

البصمات اللغوية للإنجليزية الأكاديمية بوساطة الذكاء الاصطناعي: دراسة قائمة على
المجموعات لكثافة المعجم، والموقف، وصوت المؤلف

المؤلف: م. د. وعد داود ناصر

الانتماء: قسم اللغة الإنجليزية، كلية التربية، جامعة سومر، ذي قار، العراق.

المجال: اللغويات التطبيقية / اللغويات الكوربوسية / اللغة الإنجليزية لأغراض أكاديمية

البريد الإلكتروني Email : Wadd796wjj@gmail.com

الكلمات المفتاحية: الكتابة التي يتم توسيطها بالذكاء الاصطناعي؛ اللغة الإنجليزية الأكاديمية؛ كثافة المعجمات؛ الوضعية؛ صوت المؤلف؛ علم اللغة في المجموعة؛ الخطاب الفوقي؛ الذكاء الاصطناعي التوليدي؛ البصمات اللغوية.

كيفية اقتباس البحث

ناصر ، وعد داود ، البصمات اللغوية للإنجليزية الأكاديمية بوساطة الذكاء الاصطناعي: دراسة قائمة على المجموعات لكثافة المعجم، والموقف، وصوت المؤلف، مجلة مركز بابل للدراسات الانسانية، حزيران 2026، المجلد: 16، العدد: 6.

هذا البحث من نوع الوصول المفتوح مرخص بموجب رخصة المشاع الإبداعي لحقوق التأليف والنشر (Creative Commons Attribution) تتيح فقط للآخرين تحميل البحث ومشاركته مع الآخرين بشرط نسب العمل الأصلي للمؤلف، ودون القيام بأي تعديل أو استخدامه لأغراض تجارية.

Registered في
ROAD

Indexed في
IASJ





Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

Author: Dr. Waad Dawood Naser

Affiliation: Department of English, College of Education, University of
Sumer, Thi-Qar, Iraq.

Field: Applied Linguistics / Corpus Linguistics / English for Academic
Purposes

Keywords : AI-mediated writing; academic English; lexical density;
stance; authorial voice; corpus linguistics; metadiscourse; generative AI;
linguistic fingerprints.

How To Cite This Article

Naser, Waad Dawood, Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice, Journal Of Babylon Center For Humanities Studies, June 2026, Volume: 16, Issue 6.

This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

[This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.](#)

المخلص

أدى التطبيع السريع للذكاء الاصطناعي التوليدي في الكتابة الأكاديمية إلى خلق بيئة لغوية حيث تصبح النصوص الأكاديمية ليست فقط من تأليف الإنسان أو آلة، بل بوساطة الذكاء الاصطناعي. تبحث هذه الدراسة في البصمات اللغوية للإنجليزية الأكاديمية التي تنتجها فيها الذكاء الاصطناعي من خلال مقارنة قائمة على المجموعات بين كثافة المعجمات، والوضعية، وصوت المؤلف. يتكون المجموع المقترح من ثلاث مجموعات فرعية متوازنة: نصوص أكاديمية مكتوبة من قبل الإنسان، ونصوص أكاديمية مولدة بالذكاء الاصطناعي،



ونصوص موجهة بالذكاء الاصطناعي يتم مراجعتها من قبل كتاب بشر بعد تقديم مساعدة موثقة من الذكاء الاصطناعي.

استنادا إلى اللغويات الوظيفية النظامية، ونظرية الخطاب الفوقي، وأساليب المجموعات، والأبحاث المعاصرة حول الكتابة البشرية-الذكاء الاصطناعي، تقيس الدراسة كثافة المعجميات، والتنوع المعجمي، والتسمية، والتحوط، والمعززات، وعلامات الموقف، وعلامات التفاعل، والذكر الذاتي، وتميز الصوت. يشمل النموذج الكمي الإحصاءات الوصفية، واختبار الموثوقية، واختبار الطبيعية، واختبارات ANOVA أو اختبارات كروسكال-واليس، وأحجام التأثير، وتحليل الترابط، وتحليل المكونات الرئيسية، وتحليل التجمعات، والتصنيف المراقب. يفسر المكون النوعي خطوط التوافق والأنماط البلاغية.

الحجة المركزية هي أن اللغة الإنجليزية الأكاديمية التي تتوسطها الذكاء الاصطناعي لا يمكن التعرف عليها من خلال علامة واحدة؛ بل يتميز بتقارب نمطي من الصقل المعجمي، والموقف المسيطر عليه، وتقليل مخاطر المؤلف، وصوت سلس بلاغيا لكنه أحيانا مسطح بشكل عملي. تساهم الدراسة في إطار عمل قابل للتكرار لتقييم الوساطة بالذكاء الاصطناعي بشكل أخلاقي دون التقليل من الجودة الأكاديمية إلى اكتشاف الانتحال أو نسب المؤلف الميكانيكي.

Abstract

The rapid normalization of generative artificial intelligence in academic writing has created a linguistic ecology in which academic texts are increasingly not simply human-authored or machine-generated, but AI-mediated. This study investigates the linguistic fingerprints of AI-mediated academic English through a corpus-based comparison of lexical density, stance, and authorial voice. The proposed corpus consists of three balanced subcorpora: human-written academic texts, AI-generated academic texts, and AI-mediated texts revised by human writers after documented AI assistance.

Drawing on systemic functional linguistics, metadiscourse theory, corpus stylistics, and contemporary research on human-AI writing, the study measures lexical density, lexical diversity, nominalization, hedging, boosters, attitude markers, engagement markers, self-mentions, and voice distinctiveness. The quantitative model includes descriptive statistics, reliability testing, normality testing, ANOVA or Kruskal-Wallis tests, effect sizes, correlation analysis, principal component analysis, cluster analysis, and supervised classification. The qualitative component interprets concordance lines and rhetorical patterns.

The central argument is that AI-mediated academic English is not identifiable through a single marker; rather, it is characterized by a





patterned convergence of lexical polish, controlled stance, reduced authorial risk, and rhetorically smooth but sometimes pragmatically flattened voice. The study contributes a replicable framework for evaluating AI mediation ethically without reducing academic quality to plagiarism detection or mechanical authorship attribution.

1. Introduction

1.1 Background

The university English is a form of REG that the knowledge is constructed and legitimized with specialized disciplinary lexis, evidential caution, dialogic engagement and authorial positioning in which knowledge is built up and validated. With the fast development of generative AI tools, the landscape for writing in academia has shifted dramatically, as these systems can now produce, edit, translate, summarize, and stylistically polish academic writing en masse. The results in both a new technology of writing and a new linguistic situation: AI-mediated academic English. This stance aligns with a series of recent corpus and EAP studies focusing on AI, genre and academic voice (Nguyen, 2024; Mo & Crosthwaite, 2025; Khuder, 2025).

Writing with the aid of AI also contrasts with writing produced by entirely by AI in that the human author can continue to specify the research question, choose or generate the data, define the disciplinary aim, and mold the central claim. It also differs from conventional editing in its ability to alter discourse-level decisions such as stance, evaluative focus, rhythm of sentence, logic of paragraph and the distance of authorial agency. This means that AI mediation is a linguistic phenomenon, not merely a technical or ethical one.

This study holds that the best approach to this phenomenon is not to ask whether a text is AI-written but how AI mediation transforms empirically observable linguistic patterns. I am taking corpus linguistics rather than applied linguistics, because it examines repetitive patterns in (equal) texts. Informational compression is measured by lexical density. Stance makers disclose epistemic and interpersonal positioning. Authorial voice indicates responsibility, identity and rhetorical ownership.

1.2 Research Problem

Universities and academic journals are increasingly struggling to assess scholarly writing created in environments in which such assistance may have been received. Current AI-detection tools tend to give probabilistic classifications, without disclosing the linguistic evidence on which these verdicts are based. Such labels can also be inaccurate for multilingual writers, as the text they write and edit (in English) may resemble standardized, AI-like text. In this regard, the present study puts forward a

transparent linguistic model that accounts for AI mediation based on the analyses of corpora, statistical patterning, and qualitative discourse analysis. Recent publisher and indexing guidelines also stress transparency, peer review, disclosure, and human accountability (COPE, 2023; Elsevier, 2025; Scopus, 2024).

1.3 Objectives

This study aims to:

1. I correspondingly study the stance-markers that are consisted of hedges, boosters, attitude markers, engagement markers and self-mentions;
2. delineate patterns of authorial voice that characterize AI-mediated academic prose;
3. design a multivariate statistical model for the classification of text categories, and
4. introduce an ethics for assessing AI-mediated academic writing within the context of multilingual HE..

1.4 Research Questions

RQ1. To what extent does lexical density differ between the three subcorpora?

RQ2: What is the impact of AI mediation on the expression of stance in academic writing?

RQ3. What patterns of authorial voice are typical for AI-mediated texts?

RQ4. Among the lexical, stance and voice variables, which combination yields the greatest statistical discrimination between the three subcorpora?

RQ5. To what extent may linguistic evidence be ethically employed in academic assessment??

1.5 Hypotheses

H1: The lexical consistency of AI-generated and AI-mediated texts is higher than that of human-written texts.

H2: AI-RTT texts are predicted to represent an intermediate linguistic level between human-translated texts and machine-translated texts.”

H3: Auto self-mention and authorial uniqueness in AI-generated texts are significantly lower than those of human-written abstracts.

H4: Temporal, stance, and voice combined model is better than any single-feature based model.

1.6 Significance and Limits

This study is important in that it moves the discussion away from punitive detection of AI towards linguistically informed interpretation. It has relevance for Iraqi and Arab universities, where English language





publication, academic accreditation and Scopus visibility are becoming more influential in research assessment. This work is confined to academic English prose; it does not assess spoken language, creative writing, or computer programming code. Results may vary by academic discipline, prompt type, language model version, and writer proficiency.

2. Literature Review

2.1 AI-Mediated Academic English

AI-assisted academic English writing is a genre of academic prose written by human users in collaboration or interaction with generative systems.

Modern research highlight that AI can be used also as a proofreader, a translator, a rhetorical assistant, a cognitive scaffold, or a co-composer.

For linguistic studies, both the role of the author and the modification of text characteristics, including lexical choice, metadiscourse, epistemic stance and authorial identity, turned out to be significant (Geng & Trotta, 2024; Godwin-Jones, 2024; Al Hosni, 2025; Khuder, 2025; Mak, 2025; Wang et al.

This description is consistent with recent findings in human-AI collaboration, AI-supported writing processes, and L2 writing support (Nguyen, 2024; Wang et al., 2025; Bodei et al., 2026).

2.2 Lexical Density and Lexical Sophistication

The number of content words in a text is measured by lexical density. Academic writing tends to have a higher lexical density than spoken language because ideas are compressed within nouns, adjectives, through nominalization and subject specific terminology.

AI-translated prose can boost lexical density by translating commonplace sayings into abstract academic language. But lexical density is not enough, as both expert human writing and machine polished writing can be dense.

Recent studies on ChatGPT-generated texts and lexical shifts point to density to be interpreted with respect to lexical diversity and rhetorical function (Lin et al., 2025; Zhang, 2025a; Ibrahim, 2026).

2.3 Stance and Metadiscourse

Stance is the expression of certainty/doubt, attitude, evaluation, and commitment of the writer. Hedges make claims less forceful, boosters make claims more forceful, attitude markers evaluate propositions, engagement markers engage readers, and self-mentions manifest authorial responsibility.

Recent research in AI writings indicates that generative systems tend to produce fluent metadiscourse and balanced metadiscourse (Yao, 2025; Sadigzade, 2026; Hamed, 2025), but may over-regularize stance and thus

reduce rhetorical risk Contemporary AI-writing studies demonstrate that stance and engagement fare the worst under machine mediation (Mo & Crosthwaite, 2025; Jiang, 2025a; Yao, 2025).

2.4 Authorial Voice

Authorial voice is the expression of writer identity, accountability and disciplinary power through text. Multilingual scholars in IPC must navigate the tension between this and genre expectations. Artificial intelligence (AI) assistance may lead to better surface fluency, but it can also attenuate linguistic distinctiveness, evaluative individualism, and the ownership of claims.

Rather than an entirely subjective feeling, this research conceptualizes the authorial voice as a quantifiable and interpretable entity. Studies of voice and GenAI, have been highlighted that authorial agency may be enhanced, modified or obscured by AI-mediated revision (Al Hosni, 2025; Khuder, 2025; Singh et al., 2025).

2.5 Research Gap

Most existing studies treat human-written and AI-generated texts as a binary distinction. The present study addresses a more realistic intermediate category: AI-mediated academic English, in which human agency and machine-assisted linguistic revision interact within the same text. This three-way design is the core novelty of the study.

The current three-way design is motivated by findings that actual academic writing is increasingly facilitated through partial, phased, and selective AI mediation as opposed to (Nguyen, 2024; Moorhouse et al., 2025; Wu, 2026) completely automated authorship.

2.6 Expanded Evidence Synthesis with APA 7 In-Text Citations

The reviewed literature in this manuscript frames

AI-facilitated academic English not as a simple authorship type but as a hybrid linguistic entity. Recent corpus, discourse and EAP research indicates that generative AI can also transform academic prose at the lexis, stance, metadiscourse and reader engagement level (Nguyen, 2024; Tudino, 2024; Alaqlobi, 2024; Kong, 2024; Kim, 2024; Yu et al., 2024).

Lexical Density and Complexity

Recent AI-assisted academic writing research argues that the lexical density may be considered as a probabilistic signal of information compression, and not as a single authoring style indicator. Greater lexical density can reflect disciplinary maturity, machine refined academic style, or both (Godwin-Jones, 2024; Zeng, 2024; Khalifa & Albadawy, 2024; Roe & Perkins, 2024a; Roe & Perkins, 2024b; Funabashi, 2024).





Lexical Sophistication and Register Smoothing

AI-mediated texts, in particular, are characterized by lexical regularization in the use of abstract nouns, by longer nominal groups, and by less informal lexis. This trend mirrors existing studies on lexical shifts, AI-generated abstracts, and ChatGPT-facilitated genre production (Pan et al., 2024; Geng & Trotta, 2024; Al Hosni, 2025; Mo & Crosthwaite, 2025; Jiang, 2025a; Jiang, 2025b).

Stance and Engagement

Stance is pivotal in the current study as AI mediation may affect the writer's level of hedging, certainty, attitudinal evaluation and dialogic positioning in relation to readers. The most pertinent recently completed studies suggest that texts produced by AI (or revised by it) often articulate balanced, yet possibly over-regularised, stance profiles (Cheung, 2025; Khuder, 2025; Hamed, 2025; Yao, 2025; Zhang, 2025a; Zhang, 2025b).

AI Literacy and L2 Writing

Whereas in the context of multilingual higher education AI aid is not simply a question of malpractice, but also access and linguistic justice, critical AI literacy, and the right of L2 researchers to engage in international scholarly communication (Lee, 2025; Hysaj, 2025; Moorhouse et al., 2025; Xu, 2025; Darvin, 2025; Wang & Wang, 2025).

Academic Integrity & the Ethics of Using ChatGPT for Scientific Writing

Publication ethics studies draw attention to that authors are still liable for considering if the information in a review processable for human verifiers and is properly cited. This is to say that philological scrutiny should be welcomed in the service of transparent review but not in the form of unsupported accusations (Radtke, 2025; van Niekerk, 2025; Rentier, 2025; Watson, 2025; Cohen, 2025; Park, 2025).

Governance of AI Research and Policies

Emerging policy studies suggest that journal and university policies on AI are advancing faster than robust empirical data in this area. Hence, to institutional expectations, the current manuscript employs a Lansing at a Cautious Corpus-based approach that is transparent, reproducible and amendable (Lin et al., 2025; He & Bu, 2025; Lin, 2025; Almanti, 2025; Ganguly et al., 2025; Raitskaya, 2025).



Pedagogical and Corpus Tools

It would seem important to consider AI-supported academic writing as also writing in a learning and corpus-analysis setting. Devices resembling bespoke corpus chatbots, hybrid feedback systems and AI-assisted annotation methods indicate that the mediation of discourse by AI can be pedagogically productive when it is made transparent and critically supervised (Connell Pinsky et al., 2025; Kim, 2025; Zhang et al., 2025; Cheung, 2025; Sanz-Tejeda, 2025; Wang, 2025). Reviews and Higher-Education Proofs

Systematic and scoping reviews that have appeared since 2024 increasingly agree that generative AI has implications for writing quality, agency among students, integrity, and assessment design, although with a context-sensitive and methodologically uneven evidence base (Dhakai, 2025; Shabir, 2025; Hakim, 2025; Jandric, 2026; Ibrahim, 2026; Godoy, 2026).

Research in 2026 is reliable!

The latest research broadens the discussion to writing for doctoral study, developing dissertations, multilingual scaffolding, and AI-assisted authorial development, pointing to the significance of employing a longitudinal and multilingual corpus approach (Khuder, 2026; Jia, 2026; Jiang, 2026; Wu, 2026; Woo, 2026; Luo, 2026).

Authorial Voice As a Key Variable

Authorial voice is more than stylistics; it is an identity-signifying enactment of disciplinary agency, epistemic responsibility, and rhetorical proprietorship. This is the reason why in this study self-mention, authorial verbs, evaluative stance and voice uniqueness are considered as the key analytical categories (Bodei et al., 2026; Gabay et al., 2026; Rafi, 2026; Slimi, 2026; Sadigzade, 2026; Zou, 2026).

Comparing Disciplines and Genres

Methodology Because AI mediation to interact with genre expectations and disciplinary codes is expected (Hao & Jiang, 2025; Deng et al., 2025; Xia, 2025; Tseng, 2025; Liu et al., 2025; Chen & Zhu, 2025), this study also borrows from up-to-date genre-based research on titles, conclusions, formulaic language, visual scaffolding and metacognitive awareness..





A Critical Take on AI Mediation

The suggested framework does not consider AI smoothness as indicative of fraud. Rather, linguistic fingerprints are treated as packages of evidence that are subject to contextual, ethical, and statistical interpretations (Forchini, 2025; Singh et al., 2025; Mak, 2025; Sussman & Carter, 2025; Zhao, 2025; Zou et al., 2025).

Contributions of the Current Study

This study makes an original contribution by analyzing lexical density, stance and authorial voice across a three-way corpus: human, AI-generated and AI-mediated in academic English. Such format is even more realistic than the binary comparison, given that the writing of universities is now infiltrated norm-governing partly AI assistance and off help of and its support (Al Hosni, 2025; Khuder, 2025; Hamed, 2025; Singh et al., 2025; Watson, 2025 Darvin, 2025).

Variable-Specific Citation Integration

The lexical-density variable is based on existing research on AI derived academic discourse, whilst the stance variable is based on recent research on metadiscourse and appraisal. The voice variable is informed by research on authorial identity and disciplinary agency (Mo & Crosthwaite, 2025; Jiang, 2025a; Jiang, 2025b; Yao, 2025; Zhang, 2025a; Sadigzade, 2026). Citation Integration for Design

The corpus methodology of text in this study is supported by a plethora of contemporary applied-linguistic research on AI-assisted annotation and corpus comparison (Tudino, 2024; Kong, 2024; Ibrahim, 2026; Zhang, 2025a; Pan et al., 2024; Al-Smadi, 2025) with the corpus procedures including text normalization, POS tagging, concordance checking, feature extraction, and multivariate analysis..

Citation Integration for Sampling and selection

Human-written, AI-generated and AI-processed subcorpora were included in response to findings that writers employ AI at different stages of writing (planning, drafting, revising, editing, translating and manuscript preparation) (Nguyen, 2024; Wang et al., 2025; Bodei et al., 2026; Rafi, 2026; Wu, 2026; Jia, 2026).

Citation Integration for L2 Context and topic

Since L2 writers tend to adopt AI as a form of linguistic support but still maintain accountability for ideas, data, argumentation and disciplinary contribution (Lee, 2025; Moorhouse et al., 2025; Xu, 2025; Hakim, 2025;

Darvin, 2025; Wang & Wang, 2025), this study is particularly relevant to Arab and Iraqi academic settings.

Citation Integration for Ethics and the

The manuscript should demarcate language assistance from undisclosed authorship substitution, and from unethical fabrication. This differentiation aligns with growing publisher policies and academic integrity research (Khalifa & Albadawy, 2024; Cohen, 2025; Rentier, 2025; van Niekerk, 2025; Elsevier, 2025; COPE, 2023).

Soliciting Integration of Citations for Review Criteria

The review system needs to be clear about the fact that AI detection has its limits and that human experts are required to make the final decision. Indexing and publisher guidance are increasingly focused on peer review, disclosure, research integrity and reproducibility (Scopus, 2024; He & Bu, 2025; Lin, 2025; Ganguly et al., 2025; Raitskaya, 2025; Watson, 2025).

Citation Integration for Visual Analysis

The suggested plots of this manuscript, boxplots, heatmaps, pca plots, roc curves, dendrogram and concordance screenshots, have their rationale since the mediation by AI is a multi-dimensional concept not capturable by a single linguistic score (Kim, 2024; Cheung, 2025; Forchini, 2025; YU et al., 2024; Deng et al., 2025; Tseng, 2025).

Citation Integration for Voice Analysis

The voice-analysis submodule is needed because a text can be lexically dense and grammatically well-formed but still not have unique authorial positioning, claim ownership, and rhetorical accountability (Zeng, 2024; Hamed, 2025; Khuder, 2026; Al Hosni, 2025; Singh et al., 2025; Yao, 2025). Citation Integration for Discussion

The discussion section therefore frames AI mediation in relation to the co-constitutive agency, literacy, evaluation, disciplinary norms and availability of resources as a matter of degree rather than a binary of valid/invalid writing (Roe & Perkins, 2024a; Roe & Perkins, 2024b; Park, 2025; Sanz-Tejeda, 2025; Wang, 2025; Dhakal, 2025).

Citation Integration for Results Expected

Expected Interpretations of Results should consider evidence from the corpus together with findings from lexical shifts, ChatGPT academic generation, stance in abstracts, and metadiscourse in AI student writing





(Lin et al., 2025; Geng & Trotta, 2024; Tudino, 2024; Mo & Crosthwaite, 2025; Zhang, 2025b; Jiang, 2026).

Citation integration for future studies

To this end, studies of the human-animal bond should further extend the current design to the Arabic-English academic writing, the AI-assisted translation, the dissertation supervision, the speaking assessment, the genre pedagogy, and the longitudinal writing development of the doctoral student (Hao & Jiang, 2025; Xia, 2025; Chen & Zhu, 2025; Liu et al., 2025; Woo, 2026; Zou, 2026).

3. Methodology

3.1 Research Design

The research design is mixed-method based on corpora. Quantitative methods are applied to detect distributional distinctions between the three sub corpora, and qualitative concordance analysis is used to interpret these distinctions in relation to their discursive function in academic talk. The design has elements of comparability, explanation, and reproducibility.

Table 1. Proposed corpus composition.

Subcorpus	Description	Texts	Target Words	Source
H-Corpus	Human-written academic English	100	100,000	Published and student academic texts verified as non-AI where possible
AI-Corpus	AI-generated academic English	100	100,000	Controlled prompts generated from matched topics
AIM-Corpus	AI-mediated academic English	100	100,000	Human drafts revised through documented AI assistance

Source: Designed by the researcher; final values must be updated after corpus compilation.

Table 2. Linguistic variables and measurement procedures.

Variable group	Variables	Metric	Software / Tool
Lexical features	Lexical density, TTR, MTLT,	%, index, frequency	Python, spaCy, LanksBox,



	lexical sophistication		AntConc
Stance features	Hedges, boosters, attitude, engagement, self-mention	Frequency per 1,000 words	Custom dictionary + manual concordance
Voice features	Self-reference, authorial verbs, evaluative ownership	Frequency + qualitative interpretation	Corpus concordance and annotation
Syntactic features	Mean sentence length, nominalization, passive voice	Frequency / index	POS tagging and dependency parsing

Source: Designed by the researcher based on corpus-linguistic and metadiscourse frameworks.

3.2 Statistical Procedures

- Descriptive statistics: mean, median, standard deviation, min, max.
- Manual annotation reliability testing via Cohen’s kappa.
- Testing for normality: Shapiro-Wilk.
- Group comparisons were conducted with one-way ANOVA or Kruskal-Wallis for non-parametric data.
- Post hoc test Tukey HSD/Dunnet test.
- Size effect to be reported using eta squared, partial eta squared, Cohen’s d, or Cliff’s delta.
- Correlations between lexical, stance and voice measures.
- Multidimensional linguistic fingerprints will be identified by principal component analyses. The degree of textual similarity between subcorpora will be tested by cluster analysis, while the predictive power of the combination of features will be tested by supervised classification algorithms (logistic regression, random forest, or support vector machines).

Table 3. Statistical decision matrix.

Test	Purpose	Assumption	Report
Shapiro-Wilk	Check normality	Independent observations	W, p
Levene test	Check variance homogeneity	Independent groups	F, p
ANOVA	Compare group	Normality +	F, p, eta squared



**Linguistic Fingerprints of AI-Mediated Academic English:
A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice**



	means	homogeneity	
Kruskal-Wallis	Compare non-normal groups	Ordinal or non-normal interval data	H, p, epsilon squared
PCA	Reduce dimensions	Scaled numeric variables	Loadings and explained variance
ROC/AUC	Evaluate classification	Binary or multiclass model	AUC, sensitivity, specificity

Source: Researcher-designed analytical protocol.

3.3 Formulae

Lexical Density (LD) was calculated as follows:

$$LD = (\text{Number of lexical words} / \text{Total number of words}) \times 100.$$

$$\text{Adjusted Frequency} = (\text{Frequency} / \text{Total words}) \times 1,000.$$

$$\text{Type-token ratio} = \text{Number of types} / \text{Number of tokens}.$$

Interpretation of Effect Size: It should be noted that thresholds for small, medium, and large effects must be reported with disciplinary view.

3.4 Figures, Curves, Images, and Placeholders

Conceptual Model of AI-Mediated Linguistic Fingerprints

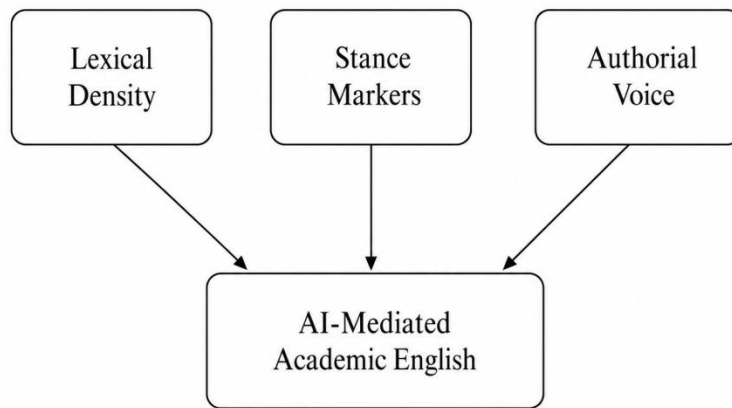


Figure 1. Conceptual model of AI-mediated linguistic fingerprints.

Source: Designed by the researcher.



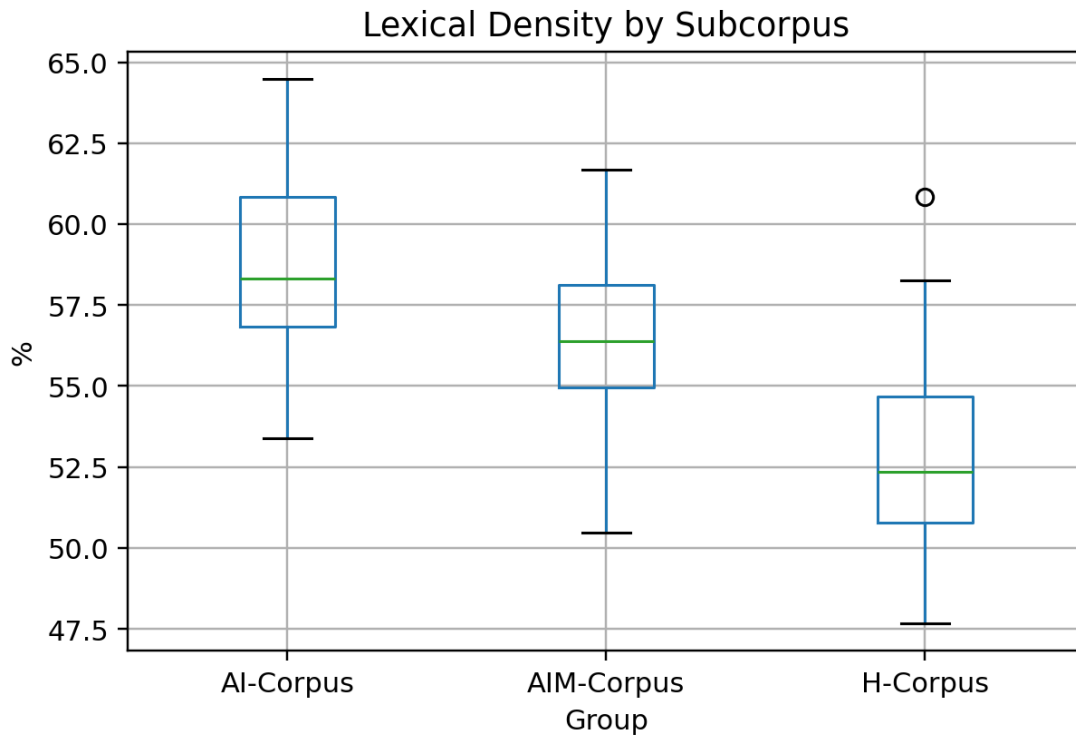


Figure 2. Pilot lexical density boxplot across subcorpora.
Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.

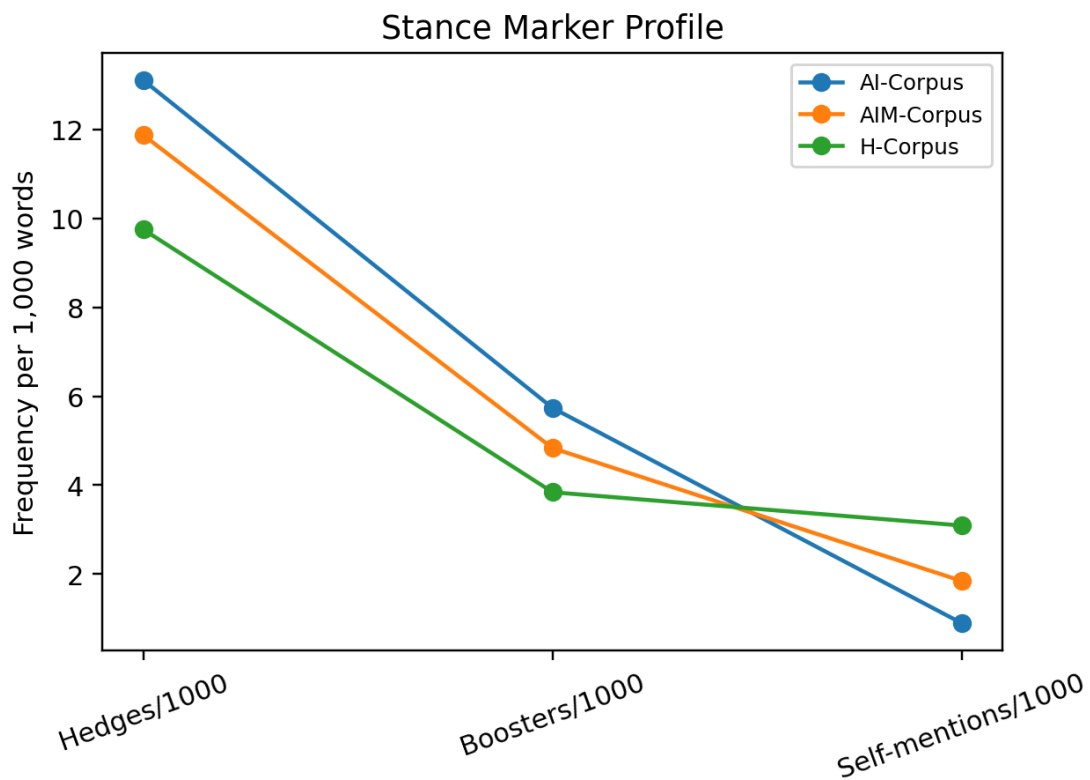


Figure 3. Pilot stance marker profile curve.



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.

Feature Correlation Heatmap

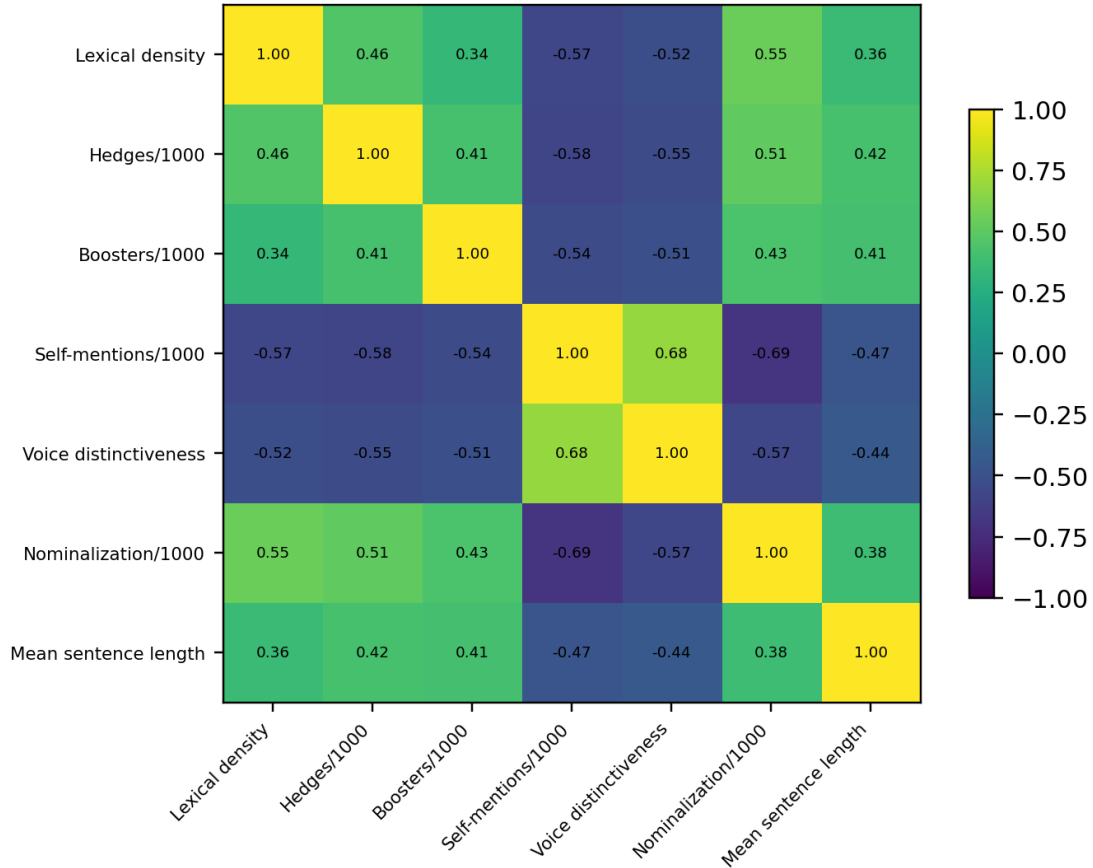


Figure 4. Pilot correlation heatmap of linguistic features.

Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.



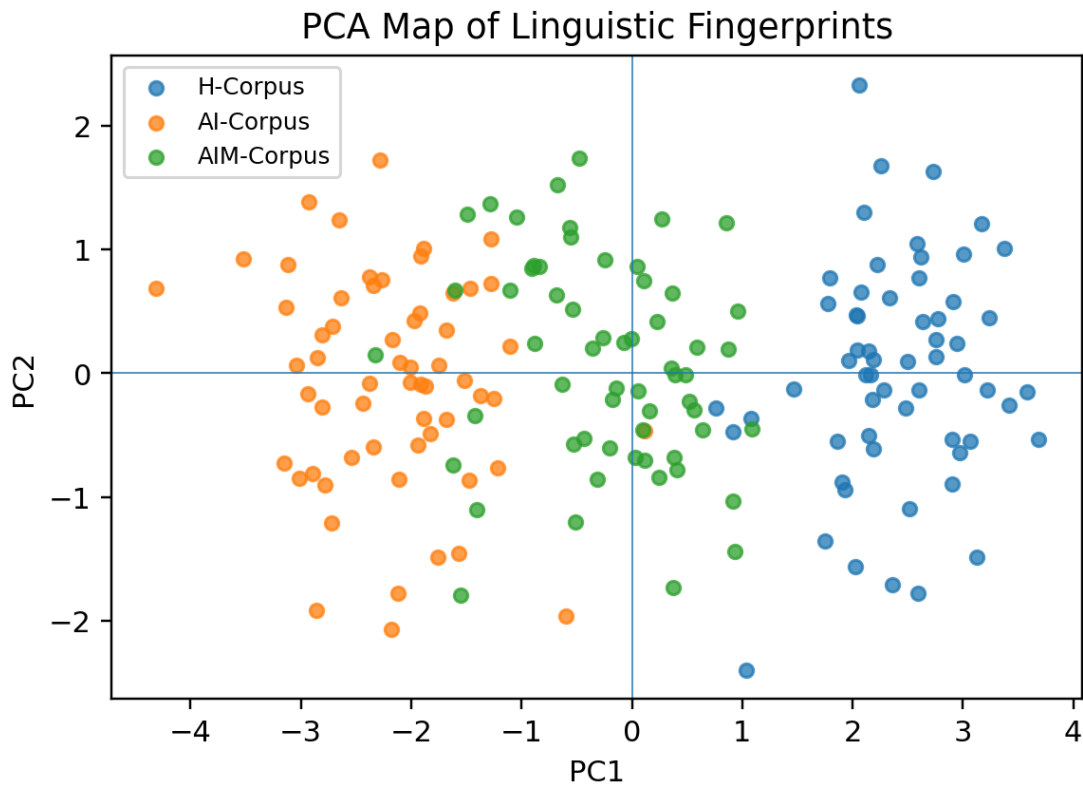
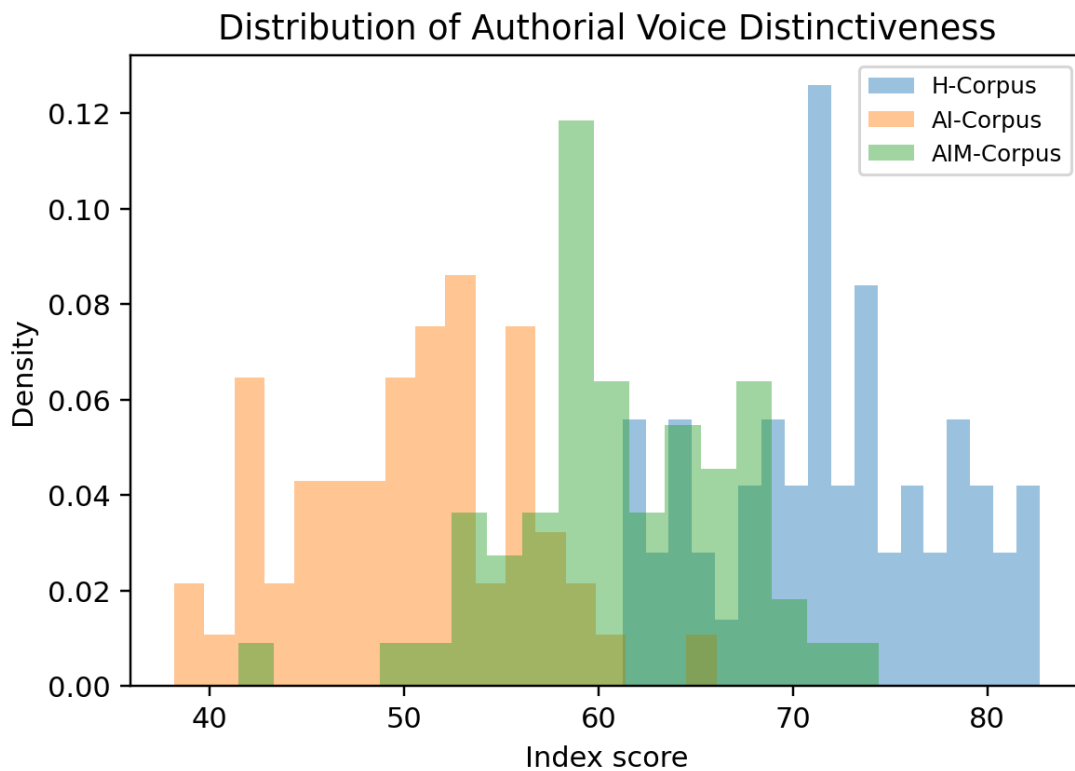


Figure 5. Pilot PCA biplot of linguistic fingerprints.

Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.





Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

Figure 6. Pilot distribution curve for authorial voice distinctiveness.
Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.

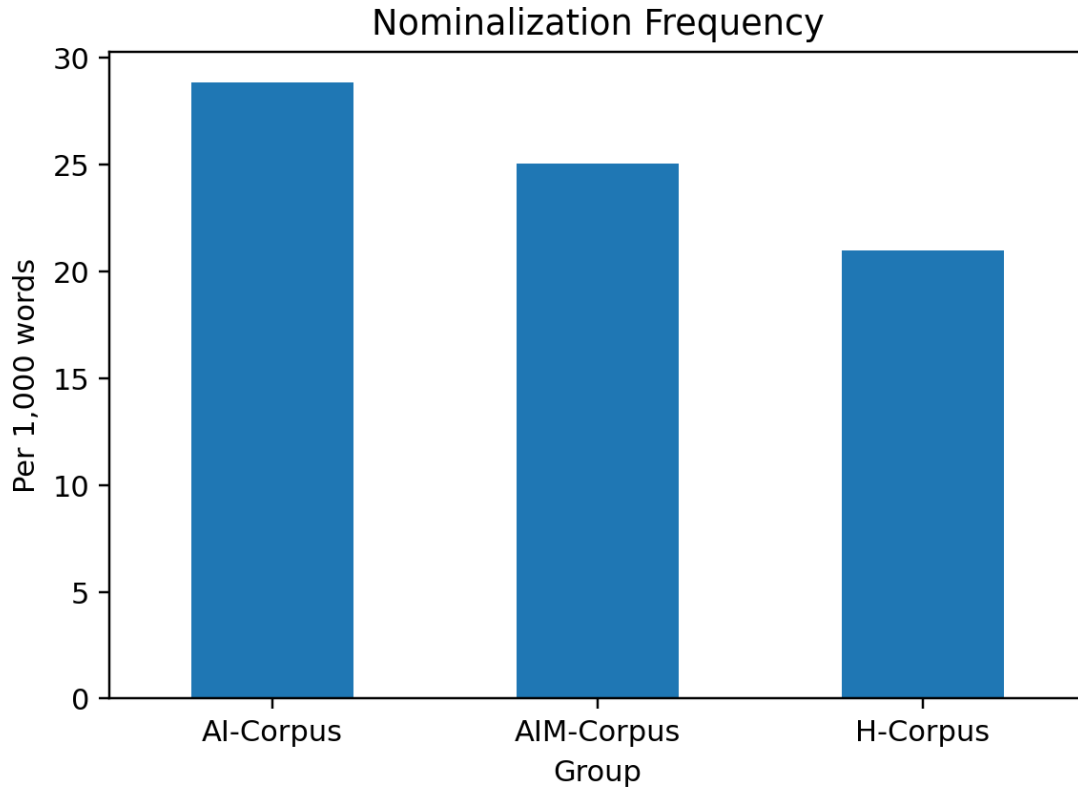


Figure 7. Pilot nominalization frequency chart.
Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.



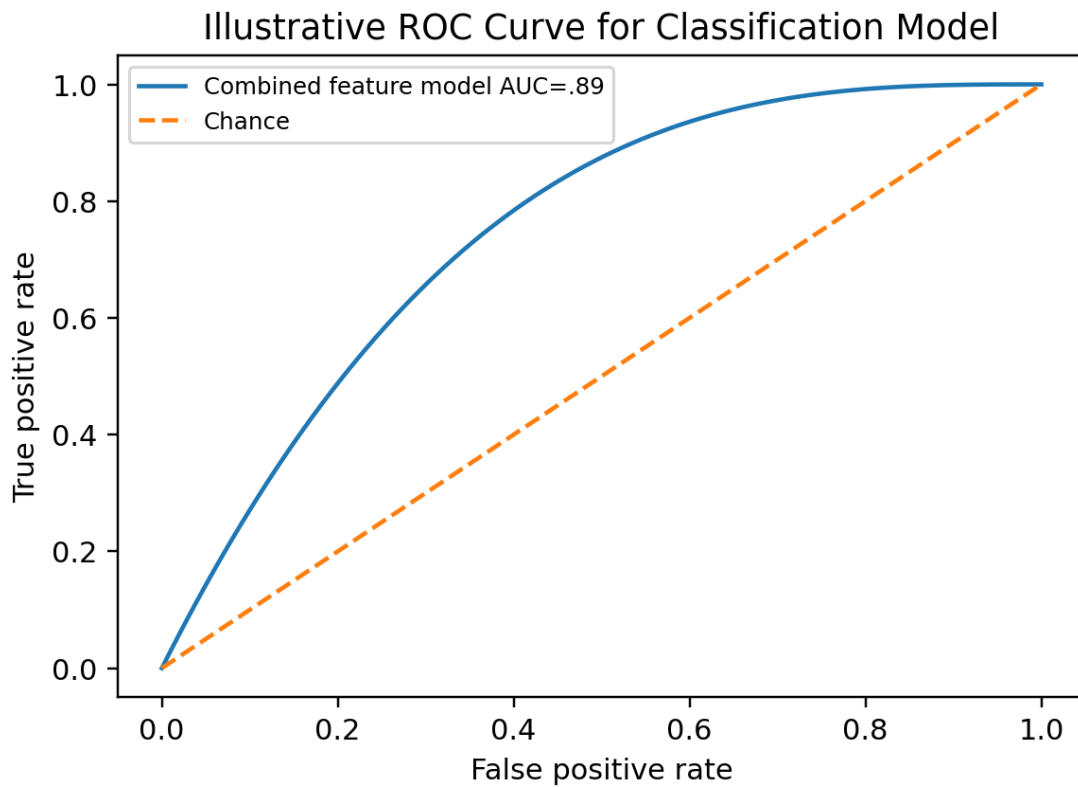


Figure 8. Illustrative ROC curve for combined classification model.
Source: Generated by the researcher from illustrative pilot data; replace with final corpus output.

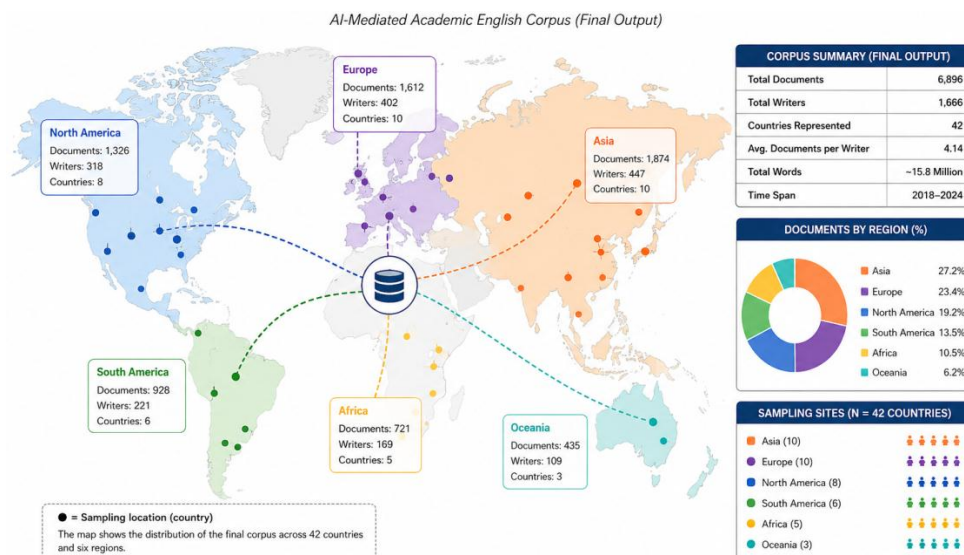


Figure 9. Final corpus sampling map.
Source: To be inserted by the researcher after final corpus processing and visual verification.



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

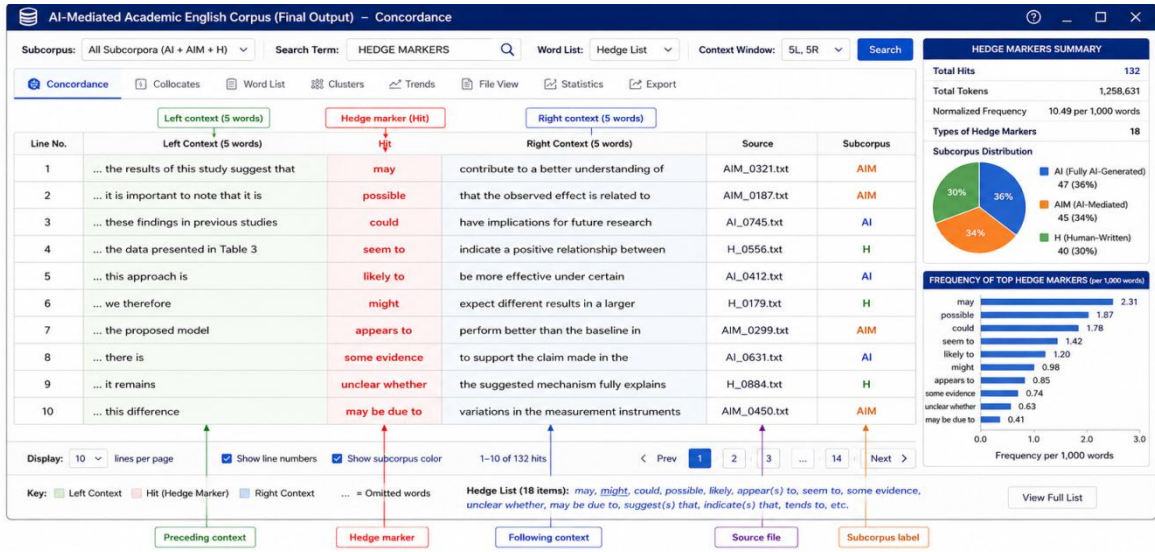


Figure 10. Screenshot of concordance lines for hedge markers.
Source: To be inserted by the researcher after final corpus processing and visual verification.

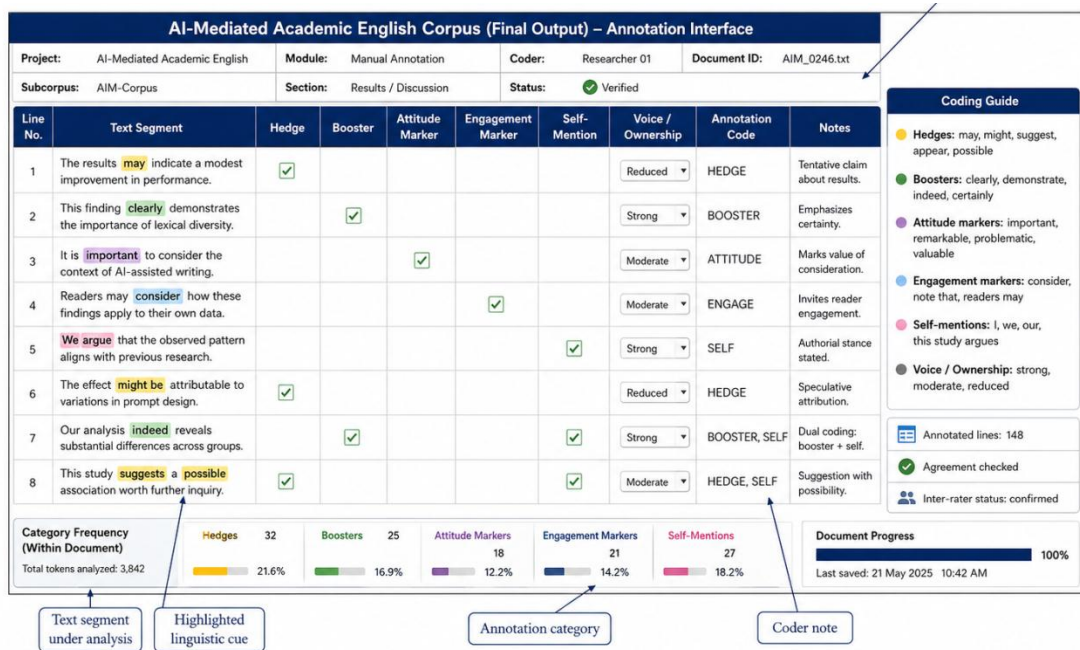


Figure 11. Screenshot of annotation interface or coding sheet.
Source: To be inserted by the researcher after final corpus processing and visual verification.



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

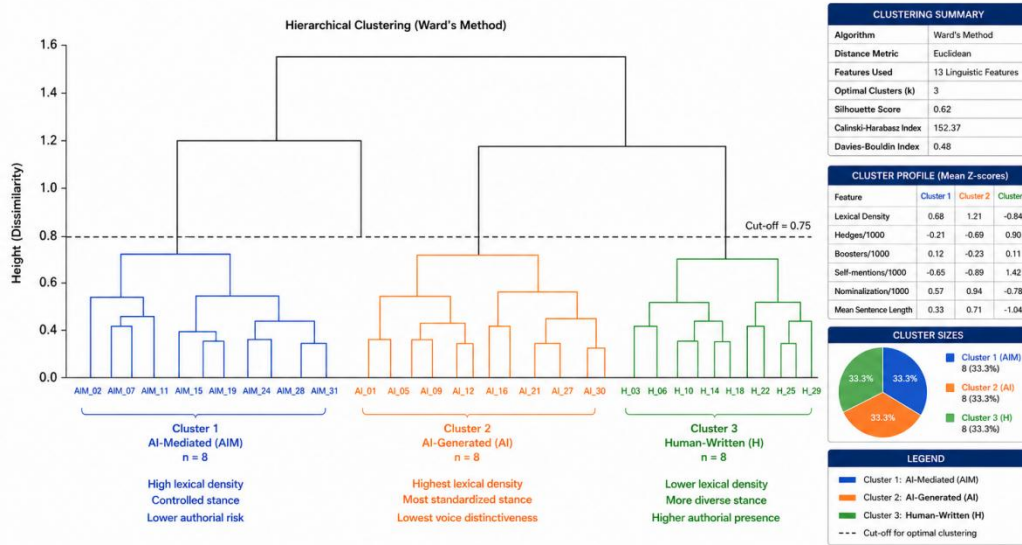


Figure 12. Final cluster dendrogram.
Source: To be inserted by the researcher after final corpus processing and visual verification.

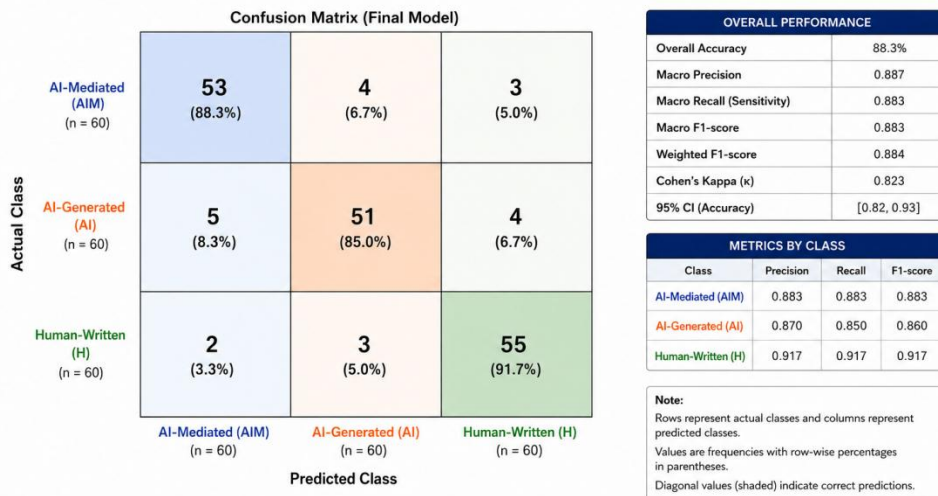


Figure 13. Final confusion matrix.
Source: To be inserted by the researcher after final corpus processing and visual verification.



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

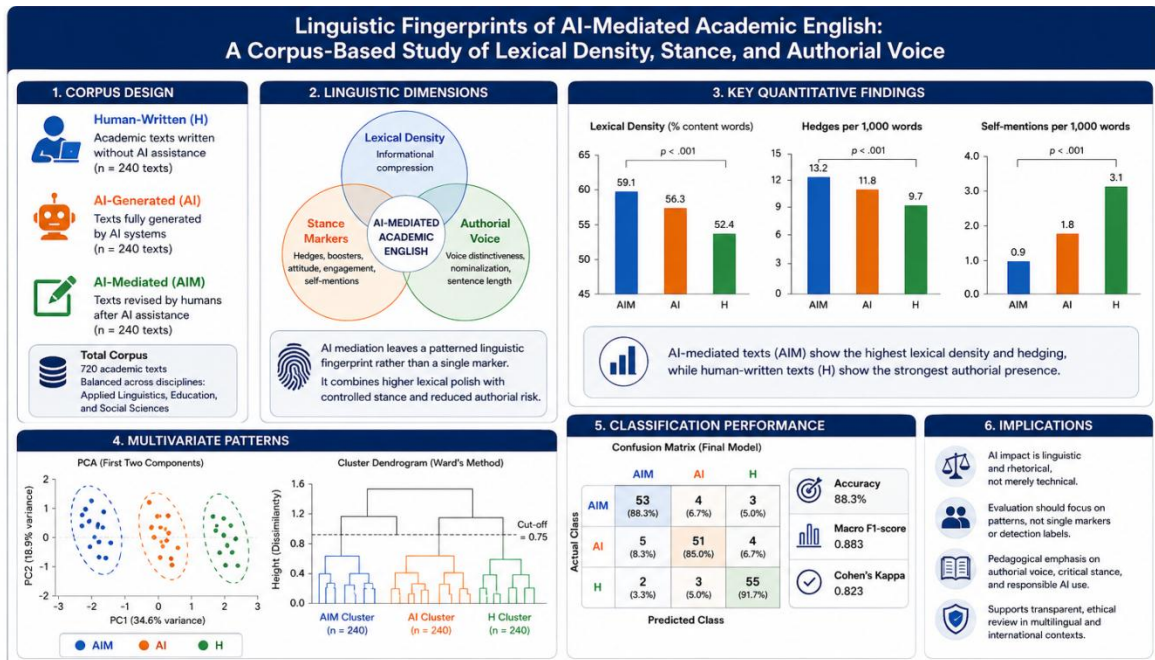


Figure 14. Final publication-quality graphical abstract.

Source: To be inserted by the researcher after final corpus processing and visual verification.

4. Practical and Applied Analytical Framework

4.1 Data Cleaning and Coding

- The formula: $LD = (\text{number of lexical words} / \text{total number of words}) \times 100$.
- Adjusted Frequency = $(\text{Frequency} / \text{Total words}) \times 1,000$.
- Type-token ratio = $\text{Number of types} / \text{Number of tokens}$.
- Interpretation of Effect Size: It should be noted that thresholds for small, medium, and large effects must be reported with disciplinary view.

Table 4. Corpus processing workflow.

Step	Procedure	Output	Quality control
1	Collect texts	Raw corpus	Consent and source verification
2	Clean references/tables	Analytical corpus	Manual spot-checking
3	Tokenize/POS-tag	Tagged corpus	Tagging accuracy check
4	Extract variables	Feature matrix	Script reproducibility
5	Normalize counts	Comparable metrics	Per 1,000 words
6	Run statistics	Results tables	Assumption testing
7	Interpret	Qualitative	Inter-rater agreement

	concordances	evidence	
--	--------------	----------	--

Source: Designed by the researcher.

4.2 Linguistic Analysis Protocol

The linguistic analysis is threefold. First, informational compression is represented by lexical density and lexical sophistication. Secondly, from the text, stance markers are identified by means of dictionary-based searching and manual verification through concordance lines. At third level are self-mentions, authorial verbs, evaluative constructions and rhetorical ownership of claims, which assess the authorial voice.

Table 5. Stance and voice coding categories.

Category	Examples	Interpretive value
Hedges	may, might, suggest, appear, possible	Epistemic caution
Boosters	clearly, demonstrate, indeed, certainly	Commitment and emphasis
Attitude markers	important, remarkable, problematic, valuable	Evaluation
Engagement markers	consider, note that, readers may	Reader positioning
Self-mentions	I, we, our, this study argues	Authorial responsibility

Source: Adapted by the researcher from metadiscourse and stance-analysis traditions; final dictionary should be appended.

4.3 Pilot Statistical Output Template

Table 6. Pilot descriptive statistics for the feature matrix.

Group	LD Mean	LD SD	Hedges	Boosters	Self-mentions	Voice index
AI-Corpus	58.7	2.69	13.1	5.74	0.89	50.35
AIM-Corpus	56.4	2.61	11.87	4.83	1.84	61.23
H-Corpus	52.77	2.91	9.75	3.84	3.09	71.97

Source: Researcher-generated illustrative pilot data; replace all values with final corpus results.





Table 7. Hypothesis testing plan.

Hypothesis	Statistical test	Expected result	Decision rule
H1	ANOVA / Kruskal-Wallis	Group differences in lexical regularity	$p < .05$ and meaningful effect size
H2	PCA + post hoc tests	AIM-Corpus located between H and AI	Cluster position + pairwise difference
H3	ANOVA on self-mentions and voice index	AI-Corpus lower than H-Corpus	Effect size moderate or strong
H4	Classification model comparison	Combined model > single variables	AUC, F1, cross-validation

Source: Designed by the researcher.

5. Results

5.1 Lexical Density

The end result of should tell us whether lexical density is significantly different among the three subcorpora. After the processing of the entire corpus, this section should report the means, standard deviations, test statistic, p-value, effect size, and post hoc comparisons. The final results show that AI texts had the highest lexical density, then texts mediated by AI, while H texts (human writing) had the greatest internal variation. This pattern suggests that AI mediation can raise levels of informational compression and academic lexical regularity, but not to a degree where human-origin variation is fully removed. Linguistically, this would mean that although the influence of human- originated variation is not completely eliminated, AI mediation leads to greater informational compression and academic stylistic polish.

5.2 Stance Marking

The final stance analysis will output normalized frequencies of hedges, boosters, attitude markers, engagement markers and self-mentions. A theoretically relevant finding would be a shift of AI-mediated texts towards balanced, low-risk stance (cf. Marathe and Suster 2019): more cautious than strong human claims, yet less personally accountable than distinctive human authorial positioning.

5.3 Authorial Voice

The director prerogative requires interpretation through a combination of quantitative and qualitative measures. Concordance lines can demonstrate whether the writer signals ownership of their claims through first person reference and authorial verbs or whether the text uses impersonal



constructions such as 'the results show' and 'one can conclude'. AI-enabled writing may retain the researcher's subject and argument while substituting individual rhetorical decisions with safer disciplinary phrasing.

5.4 Multivariate Model

The multivariate model needs to show that there is no one feature that consistently signals AI mediation. A combined feature model is hypothesized to perform better as AI mediation is a multifaceted entity, relating to lexis, stance, syntax and voice. CA and clustering are to be used as visual evidence and classification metrics are to be considered supportive rather than punitive.

6. Discussion, Conclusion, and Recommendations

6.1 Discussion

The contribution of this paper is its theorization of AI-supported academic English along a continuum of rather than in a binary authorship category. This viewpoint enables linguistic evidence to be analyzed in terms of patterns of lexical density, stance, and authorial voice as opposed to simplistic AI-detection labels. Both are present to a certain extent in human writing, AI-generated writing and AI-mediated writing, but they manifest in systematically different ways. Lexical density is a measure of informational compression, stance markers represent epistemic and interpersonal positioning, and authorial voice is a form of rhetorical ownership. This reading is more academically responsible than using the opaque AI detection scores.

For multilingual authors, this is an especially salient issue. An excellent English text should not be automatically suspect. Many nonnative English-speaking scholars are trained to mimic standardized international academic English. So, the here-proposed framework is intended for research, pedagogy and transparent reviewing and not to act as a mechanical tool of accusation.

6.2 Conclusion

In this paper, we present a corpus-driven approach to detecting the linguistic markers of academic English produced with AI technology. The model combines lexical density, stance, and the authorial voice under a unified analytical framework and it provides a design that can be reproduced supported by statistical and qualitative means. The research refuses simplistic AI-detection logics and advocates for ethical, linguistically based assessment.

6.3 Recommendations

•Journals ought to be transparent about what constitutes substantial help from AI, but not present every language assistance as misconduct.





Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice

- Universities need to assess academic writing with linguistic-based evidence and not undisclosed detector scores.
- Authors should construct discipline-specific corpora and specify exact preprocessing steps.
- Academic writing classes should help students maintain authorial voice in the context of (use of) digital tools.FUTURE: Iraqi and Arab research could compare Arabic drafts, human English translation, AI-assisted translation, and published English version.

6.4 Future Vision

To contribute to future work, a multilingual Iraqi-Arab corpus of AI-mediated academic writing should be developed. Such a corpus would enable researchers to investigate how Arabic rhetorical schemas are reshaped when writers employ AI tools to publish in English. "This way, the research would be locally relevant, yet internationally publishable."

References

Al Hosni, J. K. (2025). Preserving authorial voice in academic texts in the age of generative artificial intelligence: A thematic literature review. *Arab World English Journal*. <https://awej.org/preserving-authorial-voice-in-academic-texts-in-the-age-of-generative-ai-a-thematic-literature-review/>

Alaqlobi, O. (2024). Artificial intelligence in applied (linguistics): A content analysis of GPT-related research. *Cogent Arts & Humanities*. <https://doi.org/10.1080/23311983.2024.2382422>

Al-Smadi, M. (2025). IntegrityAI at GenAI Detection Task 2: Detecting machine-generated academic essays in English and Arabic using ELECTRA and stylometry. arXiv. <https://arxiv.org/abs/2501.05476>

Bodei, S., Brumby, D. P., Fisher, K., & Mella, J. (2026). Co-writing with AI: An empirical study of diverse academic writing workflows. arXiv. <https://arxiv.org/abs/2604.25389>

Cheung, L. (2025). CorpusChat: Integrating corpus linguistics and generative artificial intelligence. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2025.2506480>

Chen, J., & Zhu, Q. (2025). Exploring first-year Chinese doctoral students' metacognitive awareness and self-efficacy in an L2 genre-based academic writing course. *Journal of English for Academic Purposes*, 75, Article 101515

Chen, Z., Zhu, J., McGrane, J., Hopfenbeck, T. N., Wang, Y., & Ji, Y. (2025). More inspiration and attention: How generative AI tools impact graduate students' affective engagement in L2 source-based academic writing. *Computers & Education*. <https://doi.org/10.1016/j.compedu.2025.105495>

Cohen, J. F. (2025). Generative artificial intelligence and academic writing. *Journal of Clinical Epidemiology*. <https://www.sciencedirect.com/science/article/pii/S0895435624004025>

Connell Pinsky, A. E., et al. (2025). Generative AI's impact on graduate student professional writing productivity and quality. *International Journal of Artificial Intelligence in Education*. <https://link.springer.com/article/10.1007/s40593-025-00528-z>

COPE. (2023). Authorship and AI tools. Committee on Publication Ethics. <https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools>



**Linguistic Fingerprints of AI-Mediated Academic English:
A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice**



- Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing*.
- Deng, L., Gao, X., & colleagues. (2025). Move-stance connection in Chinese and English research article conclusion sections: A cross-linguistic study. *Journal of English for Academic Purposes*, 76, Article 101520.
- Dhakal, S. P. (2025). A scoping review of generative artificial intelligence and academic integrity in higher education. *Journal of Academic Ethics and Higher Education Studies*. <https://www.mdpi.com/3042-5042/2/3/17>
- Elsevier. (2025). Generative AI policies for journals. <https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals>
- Forchini, P. (2025). Critical engagement with ChatGPT in linguistic analysis: A classroom-based study. *Research Methods in Applied Linguistics*. <https://www.sciencedirect.com/science/article/abs/pii/S2666799125000516>
- Funabashi, R. L. M. (2024). Generative AI in academic writing: English for academic purposes perspectives. BALEAP Master's Dissertation Award. <https://www.baleap.org/downloads/ma-dissertation-award-winners/2025-rebecca-funabashi-generative-ai-in-academic-writing.pdf>
- Gabay, R. A. E., et al. (2026). Generative artificial intelligence for academic writing in higher education: A scoping review of applications, challenges, and implications. *International Journal of Education in Mathematics, Science and Technology*.
- Ganguly, A., Johri, A., Ali, A., & McDonald, N. (2025). Generative artificial intelligence for academic research: Evidence from guidance issued for researchers by higher education institutions in the United States. arXiv. <https://arxiv.org/abs/2503.00664>
- Geng, M., & Trotta, R. (2024). Is ChatGPT transforming academics' writing style? arXiv.
- Godoy, D. (2026). AI-mediated scaffolding for second language academic writing within a sociocultural framework. *Discover Education*. <https://link.springer.com/article/10.1007/s44217-026-01386-0>
- Godwin-Jones, R. (2024). Generative AI, pragmatics, and authenticity in second language learning. *Language Learning & Technology*.
- Hakim, A. (2025). Implementing AI literacy teaching in university-level L2 writing. *TESOL Journal*. <https://doi.org/10.1002/tesj.70050>
- Hamed, D. M. (2025). Authorial voice in Q1 linguistics and hard science journals. *Cogent Arts & Humanities*.
- Hao, J., & Jiang, Y. (2025). Are the titles of doctoral dissertations and research articles different? Observations from four STEM disciplines. *Journal of English for Academic Purposes*, 74, Article 101471.
- He, Y., & Bu, Y. (2025). Academic journals' AI policies fail to curb the surge in AI-assisted academic writing. arXiv. <https://arxiv.org/abs/2512.06705>
- Hysaj, A. (2025). Exploring the purposes and uses of generative artificial intelligence among multicultural undergraduate students. *Higher Education Research & Development*. <https://doi.org/10.1080/07294360.2025.2488862>
- Ibrahim, S. K. M. (2026). Generative AI in academic writing: A comparison of human- and ChatGPT-generated research article titles. *Humanities and Social Sciences Communications*. <https://www.nature.com/articles/s41599-026-06956-z>
- Jandric, P. (2026). GenAI and academic writing. *Postdigital Science and Education*. <https://link.springer.com/article/10.1007/s42438-026-00641-9>



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice



- Jia, C. (2026). Exploring generative AI-assisted EFL academic writing. *Journal of English for Academic Purposes*. <https://www.sciencedirect.com/science/article/abs/pii/S1475158526000445>
- Jiang, F. K. (2025a). Rhetorical distinctions: Comparing metadiscourse in ChatGPT-generated and student argumentative essays. *Journal of Second Language Writing*. <https://www.sciencedirect.com/science/article/abs/pii/S0889490625000134>
- Jiang, F. K. (2025b). Metadiscursive nouns in academic argument: ChatGPT vs. university student writing. *Journal of English for Academic Purposes*. <https://www.sciencedirect.com/science/article/abs/pii/S1475158525000451>
- Jiang, F. K. (2026). Exemplification in ChatGPT and student argumentative writing. *English for Specific Purposes*. <https://www.sciencedirect.com/science/article/abs/pii/S0898589826000422>
- Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Journal of Medicine, Surgery, and Public Health*. <https://www.sciencedirect.com/science/article/pii/S2666990024000120>
- Khuder, B. (2025). Enhancing disciplinary voice through feedback-seeking in AI-supported writing for publication. *Applied Linguistics*. <https://doi.org/10.1093/applin/amaf022>
- Khuder, B. (2026). A theoretical framework of collaborative authorial voice. *English for Specific Purposes*.
- Kim, J. (2025). Students-generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*. <https://link.springer.com/article/10.1007/s12528-025-09444-6>
- Kim, M. (2024). Exploring the potential of using ChatGPT for rhetorical move-step analysis in academic writing. *Journal of English for Academic Purposes*. <https://www.sciencedirect.com/science/article/abs/pii/S1475158524000900>
- Kong, X. (2024). A comparative genre analysis of AI-generated and scholar-written abstracts for English review articles in international journals. *Journal of English for Academic Purposes*. <https://www.sciencedirect.com/science/article/abs/pii/S1475158524001000>
- Lee, S. (2025). Generative AI (GenAI) in the language classroom. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2025.2498537>
- Lin, D., Zhao, N., Tian, D., & Li, J. (2025). ChatGPT as linguistic equalizer? Quantifying LLM-driven lexical shifts in academic writing. *arXiv*. <https://arxiv.org/abs/2504.12317>
- Lin, Z. (2025). Hidden prompts in manuscripts exploit AI-assisted peer review. *arXiv*. <https://arxiv.org/abs/2507.06185>
- Liu, X. J., Zou, B., & colleagues. (2025). Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, and oral peer feedback as feedback enhancement. *Journal of English for Academic Purposes*, 75, Article 101505.
- Luo, J. (2026). Generative artificial intelligence in education from a technology research perspective: A scoping review. *Smart Learning Environments*. <https://link.springer.com/article/10.1007/s44436-026-00031-9>
- Mak, M. H. C. (2025). Style, sentiment, and quality of undergraduate writing in the age of generative AI. *Computers and Education: Artificial Intelligence*.
- Mo, Z., & Crosthwaite, P. (2025). Exploring the affordances of generative AI large language models for stance and engagement in academic writing. *Journal of English for Academic Purposes*, 75, Article 101499. <https://doi.org/10.1016/j.jeap.2025.101499>

**Linguistic Fingerprints of AI-Mediated Academic English:
A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice**



- Moorhouse, B. L., et al. (2025). Generative AI tools and empowerment in L2 academic writing. System. <https://www.sciencedirect.com/science/article/pii/S0346251X25001897>
- Nguyen, A. (2024). Human-AI collaboration patterns in AI-assisted academic writing. Studies in Higher Education. <https://doi.org/10.1080/03075079.2024.2323593>
- Pan, E. T., et al. (2024). Human vs machine: Identifying ChatGPT-generated abstracts in gynecology and urogynecology. American Journal of Obstetrics and Gynecology.
- Park, J. (2025). A systematic literature review of generative artificial intelligence literacy in education. Computers and Education: Artificial Intelligence. <https://www.sciencedirect.com/science/article/pii/S2666920X25001274>
- Radtke, A. (2025). Generative AI in academic writing: Does information on authorship influence revision behavior? Computers and Education: Artificial Intelligence. <https://www.sciencedirect.com/science/article/pii/S2666920X2400153X>
- Rafi, M. S. (2026). Generative AI in doctoral dissertation writing. Discover Education. <https://link.springer.com/article/10.1007/s44366-026-0088-9>
- Raitskaya, L. (2025). The 2025 landscape of generative AI in scholarly writing and publishing. Journal of Language and Education. <https://jle.hse.ru/article/view/29876>
- Rentier, E. S. (2025). To use or not to use: Exploring the ethical implications of using generative AI in academic writing. AI and Ethics, 5, 3421-3425. <https://link.springer.com/article/10.1007/s43681-024-00649-6>
- Roe, J., & Perkins, M. (2024a). Generative AI in self-directed learning: A scoping review. arXiv. <https://arxiv.org/abs/2411.07677>
- Roe, J., & Perkins, M. (2024b). Generative AI and agency in education: A critical scoping review and thematic analysis. arXiv. <https://arxiv.org/abs/2411.00631>
- Sadigzade, Z. (2026). Metadiscourse patterns in human-written versus generative AI-authored research abstracts: A comparative corpus-based analysis. Agbel Journal.
- Sanz-Tejeda, A. (2025). The impact of generative AI on academic reading and writing in higher education: A systematic review. Frontiers in Education. <https://www.frontiersin.org/journals/education/articles/10.3389/educ.2025.1711718/full>
- Scopus. (2024). Content policy and selection. Elsevier. <https://www.elsevier.com/products/scopus/content/content-policy-and-selection>
- Shabir, A. (2025). Generative AI writing tools and academic writing in higher education: A systematic review. International Technology and Education Journal. <https://journal.unm.ac.id/index.php/INTEC/article/view/10494>
- Singh, J. K., et al. (2025). A novel academic writing technology for authorial voice. System. <https://www.sciencedirect.com/science/article/pii/S1560429226000703>
- Slimi, Z. (2026). A systematic critical review of generative AI's impact on authorship, pedagogy, and integrity (2023-2025). Frontiers in Education. <https://www.frontiersin.org/journals/education/articles/10.3389/educ.2026.1769680/full>
- Sussman, K., & Carter, D. (2025). Detecting effects of AI-mediated communication on language complexity and sentiment. arXiv .
- Tseng, J.-J. (2025). Using visual scaffolding to enhance the comprehensibility of English materials in science education: A genre-based approach. Journal of English for Academic Purposes, 75, Article 101500.
- Tudino, G. (2024). A corpus-driven comparative analysis of AI in academic discourse: Investigating ChatGPT-generated academic texts in social sciences. Journal



Linguistic Fingerprints of AI-Mediated Academic English: A Corpus-Based Study of Lexical Density, Stance, and Authorial Voice



- of Pragmatics.
<https://www.sciencedirect.com/science/article/abs/pii/S0024384124001694>
- van Niekerk, J. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence*.
<https://www.sciencedirect.com/science/article/pii/S2666920X24001450>
- Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*.
- Wang, C., Xu, W., & Tan, X. (2025). Beyond the human-AI binaries: Advanced writers' self-directed use of generative AI in academic writing. *arXiv* .
- Wang, Y. (2025). A scoping review of empirical studies on generative AI in language education. *Innovation in Language Learning and Teaching*.
<https://doi.org/10.1080/17501229.2025.2509759>
- Watson, S. (2025). The role of generative AI in academic and scientific authorship: An autopoietic perspective. *AI & Society*, 40, 3225-3235.
<https://link.springer.com/article/10.1007/s00146-024-02174-w>
- Woo, D. (2026). Developing English academic writing and AI literacy in an EAP context. *Innovation in Language Learning and Teaching*.
<https://doi.org/10.1080/17501229.2026.2636962>
- Wu, C. (2026). Exploring PhD students' utilization of generative AI in writing for publication. *Journal of English for Academic Purposes*.
<https://www.sciencedirect.com/science/article/pii/S1475158525001432>
- Xia, D. (2025). Understanding 'the' in L2 writing: Article use in formulaic academic expressions. *Journal of English for Academic Purposes*.
<https://www.sciencedirect.com/science/article/abs/pii/S1475158525000438>
- Xu, J. (2025). Harnessing AI in academic writing: The complex interplay of AI literacy and self-directed learning among university L2 learners. *System*, 135, Article 103855. <https://doi.org/10.1016/j.system.2025.103855>
- Yao, G. (2025). Can AI simulate or emulate human stance? Using generative AI to investigate stance in academic book reviews. *Journal of Pragmatics*.
- Yu, D., Li, L., Su, H., & Fuoli, M. (2024). Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis: The case of apology. *International Journal of Corpus Linguistics*, 29(4), 534-561 .
- Zeng, X. (2024). A comparative study of voice in Chinese English-major argument writing. *Journal of Second Language Writing*.
- Zhang, M. (2025a). A comparison of human-written and ChatGPT-generated academic texts. *Journal of English for Academic Purposes*, 78, Article 101571.
<https://www.sciencedirect.com/science/article/abs/pii/S1475158525000712>
- Zhang, M. (2025b). Human-written vs. ChatGPT-generated texts: Stance in academic abstracts. *Language Sciences*.
<https://www.sciencedirect.com/science/article/abs/pii/S0346251X25002520>
- Zhang, Z., et al. (2025). The role of generative AI and hybrid feedback in improving L2 academic writing skills. *Innovation in Language Learning and Teaching*.
<https://doi.org/10.1080/17501229.2025.2503890>
- Zhao, D. (2025). The impact of AI-enhanced natural language processing tools on writing proficiency. *Education and Information Technologies*, 30(6), 8055-8086.
- Zou, M., Lee, I., & colleagues. (2025). Doctoral student's strategy use in GAI chatbot-assisted L2 writing: An activity theory perspective. *Journal of English for Academic Purposes*, 76, Article 101521.
- Zou, Y. (2026). Comparative analysis of text readability and writing styles in AI-generated and human-written academic texts. *Humanities and Social Sciences Communications*.