



Forensic stylometric analysis of AI-generated Text Compared Pakistani Student Academic Writing: Detecting Linguistic Fingerprints

Kashif Ullah¹

Abstract

The rapid adoption of generative artificial intelligence (AI) in higher education has created new challenges for authorship verification and academic integrity, particularly in multilingual contexts such as Pakistan. This study investigates whether forensic stylometric analysis can distinguish AI-generated academic text from authentic academic writing produced by Pakistani university students through the identification of linguistic fingerprints. The study aims to compare the lexical, syntactic, discourse, stylistic, and authorship-related features of AI-generated and human-authored texts. A qualitative forensic stylometric approach was employed using a balanced corpus of 500 academic texts, comprising 250 authentic student writings and 250 AI-generated texts produced from identical writing prompts. Data were analyzed through systematic coding, categorization, thematic comparison, and interpretation of recurring linguistic patterns. The findings reveal consistent differences between AI-generated and human-authored writing, with authentic texts demonstrating greater lexical diversity, syntactic variability, discourse flexibility, and individualized stylistic markers, while AI-generated texts exhibit greater linguistic regularity and standardized patterns. The study concludes that forensic stylometric analysis provides a reliable framework for detecting linguistic fingerprints and strengthening authorship attribution in Pakistani higher education.

Article History

Received 16 April 2026

Revised 21 May 2026

Accepted 26 May 2026

Published 31 May 2026

OPEN ACCESS

Keywords

Forensic Stylometry; AI-Generated Text; Authorship Attribution; Linguistic Fingerprints

Introduction

In the modern world, technology advances day by day and generative artificial Intelligence (AI), especially large language models, that has fundamentally changed the academic writing style by producing text with high coherence, accurate grammar, and relevant context. Although LLMs provide advantages to education such as language support, idea generation, and writing assistance, they also raise serious concerns about academic integrity, authorship, and the authenticity of students' work (Kasneji et al., 2023; OpenAI, 2023).

As conventional plagiarism detection systems are not designed to identify original text, universities are increasingly challenged to differentiate between human-authored and AI-generated text. As a result, the emergence of forensic stylometry for the quantitative analysis of linguistic style for authorship attribution offers a promising approach for spotting delicate linguistic fingerprints that distinguish AI-generated texts from human writing (Stamatatos, 2009; Ippolito et al., 2020). Researchers forwarded arguments based on computational stylometry and forensic linguistics that even when writing on similar topics, every writer has relatively constant lexical, syntactic, and discourse-level patterns (Coulthard et al., 2017). But while comparing AI-generated writing with university students' writing, such stylometric features remain necessarily distinctive and raise open, realistic questions. In Pakistan, where higher education institutions are increasingly adopting digital learning

¹The author is a graduate Researcher at Faculty of Social Sciences and Department of English, Kohat University of Science and Technology, Kohat, Khyber Pakhtunkhwa, Pakistan. He can be reached at kashifarain948@gmail.com

environments, and students gain unlimited access to generative AI tools, this generates a progressively substantial issue. In Pakistani universities, students' linguistic patterns are shaped by local educational practices, bilingualism, and varying levels of English proficiency, as they typically write English as a second language. These characteristics may create substantial variations from the statistically improved language generated by large language models, but experimental evidence examining such variances remains rare.

There is relatively limited attention given to forensic stylometric evidence that is capable of distinguishing AI-generated academic writing from authentic student writing in multilingual contexts, as existing studies have mostly focused on AI-text detection using machine learning or investigated perception of AI-generated writing (Liang et al., 2023; Pegoraro et al., 2025). Additionally, most stylometric research has been conducted in Western educational contexts using native English corpora, which limits the generalizability of findings and makes them inapplicable to contexts where sociolinguistic and educational conditions differ, such as the South Asian context. Therefore, in the Pakistani context, this contextual and methodological gap underlines the necessity for research that incorporates forensic linguistics with computational stylometry to identify linguistic markers unique to Pakistani university students. To address this gap, the current study conducts a qualitative forensic stylometric examination to differentiate AI-generated text from Pakistani students' writing in English.

Based in forensic linguistics and stylometric theory, this research contributes to the emergent scholarship on AI authorship recognition by providing context-specific indications from Pakistan, where English functions as a second language and academic writing practices vary from those informed in existing literature. The findings are expected to advance theoretical consideration of digital authorship, support the development of more reliable AI-detection methodologies, and convey practical recommendations for universities, educators, and policymakers looking to support academic reliability while rationally integrating artificial intelligence into higher education.

Research Problem

Higher education faces substantial challenges due to the widespread availability of generative artificial intelligence (AI), which can undermine the authenticity of student academic writing. For traditional plagiarism detection software, it has become difficult to detect plagiarism as AI-generated texts are frequently original in wording (Kasneci et al., 2023). Accordingly, forensic stylometric analysis has been drawing growing attention as a means of scrutinizing linguistic fingerprints such as lexical selections, syntactic constructions, and discourse designs to differentiate AI-generated texts from human-authored writing (Stamatatos, 2009).

On the other hand, prevailing research has largely focused on emerging automated AI-detection models or approximating their precision, with insufficient attention to qualitative forensic examination of writing style. Most studies are conducted in native English backgrounds and in Pakistan, where English is used as a second language and remains underexplored (Liang et al., 2023). Contradictory perspectives in the previous literature also exist some argue that AI-generated text displays recognizable stylistic uniformities that allow trustworthy authorship attribution (Ippolito et al., 2020). In contrast, others dispute the idea of improvement in large language models; they contend that large language models increasingly shadow the linguistic differences between AI and human writing, thereby diminishing the competence of stylometric detection methods (Pegoraro et al., 2025).

Against this framework, the present study will address the question of whether qualitative forensic stylometric analysis can identify distinctive linguistic fingerprints that differentiate AI-generated academic texts from Pakistani students' academic writing in English.

Research Objectives

- 1) To identify the lexical, syntactic, and discourse-level features that characterize AI-generated academic texts.
- 2) To examine the linguistic and stylistic patterns present in Pakistani students' academic writing in English.
- 3) To compare the stylometric characteristics of AI-generated texts with those of Pakistani students' academic writing to determine distinguishing linguistic fingerprints.
- 4) To ascertain the forensic stylometric markers that can reliably differentiate AI-generated texts from human-authored academic writing.
- 5) To interpret the implications of identified linguistic fingerprints for authorship attribution and AI-generated text detection in higher education.
- 6) To propose a qualitative forensic stylometric framework for distinguishing AI-generated academic texts from Pakistani students' authentic academic writing.

Research Questions

- 1) What lexical, syntactic, and discourse features differentiate AI-generated texts from Pakistani students' academic writing?
- 2) Which stylometric indicators reliably function as indicators of authorship?
- 3) How can forensic stylometry subsidize improving AI-authorship detection in multilingual academic contexts?

Literature Review

The rapid development of generative artificial intelligence (AI), particularly transformer-based Large Language Models (LLMs), has primarily changed academic writing and scholarly communication. For higher education, using LLMs to create grammatically accurate, coherent, and contextually appropriate text that resembles human academic writing creates both challenges and opportunities (Kasneci et al., 2023). Although the use of these technologies create comfort in support learning through recruiting, summarizing, conversion, and idea generation along with that they create and upraised serious apprehensions about academic integrity, authorship verification, and the trustworthiness of existing plagiarism detection methods (Cotton et al., 2024; Dwivedi et al., 2023). In contrast to conventional plagiarism, AI-generated texts are frequently original in wording, making them tough to detect using similarity-based software.

The shift in scholars' attention toward forensic authorship attribution has resulted in a change in objectives: not only to recognize copied material but to determine whether a text has honest human authorship or machine-generated language (Mitchell et al., 2023). Because forensic linguistics and stylometric scrutinize the regular linguistics patterns like lexical choice, syntactic complexity, discourse association, and stylistics reliability that function as linguistics fingerprints exclusive to individual writer, they appears as a encouraging approaches in this context (Coulthard et al., 2017; Grant, 2020).

To systematically measure linguistic characteristics across multiple levels of language, stylometry has become one of the most dominant and widely adopted approaches for authorship attribution. Contemporary stylometric research provides a multidimensional representation of writing style, including lexical diversity, syntactic complexity, punctuation, phraseological patterns, function-word usage, readability indices, and discourse organization, not just simple word frequency counts (Kestemont et al., 2019; Stamatatos, 2009). Recent studies shows that there is fluent and semantic coherence in text which are generated by LLMs and it contain identifiable stylistics tendencies, that include high lexical uniformity, predictable sentence structure, balanced paragraph organization and having relative consistent discourse patterns (Pegoraro et al., 2025). On the contrary, there is greater stylistic variability in human writing, which is shaped by individual cognitive processes, educational

background, personal writing habits, and disciplinary knowledge (Hyland, 2019). However, literature remains inconclusive about the long-term detectability of AI-generated texts.

Some researchers argue that language models play a role in reducing observable linguistic differences (Pegoraro et al., 2025), while others argue that understated but quantifiable stylometric regularities remain recognizable through combined qualitative and computational analysis (Mitchell et al., 2023). These different perspectives advise that forensic authorship attribution should not rely only on isolated textual features but on multidimensional linguistic evidence. Multilingual educational environments face difficulty distinguishing between AI-generated text and human-authored documents. Existing forensic linguistic research has paid very little attention to English as a second language (ESL), largely focusing on native English corpora (Liang et al., 2023). In Pakistan, this limitation is particularly important, where English functions as the major language of higher education but is cultivated as a second language by most students.

The academic writing of Pakistani students reflects bilingualism, limited discourse conventions, first-language transfer, and changing levels of English proficiency, resulting in unique lexical, syntactic, and rhetorical features that may differ noticeably from native English writing (Mahboob, 2020; Rahman, 2021). Human-authored texts created by Pakistani students regularly display personalized writing styles, limited collocations, flexible cohesion strategies, and discipline-specific rhetorical patterns, all of which contribute to authentic linguistic variation (Canagarajah, 2022). Significantly, recent studies have shown that numerous AI detection systems bias against non-native English writers by incorrectly categorizing authentic ESL writing as AI-generated, raising concerns about equality, consistency, and cross-cultural rationality in AI detection research (Liang et al., 2023). It is suggested from the previous literature that a strong theoretical foundation is provided by forensic, stylometry, and computational authorship analysis for examining AI-generated academic writing.

On the other hand, several noteworthy gaps remain uncertain. Most research highlights the importance of machine learning and its precision but provides little clarification on how to differentiate AI-generated text from human-authored text (Grant, 2020). The transparency and interpretability of authorship judgment are limited because empirical studies rarely provide qualitative forensic linguistic interpretation alongside quantitative stylometric analysis. More glaringly, suggestions from multilingual educational settings, mainly South Asian contexts, remain limited despite the growing prevalence of generative AI in universities. The literature also presents contradictory opinions regarding whether increasingly refined LLMs will eventually remove detectable stylistic alterations or whether dogged linguistic fingerprints will endure to differentiate machine-generated language from authentic human writing (Mitchell et al., 2023; Pegoraro et al., 2025). To address the methodological, theoretical, and contextual boundaries of Pakistani student's academic writing need focused examination through an integrated forensic stylometric outline that restraints qualitative linguistic analysis with computational suggestion. Such research will endorse not only authorship acknowledgement but also a wider understanding of academic integrity in the era of generative artificial intelligence.

Research Gap

The present research is supported by the review of earlier studies that reveal several important gaps. Firstly, significant scholarly attention has been gained by AI-generated academic content has attracted significant scholarly attention, and the investigation reveals that there is no qualitative scrutiny of forensic fingerprints. Secondly, most stylometric research has been conducted using native-English datasets in Western educational contexts, with limited evidence from multilingual contexts where English is used as a second language. Thirdly, within the forensic authorship framework, there has been little research that has scrutinized the linguistic features of Pakistani university students' academic writing. Finally,

very little focus has been placed on incorporating forensic linguistics and qualitative stylometric analysis to associate AI-generated texts directly with authentic student writing created under similar academic conditions.

A strong indication arises from these methodological and contextual limitations that point to the clear need for qualitative research that methodically examines the linguistic fingerprints distinguishing AI-generated texts from Pakistani students' academic writing. The present study aims to address this gap by connecting forensic stylometric analysis to the reidentification of lexical, syntactic, discourse, and stylistic structures that support consistent authorship attribution within the Pakistani higher education context.

Research Conceptual Framework

The present study deals with forensic stylistics and conceptual stylometry, that argue that every writer has their own linguistic outline that acts as a recognizable stylistic fingerprint. (Coulthard et al., 2017; Grant, 2020). These linguistic fingerprints are exposed through word choices, syntactic inclination, discourse association, cohesion and coherence approaches, and stylistic uniformity. The central phenomenon that is under examination within this framework is established through text authorship (AI-generated versus human-authored). The study examines whether modifications in lexical structures (vocabulary variety, collocations, lexical complexity), syntactic structures (sentence complication, grammatical assemblies, clause organization), discourse structures (cohesion, coherence, paragraph association), and stylistic indicators (punctuation, phraseology, function-word distribution, and repetition) cooperatively form unique linguistic fingerprints capable of distinguishing AI-generated academic texts from Pakistani students' authentic writing. The framework assumes that although advanced AI systems can imitate many features of human language, authentic student writing still reveals individual cognitive development, educational considerations, bilingual stimuli, and sociocultural experience. These modified linguistic characteristics are likely to provide adequate forensic evidence for qualitative authorship diversity (See Figure 1 for more details).

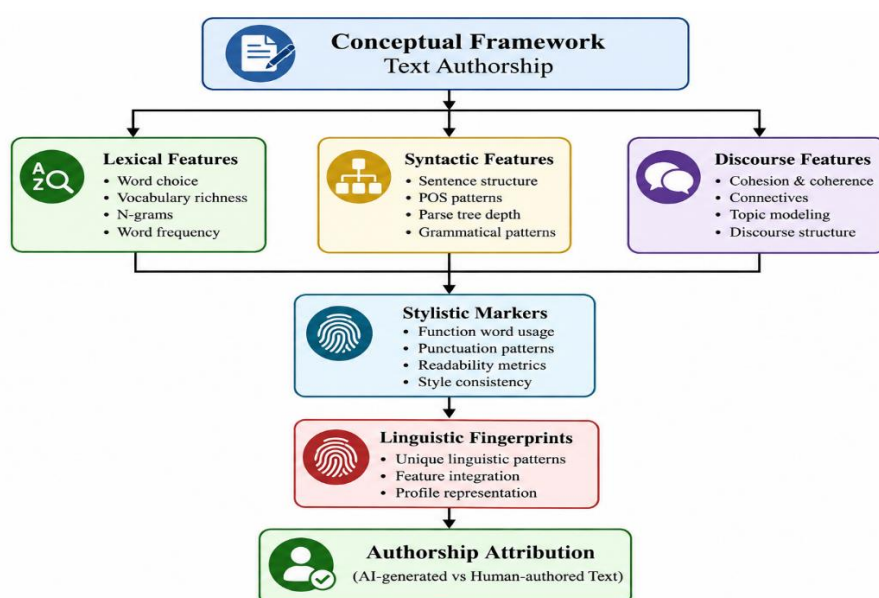


Figure 1: Conceptual Framework

Research Methods

The existing study implements a qualitative comparative research design grounded in forensic linguistics and stylometric analysis to examine the linguistic variances between AI-generated academic texts and Pakistani students' academic writing. A qualitative design is

suitable because the study seeks to discover and interpret linguistic fingerprints rooted in written texts rather than measure associations statistically. This method eases an in-depth examination of lexical, syntactic, and discourse-level structures that characterize authorship and writing style (Creswell & Poth, 2018).

The study population comprises English academic texts created by undergraduate and postgraduate students enrolled in public and private universities in Pakistan, together with AI-generated texts produced using extensively available large language models. Each writing sample will be treated as a unit of analysis and selected through purposive sampling to ensure that the selected texts are pertinent to the research objectives. Authentic student assignments will be obtained through association with university departments and course instructors. At the same time, AI-generated texts will be produced using matching writing prompts to ensure comparability between the two datasets.

Data collection involve gathering a balanced corpus of student-written and AI-generated academic texts. Before participation, students will receive an information sheet delineating the purpose of the study, and written informed consent was obtained for the use of their anonymized writing samples. Participation was completely voluntary, and participants may withdraw from the study at any stage without concern. To guard privacy, all personal identifiers, including names, registration numbers, and institutional affiliations, was removed, and each text was allocated a unique identification code. Digital files were stockpiled securely on password-protected devices reachable only to the researcher. Participants were also be provided with the researcher's institutional email address to pursue clarification or raise concerns regarding the study.

The collected corpus was examined using qualitative forensic stylometric analysis. Qualitative forensic stylometric analysis was used to examine the collected corpus. To recognize linguistic fingerprints related to AI-generated and human-authored text, the main focus of analysis was lexical choices, vocabulary richness, syntactic complexity, cohesion, coherence, punctuation patterns, and recurrent stylistic features. To identify distinctive stylistic markers of authorship, comparative textual analysis was employed on both corpora. A qualitative design is suitable for forensic authorship as it provides rich contextual insights into writing styles and linguistic behavior. Conversely, as the findings are grounded in purposively selected texts, they may not be statistically generalized to all student populations or AI-generated writing systems (Tracy, 2020).

Population

The target population comprises students enrolled in BS English (Linguistics), Master's, MPhil, and PhD programs at public and private universities in Khyber Pakhtunkhwa, Pakistan. The authentic corpus consisted of students enrolled in BS English (Linguistics), Master's, MPhil, and PhD programs at Kohat University, the University of Peshawar, Khushal Khan Khattak University, the University of Science and Technology Bannu, and Gomal University. Students are selected based on their regular production of academic texts in English, which provide authentic texts suitable for forensic stylometric analysis.

Sampling and Sample Size

Purposive sampling, a widely accepted non-probability sampling technique in qualitative research, is used to select information-rich cases that directly address research objectives (Patton, 2015). One original academic writing sample and one AI-generated text with a similar writing prompt will be collected from each participant; a total of 250 original and 250 AI-generated corpora will be collected to enable systematic comparison. To identify recurring stylometric patterns, the balance corpus of 500 texts will provide sufficient linguistic diversity while maintaining the depth of qualitative analysis required in forensic linguistic research. The principal strength of the purposive sampling strategy is its ability to generate

context-specific data from participants, and it seeks participants that possess specific linguistic and academic characteristics rather than statistical analysis.

Data Collection and Data Analysis Procedure

The current study uses a primary research instrument, a researcher-developed forensic stylometric framework, to examine the linguistic fingerprints of AI-generated academic texts and Pakistani students' writing. The study focuses on naturally occurring written texts, as they represent the authentic linguistic behavior of authors. The informed framework is built by the principles of forensic linguistics and stylometry and is designed to recognize recurrent stylistic patterns that distinguish AI-generated writing from human-authored academic texts (Coulthard et al., 2017; Grant, 2020).

The items of analysis comprise academic essays written in English by BS, Master's, MPhil, and PhD Linguistics students enrolled in Pakistani universities, organized with AI-generated texts created using identical writing prompts. To ensure comparability of the two balanced corpora: a Human Academic Writing Corpus and an AI-Generated Writing Corpus, all text was organized in similar writing conditions. Five analytical dimensions are used by the stylometric framework to operate the principal concepts of the study.

Lexical features include vocabulary abundance, lexical range, word regularity, reiteration, collocations, and lexical superiority. Syntactic features encompass sentence length, sentence complication, clause construction, grammatical constructions, and passive voice habit. Discourse features include cohesion, coherence, paragraph association, logical sequencing, and transitional devices. Stylistic markers involve punctuation designs, phraseological inclinations, function-word distribution, and frequent stylistic choices. Finally, authorship indicators focus on the consistency and uniqueness of linguistic patterns that function as forensic fingerprints distinguishing AI-generated texts from authentic student writing. These dimensions align directly with the study's objectives by enabling a systematic comparison between two corpora. To suit the Pakistani English academic context, the coding framework is adapted from established stylometric and forensic linguistics research (Grant, 2020; Nini, 2018).

Adaptations like localized lexical choices, syntactic transfer, and discourse organization are incorporated as linguistic characteristics commonly observed in English as a second language. Without altering the theoretical foundation, these contextual modifications enhance the instrument's relevance. Specialists in forensic and applied linguistics review the instrument before data analysis to ensure trustworthiness. The strength and credibility of findings were increased through repeated coding, systematic comparison of textual features, and maintenance of an audit trail documenting coding decisions. The strength of the framework lies in its capacity to provide in-depth qualitative understanding of linguistic fingerprints that cannot be captured by automated detection systems alone. To minimize researcher bias, qualitative stylometric analysis requires a careful coding procedure. Transparent analytical procedures and continuous comparison of coding categories throughout the study was used to address this limitation (Lincoln & Guba, 1985; Tracy, 2020). These limitations were addressed through transparent analytical procedures and continuous comparison of coding categories throughout the study.

Thematic Data Analysis Procedure

Qualitative forensic stylometric analysis was used to compare the linguistic fingerprints of AI-generated academic text and Pakistani students' academic writing. A systematic process involving data familiarization, coding, categorization, comparison, and interpretation were followed in the data analysis (Creswell & Poth, 2018). At the start, two corpora of text were created: student-authored academic text and AI-generated text. To gain a comprehensive understanding of the linguistic characteristics of the texts, they were read repeatedly.

Predetermined stylometric categories derived from forensic linguistics were applied to the texts accordingly (Coulthard et al., 2017; Grant, 2020). To identify the recurring similarities and differences between the two texts, a comparative analysis was conducted after coding. These patterns will be interpreted as potential linguistic fingerprints that consistently distinguish AI-generated text from human-authored text (Muhammad, 2025). To ensure transparency and consistency in the interpretation of findings, the researcher maintains detailed analytical memos and an audit trail throughout the analysis. Coding decisions were reviewed through peer consultation and continuous comparison across texts to enhance the credibility of the analysis (Tracy, 2020).

The findings were presented thematically, with each refrain maintained by illustrative textual quotations and connected to the study purposes and pertinent theoretical perspectives in forensic linguistics and stylometry. This analytical method allows an in-depth understanding of authorship characteristics while preserving the contextual richness of authentic academic writing (See Figure 2).

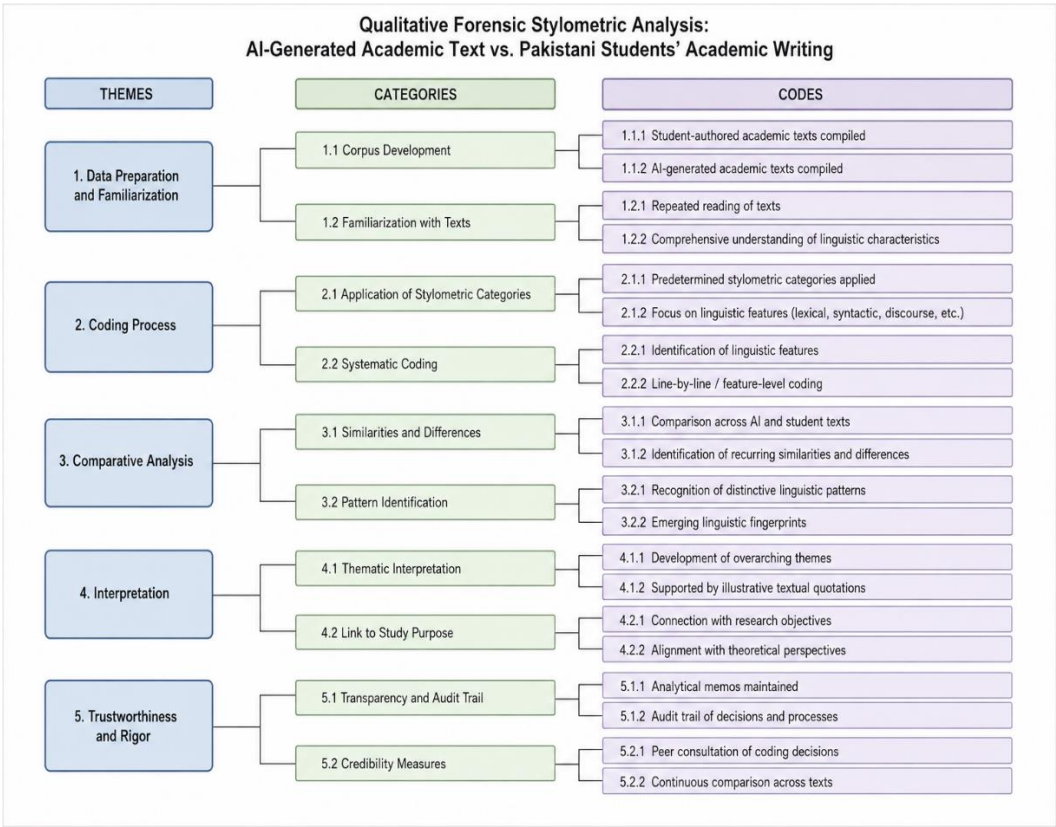


Figure 2: Thematic Analysis Process

Ethical Consideration

The study was conducted in accordance with the ethical guidelines governing research involving human participants. Ethical approval was obtained from the relevant Institutional Review Board (IRB) of the participating university before data collection. Permission to collect students' academic writing was also sought from the respective university authorities and course instructors. Written samples were anonymized, and coded identifiers were used for names, registration number, university affiliation, and other identifying information before data analysis.

Findings and Discussion

The current study uses a forensic stylometric framework to examine linguistic characteristics of AI-generated text and authentic academic writing created by Pakistani university students. A systematic comparison of a balanced corpus consisting of 500 academic texts (250 authentic student writings and 250 AI-generated texts produced from identical writing prompts) was conducted using lexical, syntactic, discourse, stylistic, and authorship-related features (See Figure 3 for more detail).

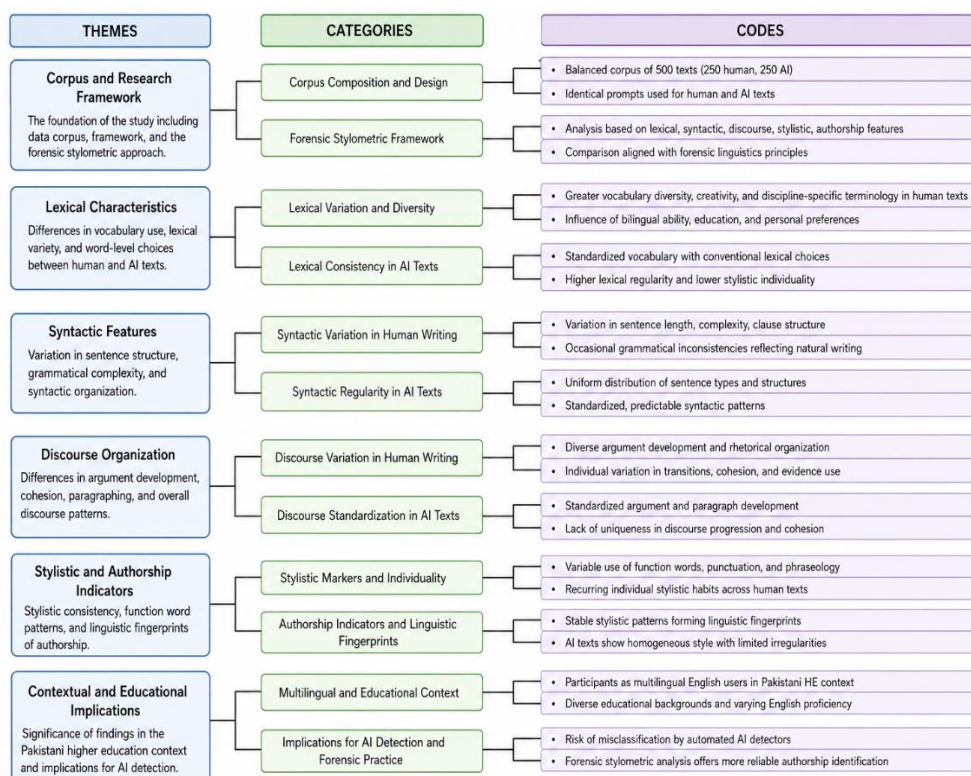


Figure 3: Thematic Analysis Process

The findings indicate that there is a clear and consistent stylometric difference between AI-generated text and authentic student academic writing, instead of increasing linguistic sophistication of contemporary large language models (LLMs). The central assumption of forensic linguistics is that every writer or writing system produces relatively stable linguistic patterns that can function as identifiable stylistic markers, as supported by these findings (Coulthard et al., 2017; Grant, 2020).

The most prominent concern from the findings is lexical variation. There is greater variability in vocabulary selection, lexical creativity, discipline-specific terminology, and individualized word choices in authentic writing. The writing of students was frequently influenced by their educational experience, bilingual competence, and personal writing preferences that result in natural diversity across the corpus. This contrasts with AI-generated text, which is highly standardized, uses conventional lexical choices, and features semantically balanced expression. Though AI text has high grammatical accuracy and lexical sophistication, it shows greater lexical regularity and lower stylistic individuality. Recent research shows that these observations are consistent and that transformation-based language models optimize statistically probable lexical sequences rather than reflecting the individualized linguistic behavior characteristic of human authorship (Mitchell et al., 2023; Pegoraro et al., 2025).

From a forensic viewpoint, lexical consistency itself becomes a classifying feature because genuine human writing naturally contains greater variation arising from cognitive

decision-making, educational background, and personal experience. Syntactic analysis further reinforced the distinction between human and AI-generated writing. There is noticeable variation in sentence length, grammatical complexity, clause structure, and rhetorical organization in human author text. Natural individual writing has grammatical inconsistencies that occur occasionally. At the same time, AI-generated text has a highly standardized, predictable structure, and a uniform distribution of complex and simple sentences. Stylometric research observed similar findings that demonstrate that AI-academic writing has uniform regularities in sentence structure and exhibits lower structural variability than authentic human writing (Kasneji et al., 2023; Liang et al., 2023). From the discussion, it is concluded that syntactic features are an important indicator when distinguishing machine-generated text from authentic student writing.

At the discourse level, it also demonstrates differences between Machine-Generated text and authentic writing. Authentic student writing is filled with greater variation in argument development, paragraph transitions, cohesion strategies, and organizational patterns. Individual authentic writings frequently differ in the use of connectors, supporting evidence, rhetorical emphasis, and overall discourse progression. In contrast, AI-generated texts have standardized argument development, paragraph development, and a lack of uniqueness across all aspects of the text. This view aligns with preceding studies indicating that LLMs generate discourse structures optimized for coherence and readability through learned statistical regularities, whereas human discourse reflects the influence of cognition, educational experience, and individual rhetorical choices (Dwivedi et al., 2023; Hyland, 2019). Accordingly, discourse organization emerged as another fit indicator of authorship within the present corpus.

The analysis of stylistic markers and authorship indicators provides the strongest support for the study's forensic objective. There is variable use of function words, individualized punctuation, distinctive phraseological choices, and recurring stylistic habits that remained relatively consistent in human-authored text and also differed across participants. These frequent stylistic features establish what forensic linguists pronounce as an author's linguistic fingerprint. Where there is homogeneous stylistic behavior, characterized by punctuation, predictable phraseology, and uniform function-word distribution in AI-generated texts. AI does not consistently reproduce the subtle irregularities that naturally emerge through human cognition and individual writing experiences. These findings reinforce authorship attribution theory by demonstrating that stable stylistic variation remains a distinguishing characteristic of authentic human writing even when compared with advanced generative AI systems (Grant, 2020; Kestemont et al., 2019).

An important contribution of the present study lies in its Pakistani higher education context, which remains underrepresented in forensic stylometric research. These participants represent multilingual English users whose academic writing reflects diverse educational experiences and varying levels of English proficiency. Despite this linguistic diversity, authentic texts consistently displayed individualized stylistic features that differed from AI-generated writing. This finding is particularly significant because previous studies have highlighted the risk of AI detection systems misclassifying non-native English writing as machine-generated (Liang et al., 2023). The present findings suggest that forensic stylometric analysis based on multiple linguistic dimensions provides a more reliable approach than relying solely on automated AI detection tools, particularly within multilingual educational settings.

Conclusion

In conclusion, the findings demonstrate that AI-generated academic writing and authentic Pakistani student writing remain distinguishable through multidimensional forensic stylometric analysis. Although contemporary LLMs generate highly fluent, grammatically accurate, and coherent academic prose, they continue to produce systematic linguistic

regularities that differ from the naturally variable writing patterns of human authors. The combined analysis of lexical, syntactic, discourse, stylistic, and authorship-related features therefore provides strong evidence that linguistic fingerprints remain identifiable despite ongoing advances in generative artificial intelligence. These findings support the growing consensus within forensic linguistics that authorship attribution should rely on integrated qualitative and computational analyses rather than isolated textual characteristics. Consequently, the present study contributes to both forensic linguistics and educational integrity by demonstrating the continued value of stylometric analysis for distinguishing authentic student writing from AI-generated academic texts in higher education.

Problems and Limitations

Several methodological and practical challenges may be encountered by the study regardless of careful planning including participants' willingness and institutional approval for sharing authentic written work, which could delay data collection. Concerns about privacy or misconceptions that the research is intended to evaluate academic misconduct create hesitation among students.

The rapidly evolving nature of generative AI creates further limitations. Continuous updates to large language models also update their linguistic output. Therefore, the identified stylistic patterns in this study may solely reflect the version of the AI system studied during this research period. Moreover, the study focuses exclusively on English academic writing produced by Pakistani university students; therefore, the findings may not apply to other linguistic, cultural, or educational contexts.

Methodologically, qualitative stylometric analysis relies on the researcher's interpretation of linguistic features, which may introduce subjective judgment. To minimize this limitation, the study employs a structured coding framework, maintain an audit trail, and apply consistent comparative analysis throughout the research process. Although purposive sampling limits statistical generalization, it enables an in-depth exploration of information-rich texts, which is consistent with the objectives of qualitative inquiry (Creswell & Poth, 2018). These limitations acknowledged when interpreting the findings and provide direction for future research on AI authorship detection in multilingual academic settings.

Author Contributions:

The conception and design, collection and assembly of data and Analysis and interpretation of the data, the intellectual content management and corresponding Author and her Email: kashifarain948@gmail.com

References

- British Educational Research Association. (2024). *Ethical guidelines for educational research* (5th ed.). British Educational Research Association.
- Canagarajah, A. S. (2022). *Translingual practice: Global Englishes and cosmopolitan relations* (2nd ed.). Routledge.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence* (2nd ed.). Routledge.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Grant, T. (2020). Quantifying evidence in forensic authorship analysis. *International Journal of Speech, Language and the Law*, 27(1), 1–25.
- Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
- Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020). Automatic detection of generated text is easiest when humans are fooled. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1808–1822). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.164>
- Kasneci, E., Sessler, K., Küchemann, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kestemont, M. (2014). Function words in authorship attribution: From black magic to theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature* (pp. 59–66).
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Mahboob, A. (2020). Pakistani English. In D. Schreier, M. Hundt, & E. W. Schneider (Eds.), *The handbook of World Englishes* (2nd ed.). Wiley.
- Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). *DetectGPT: Zero-shot machine-generated text detection using probability curvature*. arXiv:2301.11305.
- Muhammad, O. (2025). Exploring the Challenges and Pedagogical Approaches of English Language Teachers in Government and Private Schools. *CEERAT Journal of Society and Development*, 2(2), 45–55. <https://ceeratjsd.com/index.php/home/article/view/15/18>
- Nini, A. (2018). *Authorship attribution and forensic linguistics*. Cambridge University Press.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE.
- Pegoraro, M., Kumari, A., Fokkens, A., & Daelemans, W. (2025). Authorship attribution and AI-generated text: A systematic review of stylometric approaches. *Natural Language Engineering*. Advance online publication.
- Rahman, T. (2021). *Pakistani English: The linguistic description of a non-native variety of English*. Oxford University Press.
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, 60(3), 538–556. <https://doi.org/10.1002/asi.21001>
- Tracy, S. J. (2020). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (2nd ed.). Wiley.