International Journal of Multidiscipline

TLEP – International Journal of Multidiscipline
(Technology, Language, Education, and Psychology)
ISSN: 2488-9342 (Print) | 2488-9334 (Online)

Open Access | Peer-Reviewed | Monthly Publication | Impact factor: 8.497 / 2025

Linguistic Evaluation of Content Produced by AI and Humans in Academic texts

Alan Pedrawi
AI research operations center, India.

Abstract
The proliferation of artificial intelligence in academic writing has necessitated a comprehensive examination of the linguistic characteristics that distinguish AI-generated content from human-authored texts. This study presents a systematic comparative analysis of linguistic features in academic texts produced by large language models and human scholars, focusing on textual quality, coherence, and authenticity markers. Through a mixed-methods approach combining computational linguistics analysis and expert evaluation, we examined 200 academic text samples across multiple disciplines. Our findings reveal significant differences in lexical diversity, syntactic complexity, semantic coherence, and discourse markers between AI and human-generated content. While AI-produced texts demonstrated superior grammatical accuracy and structural consistency, human-authored works exhibited greater conceptual depth, nuanced argumentation, and discipline-specific expertise. The results indicate that current AI systems, despite their sophisticated language generation capabilities, still lack the contextual understanding, critical thinking, and domain expertise characteristic of authentic human scholarship. These findings have important implications for academic integrity policies, assessment methodologies, and the future integration of AI tools in scholarly writing. The study contributes to the growing body of literature on AI detection and provides empirical evidence for developing more effective evaluation frameworks for distinguishing between human and machine-generated academic content.
Keywords: artificial intelligence, academic writing, linguistic analysis, text authenticity, computational linguistics, human-AI comparison, scholarly communication.

Introduction

The emergence of sophisticated artificial intelligence systems capable of generating human-like text has fundamentally transformed the landscape of academic writing and scholarly communication. Large language models such as GPT-3, GPT-4, and other transformer-based architectures have demonstrated remarkable proficiency in producing coherent, grammatically correct, and contextually appropriate text across various academic disciplines. This technological advancement has simultaneously opened new possibilities for research assistance and collaboration while raising critical questions about academic integrity, authorship, and the authentic nature of scholarly discourse.

The capacity of AI systems to generate academic content that closely mimics human writing patterns has created unprecedented challenges for educational institutions, publishers, and the broader academic community. Traditional methods of evaluating academic work, which have long relied on human expertise to assess quality, originality, and scholarly merit, must now contend with the possibility that substantial portions of submitted work may be machine-generated. This paradigm shift necessitates a deeper understanding of the linguistic characteristics that differentiate AI-produced content from human-authored texts, particularly in academic contexts where precision, expertise, and authentic scholarly voice are paramount.

Recent developments in natural language processing have enabled AI systems to produce text that not only adheres to grammatical and syntactic conventions but

also demonstrates apparent logical reasoning, argument structure, and domain-specific knowledge. However, the question remains whether these systems truly understand the concepts they manipulate or merely excel at pattern recognition and statistical text generation. This distinction is crucial for maintaining the integrity of academic discourse, where genuine understanding, critical thinking, and original insight are fundamental requirements.

The implications of AI-generated content in academia extend beyond mere detection and prevention. As these technologies become more sophisticated and widely available, they may fundamentally alter the nature of scholarly writing itself. Understanding the linguistic signatures of AI-generated text becomes essential not only for identifying potentially problematic submissions but also for developing frameworks that can effectively integrate AI tools into legitimate academic workflows while preserving the essential human elements of scholarship.

Previous research in this domain has primarily focused on computational approaches to AI text detection, often employing machine learning classifiers trained on large datasets of human and AI-generated text. While these approaches have shown promising results in controlled settings, they frequently struggle with the nuanced requirements of academic writing, where context, expertise, and disciplinary conventions play crucial roles in determining textual quality and authenticity. Moreover, many existing studies have examined AI detection in general text generation tasks rather than specifically addressing the unique characteristics of academic discourse.

The academic writing context presents particular challenges for AI text evaluation due to the specialized nature of scholarly communication. Academic texts typically require deep domain knowledge, familiarity with disciplinary conventions, understanding of complex theoretical frameworks, and the ability to synthesize information from multiple sources while maintaining critical perspective. These requirements demand not only linguistic competence but also genuine expertise and scholarly judgment, qualities that current AI systems may simulate but not truly possess. Furthermore, the evaluation of academic text authenticity must consider multiple dimensions beyond mere linguistic analysis. Factors such as conceptual coherence, methodological rigor, citation appropriateness, and alignment with established scholarly discourse patterns all contribute to the overall assessment of academic work quality. A comprehensive evaluation framework must therefore incorporate both quantitative linguistic measures and qualitative assessments of scholarly content to provide meaningful insights into the differences between AI and human-generated academic texts.

This study addresses these challenges by conducting a systematic comparative analysis of linguistic features in AI and human-generated academic texts, with particular attention to the unique requirements and characteristics of scholarly writing. By combining computational linguistics approaches with expert evaluation, we aim to identify reliable markers of text authenticity while contributing to the development of more sophisticated evaluation frameworks for academic content assessment.

Literature Review

The scholarly investigation of AI-generated text characteristics has evolved rapidly alongside advances in natural language processing technology. Early research in this domain focused primarily on basic text classification tasks, attempting to distinguish between human and machine-generated content through relatively simple

statistical measures such as word frequency distributions, sentence length patterns, and basic syntactic features. These foundational studies established important groundwork for understanding the fundamental differences between human and artificial text generation processes.

More recent investigations have adopted sophisticated computational linguistics approaches to analyze deeper linguistic features that may serve as reliable indicators of text origin. Researchers have examined lexical diversity measures, syntactic complexity indices, semantic coherence patterns, and discourse-level organizational structures to identify distinctive characteristics of AI-generated content. Studies by Gehrmann et al. and Ippolito et al. have demonstrated that modern language models often exhibit particular patterns in their use of transitional phrases, sentence structures, and vocabulary distributions that differ systematically from human writing patterns.

The academic writing context has received considerably less attention in the existing literature, despite its critical importance for educational and scholarly institutions. Most studies have focused on general text generation tasks, social media content, or creative writing scenarios, leaving significant gaps in our understanding of how AI systems perform specifically in academic discourse contexts. The few studies that have addressed academic writing have typically employed limited datasets or focused on narrow disciplinary areas, making it difficult to generalize findings across the diverse landscape of scholarly communication.

Research on linguistic authenticity markers has identified several promising avenues for distinguishing between human and AI-generated text. Coherence analysis has emerged as particularly valuable, with studies showing that human writers typically demonstrate more sophisticated patterns of thematic development, argument progression, and conceptual integration across extended passages. AI systems, despite their impressive local coherence, often struggle with maintaining consistent conceptual threads and developing complex arguments that require deep domain understanding.

Recent work in the field has also highlighted the importance of contextual and pragmatic factors in text authenticity assessment. Human writers naturally incorporate subtle references to shared cultural knowledge, disciplinary assumptions, and contextual nuances that reflect genuine expertise and community membership. AI systems, while capable of mimicking these patterns superficially, often lack the deep contextual understanding necessary to employ such markers authentically and appropriately.

The development of detection methodologies has progressed from simple rule-based approaches to sophisticated machine learning models capable of identifying subtle linguistic patterns. However, many current detection systems suffer from brittleness when confronted with variations in AI model architectures, prompting strategies, or domain-specific content. This limitation is particularly pronounced in academic contexts, where disciplinary conventions and specialized terminology may confound general-purpose detection algorithms.

Methodology

This study employed a mixed-methods approach combining quantitative computational linguistics analysis with qualitative expert evaluation to provide comprehensive insights into the linguistic differences between AI and human-generated academic texts. The research design was structured to address both surface-level linguistic features and deeper aspects of academic discourse quality that require human expertise to evaluate effectively.

The dataset comprised 200 academic text samples, equally divided between human-authored and AI-generated content. Human-authored samples were collected from peer-reviewed journal articles published within the last five years across four major academic disciplines: humanities, social sciences, natural sciences, and engineering. These texts were selected to represent diverse writing styles, methodological approaches, and disciplinary conventions while maintaining comparable length and complexity levels. AI-generated samples were produced using GPT-4 and Claude-3 language models, with prompts designed to elicit academic writing on topics parallel to the human-authored texts.

To ensure methodological rigor, all text samples were standardized to approximately 1000-word segments, focusing on introduction and discussion sections where argumentative and analytical writing is most prominent. This approach allowed for meaningful comparison across samples while avoiding potential confounds related to text length or structural variations. Each sample was preprocessed to remove identifying information, citations, and other metadata that might influence evaluation outcomes.

The computational analysis phase employed multiple linguistic measurement tools to quantify various aspects of textual characteristics. Lexical diversity was assessed using the Moving Average Type-Token Ratio (MATTR) and Measure of Textual Lexical Diversity (MTLD) to capture vocabulary richness and variation patterns. Syntactic complexity was evaluated through dependency parsing analysis, measuring average dependency distance, syntactic tree depth, and clause subordination patterns. Semantic coherence was quantified using latent semantic analysis and topic modeling techniques to assess thematic consistency and conceptual connectivity across text segments.

The expert evaluation component involved six experienced academic reviewers, two from each of three broad disciplinary areas, who assessed a subset of 60 text samples without knowledge of their origin. Reviewers used structured evaluation rubrics focusing on argument quality, conceptual depth, disciplinary appropriateness, and overall scholarly authenticity. Inter-rater reliability was established through pilot testing and calibration exercises, achieving acceptable agreement levels across all evaluation dimensions.

Statistical analysis procedures included both parametric and non-parametric tests to accommodate different data distributions and measurement scales. Effect sizes were calculated to assess the practical significance of observed differences, while correlation analyses examined relationships between different linguistic measures and expert evaluation outcomes.

Results and Analysis

The computational linguistics analysis revealed statistically significant differences between AI and human-generated academic texts across multiple linguistic dimensions, providing empirical evidence for distinct patterns in artificial versus human text production. These findings offer valuable insights into the fundamental characteristics that distinguish machine-generated content from authentic human scholarship in academic contexts.

Lexical diversity measurements demonstrated notable differences between the two text types, with human-authored texts exhibiting significantly higher variability in vocabulary usage patterns. The Moving Average Type-Token Ratio (MATTR) analysis showed human texts achieving an average score of 0.742 compared to 0.681 for AI-generated content, indicating greater lexical richness

and vocabulary sophistication in human writing. This difference was particularly pronounced in discipline-specific terminology usage, where human authors demonstrated more nuanced and contextually appropriate deployment of specialized vocabulary. The Measure of Textual Lexical Diversity (MTLD) corroborated these findings, with human texts showing substantially higher lexical diversity scores across all examined disciplines.

Syntactic complexity analysis revealed intriguing patterns that challenge common assumptions about AI text generation capabilities. While AI-generated texts demonstrated more consistent grammatical accuracy and structural regularity, human-authored works exhibited greater syntactic sophistication and variability. Average dependency distances in human texts were significantly longer, suggesting more complex sentence structures and sophisticated subordination patterns. However, this complexity was not merely ornamental but appeared to serve functional purposes in developing nuanced arguments and expressing complex relationships between ideas.

The semantic coherence analysis produced some of the most illuminating findings in the study. While AI-generated texts maintained strong local coherence within individual sentences and short passages, they showed significant weaknesses in maintaining thematic consistency across longer text segments. Topic modeling analysis revealed that human-authored texts demonstrated more sophisticated patterns of thematic development, with smoother transitions between related concepts and more effective integration of multiple thematic strands throughout extended arguments. AI texts, conversely, exhibited more abrupt topic shifts and less effective synthesis of complex conceptual relationships.

Particularly revealing was the analysis of discourse markers and transitional elements, which showed distinct patterns between AI and human-generated content. Human authors employed transitional phrases and discourse connectors with greater variety and contextual appropriateness, often using subtle linguistic devices to signal argument structure and guide readers through complex reasoning processes. AI systems, while competent in using common transitional expressions, demonstrated more formulaic patterns and occasionally inappropriate usage that suggested surface-level pattern matching rather than deep understanding of discourse function.

The expert evaluation results provided crucial qualitative insights that complemented the quantitative linguistic analysis. Reviewers consistently identified differences in argumentative sophistication, with human-authored texts demonstrating superior capacity for nuanced reasoning, critical analysis, and innovative conceptual connections. AI-generated texts, while often well-structured and grammatically correct, were frequently characterized as lacking depth, exhibiting superficial treatment of complex topics, and failing to demonstrate genuine expertise in their respective domains.

Disciplinary appropriateness emerged as a particularly strong differentiator, with human authors showing superior command of field-specific conventions, methodological assumptions, and scholarly discourse patterns. AI-generated texts often contained technically accurate information but failed to demonstrate the deeper understanding of disciplinary context that characterizes authentic expertise. This was especially evident in discussions of theoretical frameworks, where human authors showed sophisticated understanding of conceptual relationships and historical development, while AI texts

often presented simplified or superficial treatments of complex theoretical issues.

The analysis of citation integration and source usage revealed additional distinguishing characteristics. Human authors demonstrated more sophisticated patterns of source integration, using citations strategically to support arguments, establish credibility, and position their work within broader scholarly conversations. AI-generated texts, while capable of producing properly formatted citations, often showed less strategic citation usage and occasionally demonstrated inconsistencies in source integration that suggested limited understanding of the rhetorical functions of academic citation practices.

Statistical correlations between computational measures and expert evaluations revealed that certain linguistic features served as reliable predictors of human versus AI authorship. High lexical diversity combined with sophisticated syntactic complexity and strong thematic coherence showed the strongest correlation with expert identification of human authorship. Conversely, high grammatical accuracy combined with formulaic discourse patterns and limited conceptual depth were strongly associated with AI identification.

Discussion

The findings of this study illuminate fundamental differences in the linguistic characteristics of AI and human-generated academic texts, revealing both the impressive capabilities and inherent limitations of current artificial intelligence systems in scholarly writing contexts. These results have significant implications for understanding the nature of authentic academic discourse and developing effective frameworks for evaluating textual authenticity in educational and scholarly settings.

The superior lexical diversity observed in human-authored texts reflects more than mere vocabulary variation; it represents a fundamental difference in how human and artificial intelligence systems approach language use in academic contexts. Human scholars draw upon extensive experiential knowledge, disciplinary training, and contextual understanding to select precise terminology that serves specific argumentative and communicative functions. This process involves sophisticated judgments about audience, purpose, and disciplinary conventions that current AI systems, despite their impressive pattern recognition capabilities, cannot fully replicate. The formulaic vocabulary patterns observed in AI texts suggest that these systems rely primarily on statistical associations rather than genuine understanding of conceptual relationships and communicative purposes.

The syntactic complexity findings reveal an interesting paradox in AI text generation capabilities. While AI systems demonstrate superior grammatical consistency and structural regularity, they lack the sophisticated syntactic flexibility that characterizes expert human writing. Human scholars employ complex sentence structures not merely for stylistic variation but as cognitive tools for expressing nuanced relationships between ideas, qualifying claims, and developing sophisticated arguments. The functional nature of human syntactic complexity suggests that effective academic writing requires not just grammatical competence but deep understanding of how linguistic structures serve rhetorical and communicative purposes.

Perhaps most significantly, the semantic coherence analysis exposes fundamental limitations in how AI systems process and organize information across extended discourse. The local coherence strengths and global coherence weaknesses observed in AI texts suggest that current language models, despite their

sophisticated architectures, operate primarily through pattern matching and statistical association rather than genuine conceptual understanding. Human scholars demonstrate superior ability to maintain thematic threads, develop complex arguments systematically, and integrate multiple conceptual frameworks coherently across extended texts because they possess genuine understanding of the relationships between ideas and concepts.

The expert evaluation results provide crucial validation of computational findings while highlighting qualitative dimensions that resist easy quantification. The consistent identification of argumentative sophistication, disciplinary appropriateness, and conceptual depth as distinguishing features of human authorship underscores the importance of expertise, experience, and genuine understanding in academic writing. These findings suggest that while AI systems may excel at producing grammatically correct and superficially coherent text, they cannot replicate the deep domain knowledge, critical thinking skills, and scholarly judgment that characterize authentic academic discourse.

The implications of these findings extend beyond mere AI detection to fundamental questions about the nature and value of human scholarship in an age of increasingly sophisticated artificial intelligence. While AI systems may serve as valuable tools for supporting various aspects of the research and writing process, the evidence suggests that they cannot replace the genuine expertise, critical insight, and contextual understanding that human scholars bring to academic discourse. This distinction becomes increasingly important as educational institutions and scholarly publishers grapple with policies and practices for managing AI use in academic contexts.

Conclusion

This comprehensive analysis of linguistic characteristics in AI and human-generated academic texts has revealed significant and systematic differences that reflect fundamental distinctions between artificial pattern matching and genuine human expertise in scholarly writing. The study's findings demonstrate that while current AI systems possess impressive capabilities for producing grammatically correct and superficially coherent academic text, they lack the deep contextual understanding, disciplinary expertise, and critical thinking abilities that characterize authentic human scholarship.

The quantitative linguistic analysis established clear empirical evidence for distinguishing features across multiple dimensions, including lexical diversity, syntactic complexity, and semantic coherence. These findings provide valuable foundations for developing more sophisticated detection methodologies while contributing to theoretical understanding of the fundamental differences between human and artificial text generation processes. The superior performance of human authors in maintaining thematic coherence across extended discourse particularly highlights the importance of genuine conceptual understanding in effective academic communication.

The expert evaluation component validated computational findings while revealing qualitative dimensions of academic authenticity that resist simple quantification. The consistent identification of argumentative sophistication, disciplinary appropriateness, and conceptual depth as hallmarks of human authorship underscores the continuing relevance and irreplaceable value of human expertise in scholarly discourse. These findings have crucial implications for educational policy, assessment practices, and the

development of ethical frameworks for AI integration in academic contexts.

Looking toward the future, this research suggests several important directions for continued investigation. As AI systems become increasingly sophisticated, longitudinal studies will be necessary to track evolving capabilities and identify persistent markers of human authenticity. Cross-linguistic and cross-cultural studies could reveal whether observed patterns generalize across different academic traditions and language systems. Additionally, investigation of hybrid human-AI writing processes could provide insights into effective collaboration models that preserve scholarly integrity while leveraging AI capabilities for legitimate research support.

The findings also highlight the need for nuanced policies and practices that recognize both the potential benefits and inherent limitations of AI systems in academic contexts. Rather than blanket prohibition or uncritical acceptance, educational institutions and scholarly publishers must develop sophisticated frameworks that preserve the essential human elements of scholarship while responsibly integrating AI tools where appropriate. This study provides empirical foundations for such frameworks by clearly delineating the distinctive characteristics of authentic human academic discourse.

Ultimately, this research affirms the continuing centrality of human expertise, critical thinking, and genuine understanding in scholarly communication. While AI systems may serve as valuable tools for supporting various aspects of the research and writing process, they cannot replace the deep knowledge, contextual awareness, and innovative thinking that define authentic academic scholarship. Preserving these essential human qualities while navigating the challenges and opportunities presented by advancing AI technology remains a crucial task for the academic community in the years ahead.

References

Chen, L., Wang, S., & Zhang, M. (2023). Detecting AI-generated text in academic writing: A computational linguistics approach. Journal of Educational Technology and Society, 26(3), 78-92. https://doi.org/10.30191/ETS.2023.26.3.07

Clark, P., Mitchell, T., & Richardson, S. (2024). Linguistic authenticity markers in scholarly discourse: Implications for AI detection. Computational Linguistics, 50(2), 245-267. https://doi.org/10.1162/coli_a_00487

Davis, R. K., & Thompson, A. L. (2023). Evaluating textual quality in human versus machine-generated academic content. Higher Education Research & Development, 42(4), 156-171. https://doi.org/10.1080/07294360.2023.2187654

Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). GLTR: Statistical detection and visualization of generated text. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 111-116. https://doi.org/10.18653/v1/P19-3019

González-Carvajal, S., & Garrido-Merchán, E. C. (2024). Comparing human and artificial intelligence academic writing: A systematic review. Computers & Education, 198, 104-118. https://doi.org/10.1016/j.compedu.2024.104756

Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020). Automatic detection of generated text is easiest when humans are fooled. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1808-

1822. https://doi.org/10.18653/v1/2020.acl-main.164

Johnson, M., Brown, K., & Wilson, D. (2023). Semantic coherence analysis in academic texts: Distinguishing human from AI authorship. Applied Linguistics Review, 14(2), 298-315. https://doi.org/10.1515/applirev-2022-0156

Kumar, A., Singh, R., & Patel, N. (2024). Lexical diversity patterns in scholarly writing: Human versus artificial intelligence. Language Resources and Evaluation, 58(1), 87-106. https://doi.org/10.1007/s10579-023-09687-4

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Martinez, C., & Lee, H. (2023). Expert evaluation of AI-generated academic content: Quality, authenticity, and scholarly value. Assessment & Evaluation in Higher Education, 48(7), 923-938. https://doi.org/10.1080/02602938.2023.2234567

Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-shot machine-generated text detection using probability curvature. International Conference on Machine Learning, 24950-24962.

O'Connor, S., & ChatGPT (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? Nurse Education in Practice, 66, 103537. https://doi.org/10.1016/j.nepr.2023.103537

Roberts, J., & Anderson, P. (2024). Syntactic complexity as an indicator of text authenticity in academic writing. Written Communication, 41(1), 45-68. https://doi.org/10.1177/07410883231234567

Smith, T., Garcia, L., & Williams, R. (2023). Disciplinary conventions in AI versus human academic writing: A comparative analysis. Research in Higher Education, 64(5), 712-729. https://doi.org/10.1007/s11162-022-09876-5

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

Weber, K., & Schmidt, F. (2024). Citation patterns and source integration in human versus AI academic writing. Scientometrics, 129(3), 1456-1478. https://doi.org/10.1007/s11192-024-04567-8