genre regenerate its conventional staging: the introductions realize the three CARS moves,³⁰ the methodology sections progress by derivation from the study itself, and the discussions focuses modality and hedged

interpretation. This genre sensitivity also qualifies detection-oriented research based on a single prompt,²³ since the frequency of a feature such as modality ranges from none in the methodology sections to 3.95 per 100 words in the discussions. At the same time, the genre most associated with argument, the argumentative essay, is the only one in which concessions remain unresolved. ChatGPT therefore adapts the form of each genre while leaving its argumentative core undeveloped, which links the genre findings to the interpretation of “safe neutrality”.

5.3 Form Without Argument: “Safe Neutrality” Across the Metafunctions

The results show that “safe neutrality” and the lack of argumentation arise not from any single feature but from the way the three metafunctions reinforce one another. Ideationally, material processes with nominalized or non-human Actors and verbal processes with non-human Sayers remove the writer as the source of claims. Interpersonally, modality stays at median and low values, disclaimers reject only extreme positions and pronouncement is absent. Textually, derived progression lists sub-topics instead of accumulating them into a developing position, and in the argumentative essays the thesis is restated unchanged after an unresolved dialectic. Each choice is register-appropriate; their combination produces the form of academic argument without its substance. Because expert writers make the same individual choices, such text is hard to identify impressionistically. What distinguishes it is the uniform distribution of these choices, which an SFL analysis makes visible and teachers can help students to notice.

5.4 Implications for Pedagogy and Assessment

If ChatGPT can produce the surface forms of academic discourse with this competence, the pedagogical task shifts from teaching students to produce correct academic text to teaching them how such text functions as a system of meaning-making, a task for which SFL provides the metalanguage. Analyzing AI-generated text metafunctionally, students can learn to ask not only *Is this text correct?* but also *How does it position itself in relation to knowledge? What relationship does it construct with its readers? How does it guide attention?* The findings offer concrete entry points: the transitivity profile shows how agency is shifted to tools and sources; the modality profile shows that hedging is a stance toward knowledge whose value depends on where it is placed; the concession analysis shows that concession becomes argument only when it is followed by a counter; and the progression analysis shows the difference between listing and arguing.

This orientation is well established in the Sydney School genre pedagogy, which holds that students who are made explicitly aware of how texts are patterned to serve social ends can take conscious control of genre, register and meaning.³¹ Its cycle of deconstruction, joint construction and independent construction adapts readily to AI-generated texts, which can be deconstructed, analyzed jointly and used as a baseline for independent texts that deliberately depart from them. Because genre literacy should also enable students to intervene in genres,³¹ the “safe neutrality” of AI text becomes a benchmark against which students learn how a text could be argued differently, making AI literacy an integral part of SFL-informed writing pedagogy.⁸ Assessment, in turn, should value what the AI does not do: take intellectual risks, commit to positions, engage with disciplinary debates and accept the risk of being wrong.

5.5 The Human Element

The value of human academic writing lies less in formal correctness than in intellectual engagement. ChatGPT can simulate the moves of academic argumentation, such as the CARS structure, the balanced dialectic and the hedged interpretation, but it does not participate in the uncertainty, risk and disciplinary commitment that animate scholarship. This is not to dismiss AI-generated text: its mastery of register conventions is useful for brainstorming, drafting,

feedback and modeling genre structure. The educator's task is to help students distinguish writing that sounds academic from writing that advances knowledge, and to read critically, understanding not only what a text says but how it means, and whether that meaning is worth making.

6. Conclusion

6.1 Summary of Findings

The study implies SFL to analyze the way in which ChatGPT generates meaning in ten responses to academic writing prompts in five genres.. It reveals a system that has mastered academic register across the three metafunctions yet simulates rather than enacts scholarly writing. In answer to RQ1, the ideational analysis shows that material process dominates (68.3% of 208 clauses), verbal processes with non-human Sayers attribute claims to sources, relational processes are comparatively rare (9.1%), and nominalization is dense (18.0 per 100 words). Interpersonally, all T-units are declarative, modality is confined to median and low values (34 tokens; no high-value modals), evaluation relies on generic appreciation, and engagement combines entertaining and disclaiming extremes with a complete absence of pronouncement.

Textually, ChatGPT favors unmarked Themes (91.6%), uses marked Themes almost only to stage concession, and organizes information mainly through derived progression (51.8%), which lists sub-topics rather than building an argument. In answer to RQ2, these patterns vary systematically by genre: the discussions are the most modalized (3.95 per 100 words) and nominalized (24.4), the argumentative essays the most material and the only genre with unresolved concessions, the literature reviews the most verbal and existential, and the methodology sections the most derived and least modalized. In answer to RQ3, these choices together construct "safe neutrality" and most clearly in the argumentative essays, lack of argumentation. SFL therefore gives teachers and students an explicit metalanguage for deconstructing AI-generated text, and assessment should reward what the AI does not do: take a position, argue for it and accept the risk of being wrong.

6.2 Limitations and Future Research

The corpus is small (ten texts, 1,934 words), so the frequencies indicate tendencies rather than generalizable patterns, and other model versions may behave differently. All prompts concerned AI in education; this aids comparison across genres but means that ChatGPT was, in effect, writing about itself, which may have encouraged its cautious stance. No matched human corpus was analyzed, so comparisons with human writing rely on previous research.^{18,19,26} Finally, the coding was carried out by a single analyst, and although explicit criteria were applied, some categories, particularly process type and thematic progression, involve interpretive judgment. Future research should analyze larger corpora from several models and disciplines, compare them directly with matched human texts, test inter-coder reliability, and examine in classrooms whether metafunctional deconstruction of AI-generated text improves students' argumentative writing.

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest Statement: The authors declare that they have no conflict of interest.

References:

1. Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring artificial intelligence in academic essay: Higher education

- student's perspective. *International Journal of Educational Research Open*, 5, Article 100296.
2. Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students' AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), 1–13.
 3. Gulumbe, B. H. (2024). Obvious artificial intelligence-generated anomalies in published journal articles: A call for enhanced editorial diligence. *Learned Publishing*, 37(4), Article e1626.
 4. Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
 5. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). London: Routledge.
 6. Banks, D. (2024). Systemic functional linguistics: Advances and applications. *Journal of World Languages*, 10(1), 1–8.
 7. Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13, Article 18617.
 8. Troyan, F. J., Herazo, J. D., & Ryshina-Pankova, M. (2022). SFL pedagogies in language education: Special issue introduction. *System*, 104, Article 102694.
 9. Alyousef, H. S. (2016). A multimodal discourse analysis of the textual and logical relations in marketing texts written by international undergraduate students. *Functional Linguistics*, 3(1), Article 3.
 10. Istiqomah, S. N., & Maman, A. (2025). Research trends in systemic functional linguistics studies of language textbooks (2020–2025). *Aphorisme: Journal of Arabic Language, Literature, and Education*, 6(2), 568–582.
 11. Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
 12. Daneš, F. (1974). Functional sentence perspective and the organization of the text. In Daneš, F. (ed.), *Papers on Functional Sentence Perspective* (pp. 106–128). Prague: Academia.
 13. Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
 14. Halliday, M. A. K., & Martin, J. R. (1993). *Writing Science: Literacy and Discursive Power*. London: Falmer Press.
 15. Halliday, M. A. K. (1998). Things and relations: Regrammaticising experience as technical knowledge. In Martin, J. R., & Veel, R. (eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science* (pp. 185–235). London: Routledge.
 16. Schlepppegrell, M. J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Mahwah, NJ: Lawrence Erlbaum.
 17. Christie, F., & Derewianka, B. (2008). *School Discourse: Learning to Write Across the Years of Schooling*. London: Continuum.
 18. Hood, S. (2010). *Appraising Research: Evaluation in Academic Writing*. Basingstoke: Palgrave Macmillan.
 19. Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
 20. Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
 21. Martin, J. R., & Rose, D. (2007). *Working with Discourse: Meaning Beyond the Clause* (2nd ed.). London: Continuum.
 22. Ali, S. (2024). Locating the intersection of generative artificial intelligence and human English writing skills: A comparative study. *Arab World English Journal*, Special Issue on ChatGPT, 112–123.

23. Mizumoto, A., Yasuda, S., & Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints. *Applied Corpus Linguistics*, 4, Article 100106.
24. Jiang, F., & Hyland, K. (2025a). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics*, 46(3), 375–391.
25. Jiang, F., & Hyland, K. (2025b). Rhetorical distinctions: Comparing metadiscourse in essays by ChatGPT and students. *English for Specific Purposes*, 79, 17–29.
26. Jiang, F., & Hyland, K. (2025c). Does ChatGPT write like a student? Engagement markers in argumentative essays. *Written Communication*, 42(3), 463–492.
27. Zein, M., Rosa, R. N., Rangkuti, R., Zein, T. T., & Fithriani, R. (2026). Interpersonal meaning in AI-human conversations: An SFL analysis of casual vs. academic interactions in ChatGPT responses. *Journal of Research on English and Language Learning*, 7(2), 363–377.
28. Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9.
29. Lai, T. K. L., & Trinh, H. L. (2025). The effects of ChatGPT on academic writing proficiency among English-majored undergraduates at a university. *VNU Journal of Foreign Studies*, 41(3), 92–109.
30. Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
31. Rose, D., & Martin, J. R. (2012). *Learning to Write, Reading to Learn: Genre, Knowledge and Pedagogy in the Sydney School*. Sheffield: Equinox.