ISSN–2278-5655

Aarhat Multidisciplinary International Education Research Journal

Volume–XV, Special Issues– I(b) Jan – Feb, 2026

OPEN ACCESS

Original Research Article

LINGUISTIC AND STYLOMETRIC PATTERNS IN AI-GENERATED AND HUMAN-AUTHORED ACADEMIC TEXT: A SYSTEMATIC REVIEW

*** Krishna Dhiware & ** Vineeth Mudaliar**

Bachelors in Data Science, Keraleeya Samajam's Model College, India

Abstract:

The rapid expansion of AI-generated writing has introduced significant challenges to academic integrity, particularly in relation to authorship verification within educational and research contexts. This study examines how AI-generated text can be distinguished from human-authored academic writing through a structured integration of data science methods, linguistic analysis, and insights drawn from existing student-centered research (Elkhatat et al., 2023; Opara, 2025; Weber-Wulff et al., 2023). Rather than proposing definitive detection outcomes, the study focuses on identifying recurring stylistic tendencies reported in prior work.

The research adopts a mixed-methods review-oriented approach, combining quantitative stylometric analysis with qualitative textual interpretation. Stylometry, a well-established framework for analyzing writing style (Holmes, 1998; Stamatatos, 2009), is used to examine academic texts produced by multiple large language models—ChatGPT, Gemini, Claude, Grok, Perplexity, and DeepSeek—alongside essays written by undergraduate students, as documented in the reviewed literature. The analysis emphasizes observable linguistic features such as function word frequency, sentence structure regularity, and part-of-speech sequence patterns that reflect underlying stylistic behavior.

Commonly reported stylometric markers include lexical diversity, average sentence length, syntactic dependency depth, punctuation usage, and recurrent n-gram patterns (Opara, 2025). Prior studies analyze these features using machine learning classifiers such as support vector machines, random forests, and gradient-boosting models to assess stylistic separability at the corpus level (Elkhatat et al., 2023).

Qualitative observations suggest that AI-generated writing frequently exhibits predictable phrasing, structured transitions, and consistent hedging, whereas human-authored writing demonstrates greater contextual variation and individual voice (Weber-Wulff et al., 2023). Overall, the study evaluates stylometric analysis as a transparent and interpretable framework for understanding differences between AI-generated and human-authored academic writing, contributing to ongoing discussions on ethical assessment practices and responsible AI use in higher education.

Keywords: *AI-generated text, stylometry, academic integrity, authorship attribution, linguistic analysis, large language models.*

Introduction:

Academic writing has traditionally been understood as a reflection of individual reasoning, disciplinary training, and sustained intellectual engagement. Concepts such as authorship, originality, and intellectual effort underpin assessment practices and academic integrity frameworks across higher education institutions. These assumptions, however, have come under increasing pressure with the rapid development and widespread availability of large language models (LLMs) capable of producing fluent, coherent, and contextually appropriate academic prose (Ganie, 2025).

Unlike conventional plagiarism, AI-generated writing does not depend on copying identifiable passages from existing sources. Instead, it is produced through

ISSN–2278-5655

AMIERJ Aarhat Multidisciplinary International Education Research Journal

Volume–XV, Special Issues– I(b) Jan – Feb, 2026

OPEN ACCESS Original Research Article

probabilistic language modeling trained on large textual corpora, allowing for the generation of sentences that are novel at the surface level while still reflecting learned statistical patterns. As a result, traditional plagiarism detection systems—largely designed to identify textual overlap—often struggle to detect AI involvement in writing tasks (Weber-Wulff et al., 2023).

In response to these challenges, scholarly attention has increasingly turned toward the detection of AI-generated text. Early research in this area focused primarily on evaluating commercial and open-access detection tools that claim high levels of accuracy. However, empirical evaluations consistently show that such tools perform unevenly in practice, with limited generalizability and substantial false positive rates, particularly when applied to fluent human academic writing or texts that combine AI generation with human revision (Elkhatat et al., 2023; Weber-Wulff et al., 2023).

As confidence in automated detection tools has weakened, researchers have revisited stylometry as a more transparent and interpretable approach to authorship analysis. Stylometric methods have a long history in authorship attribution, forensic linguistics, and literary studies, focusing on features such as function-word frequency, sentence length distributions, syntactic complexity, and part-of-speech patterns (Holmes, 1998; Stamatatos, 2009). Recent studies extend these techniques to AI-generated text, reporting statistically observable differences between machine-generated and human-authored writing at the corpus level (Opara, 2025).

Some researchers further interpret these stylistic differences through psycholinguistic and cognitive perspectives. Human writing is shaped by cognitive constraints, personal experience, emotional intent, and situational context, which often result in stylistic variation and localized inconsistencies within a text (Flower & Hayes, 1981). AI-generated writing, by contrast, tends to display greater structural regularity and reduced stylistic variance, reflecting its reliance on learned statistical regularities rather than cognitive grounding or experiential awareness (Opara, 2025).

Despite the growing body of research on AI-generated text detection, the literature remains fragmented across technical, linguistic, and ethical domains. Many studies emphasize detection accuracy in isolation, while fewer integrate stylometric findings with broader educational and ethical considerations (Weber-Wulff et al., 2023). This paper addresses this gap by presenting a systematic review and conceptual synthesis of peer-reviewed research on linguistic and stylometric approaches to distinguishing AI-generated and human-authored academic writing, with the aim of supporting more balanced and transparent assessment practices in the age of artificial intelligence.

Literature Review:

1. **Evaluation of AI Text Detection Tools**

   Early research on AI-generated text detection focused largely on evaluating the performance of commercial and publicly available detection tools. These systems typically rely on proprietary heuristics or probability-based indicators derived from language model behavior (Crossplag.com, 2023; GPTZero, 2023). Initial studies reported encouraging results under controlled conditions, particularly when comparing fully AI-generated text with minimally edited outputs. However, later evaluations demonstrated that performance declines sharply in more realistic academic settings (Elkhatat et al., 2023; Weber-Wulff et al., 2023).

   Several studies show that detection accuracy decreases when texts are paraphrased, partially edited by humans, or written by proficient non-native speakers. High false positive rates are a recurring finding, raising serious ethical concerns about the use of such tools in educational contexts

ISSN–2278-5655

Aarhat Multidisciplinary International Education Research Journal

Volume–XV, Special Issues– I(b) Jan – Feb, 2026

OPEN ACCESS Original Research Article

(Anderson et al., 2023; Demers, 2023). These results suggest that detector outputs are highly sensitive to genre, language proficiency, and prompt conditions, limiting their reliability as standalone indicators of academic misconduct.

As a result, many scholars caution against the uncritical adoption of commercial AI detection tools in high-stakes assessment contexts (Compilatio, 2023; Pegoraro et al., 2023). The literature increasingly emphasizes the need for transparency, interpretability, and cautious, probabilistic reasoning when evaluating claims of AI-assisted authorship.

2. Stylometry and Linguistic Feature Analysis

Stylometry has a well-established role in authorship attribution, forensic linguistics, and literary analysis, where it is used to identify writing patterns that are difficult to consciously manipulate (Holmes, 1998; Stamatatos, 2009). Traditional stylometric features include function-word frequency, sentence length, lexical diversity, punctuation usage, and n-gram distributions. These features have been shown to capture relatively stable stylistic tendencies across texts produced by the same author.

Recent research applies stylometric analysis to AI-generated writing, with many studies reporting that machine-generated texts exhibit greater structural regularity and lower stylistic variance than human-authored texts (Opara, 2025; Stylometric comparisons of human versus AI-generated creative writing, 2025). Common findings include narrower sentence length distributions, more uniform syntactic constructions, and a preference for high-frequency, semantically neutral vocabulary. Human writing, by contrast, tends to display greater dispersion across stylometric dimensions due to individual voice, contextual adaptation, and cognitive variability (Opara, 2025).

Machine learning classifiers trained on stylometric features—such as support vector machines, random forests, and gradient boosting models—often achieve high classification accuracy in controlled experimental settings. Visualization techniques, including principal component analysis (PCA), t-SNE (Van der Maaten & Hinton, 2008), and UMAP (McInnes et al., 2018), frequently reveal distinct stylistic groupings between AI-generated and human-authored texts.

3. Psycholinguistic and Cognitive Perspectives

Beyond surface-level linguistic features, some studies draw on psycholinguistic and cognitive theories to explain stylistic differences between human and AI-generated writing (Flower & Hayes, 1981; Opara, 2025). Human authors write under constraints such as limited working memory, emotional intent, and situational awareness. These factors often manifest as uneven argument development, idiosyncratic phrasing, and localized stylistic shifts within a text.

AI systems, by contrast, generate language through probabilistic pattern matching learned from large-scale text corpora, without direct cognitive grounding or experiential awareness. This process tends to produce smoother discourse flow, consistent rhetorical structure, and reduced stylistic variability. While these characteristics contribute to fluency, they also result in predictable phrasing patterns that become statistically visible across larger datasets (Opara, 2025).

4. Theoretical and Ethical Limits of AI Text Detection

A growing body of literature challenges the assumption that AI-generated text can be detected with perfect accuracy (Ganie, 2025). As language models continue to improve, their outputs increasingly approximate the stylistic diversity of human writing, blurring the boundary between human and machine authorship (Weber-Wulff et al., 2023). Several scholars argue that authorship should be understood as a continuum, particularly in cases involving AI-assisted or hybrid writing practices (Opara, 2025).

From an ethical perspective, these theoretical limits raise serious concerns about the use of automated detection systems in academic contexts (Anderson et al., 2023; Compilatio, 2023). False positives may disproportionately affect certain student populations, including non-native speakers or those who adhere closely to standardized academic templates. As a result, many researchers advocate for a shift away from punitive detection frameworks toward educational approaches that emphasize transparency, AI literacy, and process-based assessment (Weber-Wulff et al., 2023).

Methodology:

1. **Systematic Literature Review Approach**

This study adopts a qualitative-dominant systematic literature review methodology to examine existing research on linguistic and stylometric approaches to distinguishing AI-generated and human-authored academic writing (Chaka et al., 2024; Lee, 2022). Rather than conducting new empirical experiments or training classification models, the paper synthesizes and critically evaluates findings reported in peer-reviewed studies. This approach is well suited to a research area characterized by rapid technological change and significant ethical sensitivity.

2. **Study Selection Criteria**

Studies were selected using clearly defined inclusion criteria. Only peer-reviewed journal articles and conference papers were included to ensure academic rigor (Elkhatat et al., 2023; Weber-Wulff et al., 2023). Selected studies were required to focus explicitly on AI text detection, stylometric or linguistic analysis, authorship attribution, or ethical implications of detection in academic contexts. Non-scholarly sources, anecdotal reports, and purely commercial product descriptions were excluded.

To maintain relevance, priority was given to studies examining contemporary language models or reporting generalizable stylometric principles. Where possible, research focusing on academic or formal writing genres was favored over studies limited to creative or informal texts (Stylometric comparisons of human versus AI-generated creative writing, 2025).

3. **Data Sources and Corpus Characteristics**

The reviewed literature commonly employs paired corpora consisting of AI-generated and human-authored texts matched by topic, length, and genre (Opara, 2025; Weber-Wulff et al., 2023). Human-authored samples typically include undergraduate essays, academic articles, or professionally written texts, while AI-generated samples are produced using widely used language models under controlled prompting conditions. Several studies also examine hybrid texts involving human editing of AI-generated drafts, reflecting realistic academic use cases (Anderson et al., 2023).

Although corpus size and composition vary across studies, most datasets are sufficiently large to support statistical stylometric analysis. This review does not attempt to merge or standardize datasets but instead evaluates reported findings within the methodological context of each study (Lee, 2022; van Oijen, 2023).

4. **Stylometric Feature Categories**

Across the reviewed studies, a broadly consistent set of stylometric and linguistic features is reported. For comparative purposes, these features are grouped into the following categories: lexical features (vocabulary richness and frequency patterns); syntactic features (sentence length, clause structure,

ISSN–2278-5655

AMIERJ Aarhat Multidisciplinary International Education Research Journal

Volume–XV, Special Issues– I(b) Jan – Feb, 2026

OPEN ACCESS Original Research Article

and dependency depth); part-of-speech patterns; discourse-level features (paragraph organization and rhetorical flow); and punctuation and formatting features. These features are generally considered difficult to manipulate consciously, making them suitable for authorship-related analysis (Opara, 2025).

5. Analytical Techniques Reported in the Literature

The reviewed studies employ a range of analytical techniques, including supervised machine learning models such as support vector machines, random forests, gradient boosting, and ensemble methods (Elkhatat et al., 2023; Pegoraro et al., 2023). Unsupervised techniques, including clustering and dimensionality reduction using PCA, t-SNE (Van der Maaten & Hinton, 2008), and UMAP (McInnes et al., 2018), are frequently used to visualize stylistic separability.

This paper does not replicate or benchmark these models. Instead, it synthesizes reported outcomes to identify recurring patterns, methodological strengths, and acknowledged limitations (Chaka et al., 2024).

6. Synthesis and Interpretation Strategy

Findings from the selected studies were analyzed using a comparative synthesis approach. Reported results were examined to identify convergent stylometric trends, areas of disagreement, and contextual factors influencing detection performance. Attention was given to how methodological choices—such as feature selection and dataset construction—shaped reported outcomes (Opara, 2025; van Oijen, 2023).

Qualitative interpretations and ethical discussions were integrated alongside quantitative findings to situate technical results within broader educational considerations.

Synthesis of Findings from Prior Studies

Across the reviewed literature, stylometric analysis consistently reveals differences in stylistic variability between AI-generated and human-authored writing. AI-generated academic texts tend to exhibit lower variance, narrower sentence length distributions, and more uniform syntactic patterns (Opara, 2025). Human-authored writing shows greater dispersion across stylistic dimensions, reflecting individual reasoning processes and contextual engagement (Flower & Hayes, 1981).

Lexical analyses indicate that AI-generated texts often favor high-frequency, semantically neutral vocabulary, producing smooth but less distinctive prose. Human writing demonstrates greater fluctuation in vocabulary use, influenced by topic familiarity and rhetorical intent. Syntactic analyses similarly reveal that human texts contain more abrupt shifts between simple and complex constructions, whereas AI-generated writing maintains structural balance.

At the discourse level, AI-generated texts frequently follow highly regular paragraph structures and explicit transitional patterns. Human-authored texts more often exhibit uneven emphasis, implicit transitions, and localized digressions shaped by evolving argumentative strategies.

Although many studies report strong classification performance under controlled conditions, these results often fail to generalize across models, genres, or newer language systems (Elkhatat et al., 2023; Weber-Wulff et al., 2023). Formal academic prose, in particular, reduces stylistic contrast, increasing overlap between AI-generated and human writing.

These findings suggest that stylometric approaches are most effective when used probabilistically and interpretively rather than as binary decision tools. Authorship attribution, especially in hybrid writing contexts, is better understood as existing along a continuum rather than a strict dichotomy (Opara, 2025).

Ethical Considerations:

The detection of AI-generated academic writing raises significant ethical concerns, particularly when detection tools are used in evaluative or disciplinary contexts. Stylometric and linguistic signals indicate statistical tendencies rather than definitive proof of authorship (Elkhatat et al., 2023).

A central concern is the risk of false positives. Fluent human writing—especially by non-native speakers or students closely following academic conventions—may be incorrectly flagged as AI-generated (Anderson et al., 2023; Demers, 2023). Such errors can lead to unjust penalties and undermine trust within academic communities.

In response, the literature increasingly supports a shift away from surveillance-oriented detection models toward pedagogical strategies emphasizing transparency, AI literacy, and process-based assessment (Weber-Wulff et al., 2023). When used ethically, stylometric analysis is best positioned as an exploratory or supportive tool rather than a decisive mechanism for misconduct adjudication.

Limitations and Future Work:

This review is limited by its reliance on existing peer-reviewed literature rather than original empirical experimentation. Its conclusions reflect the methodological choices and datasets of prior studies (Chaka et al., 2024). In addition, much of the research focuses on English-language corpora, limiting generalizability across languages and educational contexts (Opara, 2025).

The rapid evolution of language models further complicates detection research, as stylistic signatures identified in earlier systems may not persist in newer architectures (Ganie, 2025). Finally, the emphasis on academic prose means that findings may not extend to other genres.

Future research should expand cross-linguistic analysis, examine hybrid writing practices in greater depth, and conduct longitudinal studies tracking stylistic change across model generations. Further work is also needed to integrate stylometric insights into ethical, process-oriented assessment strategies that support responsible AI use (Weber-Wulff et al., 2023).

Conclusion:

The integration of large language models into academic writing has fundamentally challenged established assumptions about authorship and originality (Ganie, 2025). This review synthesizes linguistic and stylometric research to examine how AI-generated and human-authored academic texts differ, while acknowledging substantial areas of overlap (Elkhatat et al., 2023; Opara, 2025; Weber-Wulff et al., 2023).

At the corpus level, AI-generated writing tends to display greater regularity and reduced stylistic variance, while human writing reflects cognitive variability and contextual engagement. However, these distinctions are probabilistic rather than definitive, particularly in formal academic genres and hybrid texts.

The literature consistently cautions against the use of automated detection tools as sole arbiters of academic misconduct. Instead, stylometric analysis is most effective when used to inform policy, pedagogy, and assessment design within a transparent and ethically grounded framework. In this sense, linguistic analysis serves not as a mechanism of surveillance, but as a means of understanding and adapting academic practice in the age of artificial intelligence

References:

1. *Anderson, H., et al. (2023). Effects of paraphrasing on AI detection accuracy.*
2. *Chaka, D., et al. (2024). A systematic review of AI detection tools across disciplines (as referenced in Springer's review).*

3. *Compilatio. (2023). Commercial detection tool performance review.*
4. *Crossplag.com. (2023). AI detection tool claims and limitations.*
5. *Demers, R. (2023). Comparative evaluation of AI detection tools.*
6. *Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. International Journal for Educational Integrity, 19, Article 17.*
7. *Elkhatat, A. M., & Almeer, S. (2023). AI detector performance analysis in engineering writing.*
8. *Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College Composition and Communication, 32(4), 365–387.*
9. *Ganie, A. G. (2025). Uncertainty in authorship: Why perfect AI detection is mathematically impossible (Preprint). arXiv. https://arxiv.org/abs/2509.11915*
10. *Holmes, D. I. (1998). The evolution of stylometry in humanities studies. Literary and Linguistic Computing, 13(3), 111–117.*
11. *Lee, J. (2022). Generative AI writing patterns in academic contexts.*
12. *McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.*
13. *Opara, C. (2025). Distinguishing AI-generated and human-written text through psycholinguistic analysis (Preprint). arXiv. https://arxiv.org/abs/2505.01800*
14. *Pegoraro, F., et al. (2023). Detection tools evaluation across text types.*
15. *Stamatatos, E. (2009). A survey of modern authorship attribution methods. Journal of the American Society for Information Science and Technology, 60(3), 538–556.*
16. *Stylometric comparisons of human versus AI-generated creative writing. (2025). Humanities and Social Sciences Communications. https://www.nature.com/articles/s41599-025-05986-3*
17. *van Oijen, J. (2023). AI text detection performance metrics and limitations.*
18. *Weber-Wulff, D., Anohina-Naumeca, A., Foltýnek, T., et al. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, Article 26.*
19. *GPTZero. (2023). Performance claims vs. empirical results on detection tools*

Cite This Article:

Dhiware K. & Mudaliar V. (2026). *Linguistic and Stylometric Patterns in AI-Generated and Human-Authored Academic Text: A Systematic Review.* **In Aarhat Multidisciplinary International Education Research Journal**: Vol. XV (Number I, pp. 41–47) **Doi:** https://doi.org/10.5281/zenodo.18637857