JPPIPA 12(4) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Evaluating AI Detection Performance Across Writing Modalities in Engineering Interaction

Meidi W. Lestari[1*], Agustina Ginting[1], Harris Aminudin[1], C. Cholish[1], Ulfa Hasnita[1]

[1] Department of Electrical Engineering, Politeknik Negeri Medan, Medan, Indonesia.

Received: November 04, 2024
Revised: August 21, 2025
Accepted: April 25, 2026
Published: April 30, 2026

Corresponding Author:
Meidi W. Lestari
meidilestari@polmed.ac.id

DOI: 10.29303/jppipa.v12i4.9637

Abstract: This study proposes a conceptual framework for AI-based detection of handwritten and technical English writing in electrical engineering students, integrating multiple input modalities, including handwritten, typed, hybrid, and transcribed texts. A neural network–based detection system was evaluated using classification metrics and statistical analysis. The results demonstrate that detection accuracy varies significantly across modalities, ranging from 0.72 to 0.89, with typed and transcribed texts achieving higher consistency. Mean detection scores increased progressively from handwritten (0.41-0.50) to typed (0.52-0.58), hybrid (0.60-0.66), and transcribed texts (0.68-0.76), indicating the influence of text standardization. Variability was highest in handwritten and hybrid categories, with standard deviation reaching 0.10–0.11 and wider confidence intervals (e.g., 0.36-0.56), reflecting classification uncertainty. Heatmap analysis further shows that precision and recall in transcribed texts can exceed 0.85-0.89, while handwritten texts drop below 0.45 in some cases. The findings reveal that L2 writing characteristics and technical language significantly affect detection outcomes, leading to potential misclassification. This study highlights the need for context-aware AI detection systems that incorporate modality-specific features, ensuring fairness and reliability in engineering education.

Keywords: AI detection; Education; Engineering; L2 writing; Writing modality

Introduction

The rapid advancement of artificial intelligence (AI), particularly large language models (LLMs), has significantly transformed academic writing practices across higher education. Tools such as generative AI systems are increasingly capable of producing fluent, coherent, and domain-specific text, raising critical concerns regarding academic integrity, authorship, and assessment validity (Smith 2007; Sokolova & Lapalme 2009; Winskel, 2014; Biber & Gray, 2016; Kyle & Crossley 2018,; Gehrmann et al., 2019; Ippolito et al., 2020). In technical disciplines such as electrical engineering, where precise terminology and structured reasoning are essential, the ability of AI systems to mimic human writing introduces new challenges in distinguishing authentic student work from machine-generated content (Pasari et al., 2021a, 2021b).

Recent developments in AI detection systems have attempted to address this issue by analyzing linguistic patterns, stylometric features, and probabilistic text characteristics to classify writing as human or AI-generated (Clark et al., 2021; Dwivedi et al., 2023; Kasneci et al., 2023; Weber-Wulff et al., 2023). However, existing detection approaches are often trained on general-domain text and may not perform reliably in specialized contexts such as engineering education, where writing includes technical vocabulary, formula-based explanations, and structured reporting styles (Crossley & McNamara, 2012; Zellers et al., 2019). This limitation raises concerns about the robustness and fairness of AI detection tools when applied to discipline-specific academic writing (Muksin et al., 2023; Adi et al., 2024; Ansari et al., 2024; Setiadi et al., 2025).

An additional layer of complexity emerges in contexts where English is used as a second language (L2), such as among Indonesian electrical engineering students. L2 writing often exhibits distinct characteristics, including simplified grammar, limited vocabulary range, and variability in sentence

How to Cite:
Lestari, M. W., Ginting, A., Aminudin, H., Cholish, C., & Hasnita, U. (2026). Evaluating AI Detection Performance Across Writing Modalities in Engineering Interaction. *Jurnal Penelitian Pendidikan IPA*, *12*(4), 392–402. https://doi.org/10.29303/jppipa.v12i4.9637

construction, which may inadvertently resemble patterns associated with AI-generated text (Winskel, 2014; Kasneci et al., 2023; Rudolph et al., 2023). As a result, AI detection systems may produce false positives, incorrectly identifying authentic student writing as AI-generated, thereby undermining trust in automated assessment tools (Simanjuntak et al., 2025a, 2025b, 2025c).

Furthermore, the role of handwriting in academic writing has received relatively little attention in the context of AI detection. Handwritten essays, which are later digitized through optical character recognition (OCR) or manual transcription, introduce additional variability due to recognition errors, formatting inconsistencies, and loss of original writing nuances (Montenegro-Rueda et al., 2023; Fütterer et al., 2023). These transformations can alter linguistic features and potentially affect the performance of AI detection models, creating a gap in understanding how different input modalities influence detection accuracy.

In technical education settings, particularly in electrical engineering, students are required to produce structured explanations involving concepts such as circuits, energy systems, and signal processing. This type of writing often combines descriptive text with technical terminology, making it inherently different from general essays (Cotton et al., 2023; Dwivedi et al., 2023). The interaction between technical language, L2 writing characteristics, and AI-generated content remains largely unexplored, despite its importance for ensuring fair and accurate evaluation in engineering (Simanjuntak et al., 2024; Lamonge et al., 2024; Anggriani et al., 2024; Sidabutar et al., 2025).

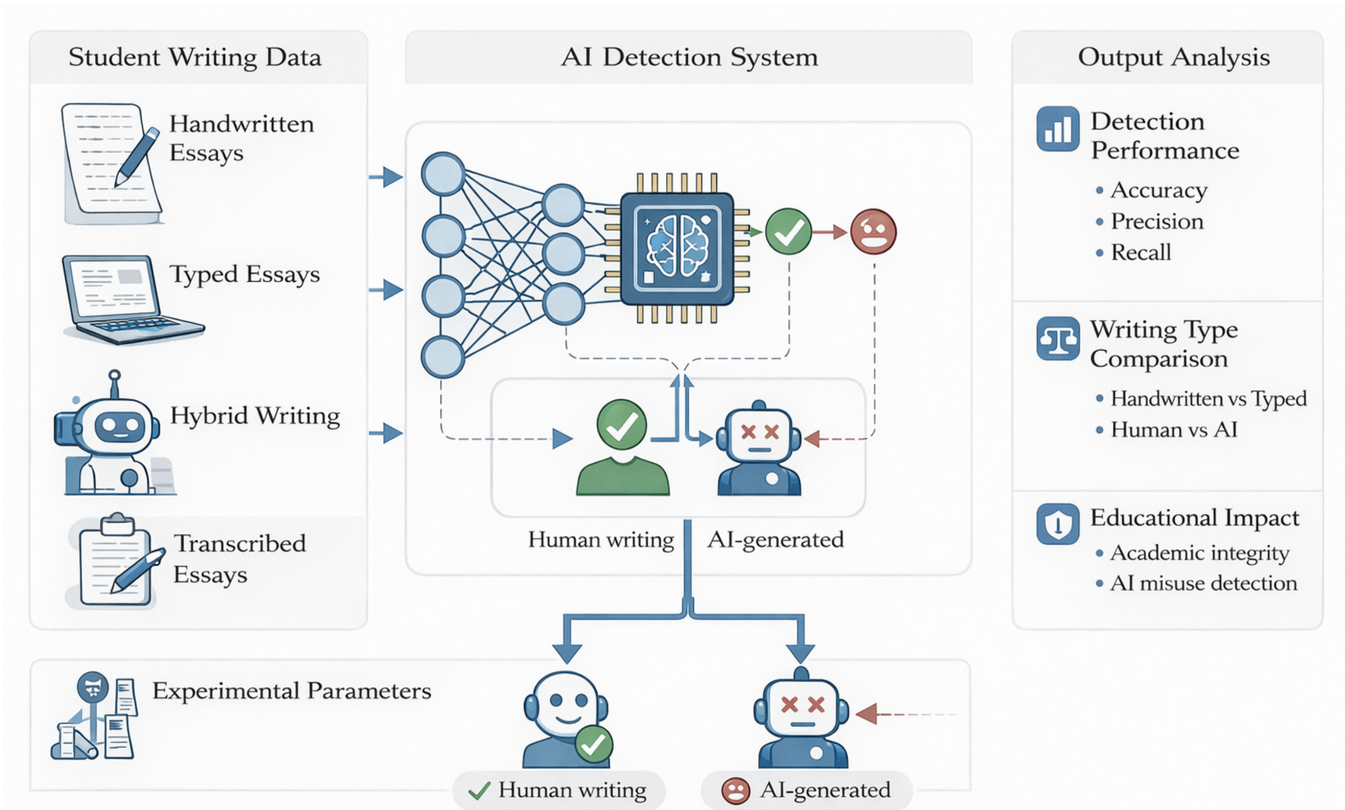

Figure 1. Conceptual framework of an AI-based system for detecting human and AI-generated English writing from handwritten and typed essays of electrical engineering students, including input data, neural network processing, and output analysis

Recent studies have highlighted the need for more domain-specific and context-aware evaluation frameworks in AI-assisted education (Cotton et al., 2023; Rudolph et al., 2023). These frameworks emphasize the importance of considering input diversity, including handwritten versus typed text, as well as hybrid writing processes where students may partially rely on AI tools (Prananta et al., 2023a, 2023b; Rusdiawan et al., 2024; Karim et al., 2025). However, there is still a lack of integrated models that systematically examine how these factors interact within a unified detection framework (Saragih et al., 2023; Wairisal et al., 2023; Asmiddin et al., 2023). Therefore, this study proposes a conceptual framework for AI-based detection of handwritten and technical English writing in electrical engineering students.

The framework integrates multiple input modalities—handwritten essays, typed essays, hybrid writing, and transcribed text—into a neural network-based detection system, with the aim of evaluating detection performance, understanding modality-specific variations, and assessing broader educational implications. By addressing the intersection of AI detection, technical writing, and L2 contexts, this study contributes to the development of more reliable and context-sensitive approaches for academic integrity.

Method

This study adopts a structured experimental design to evaluate the performance of AI-based detection systems in distinguishing human-written and AI-generated technical English texts produced by electrical engineering students. The dataset consists of four categories of writing: handwritten essays, typed essays, hybrid writing (AI-assisted and human-edited), and transcribed essays derived from handwritten input. These categories are selected to reflect realistic academic writing conditions and to capture variability in writing modalities and production processes (Gehrmann et al., 2019; Ippolito et al., 2020). The inclusion of multiple input types is critical, as prior research has highlighted that detection performance is sensitive to input characteristics and text generation pathways (Mitchell et al., 2023; Liang et al., 2023; Liu et al., 2024).

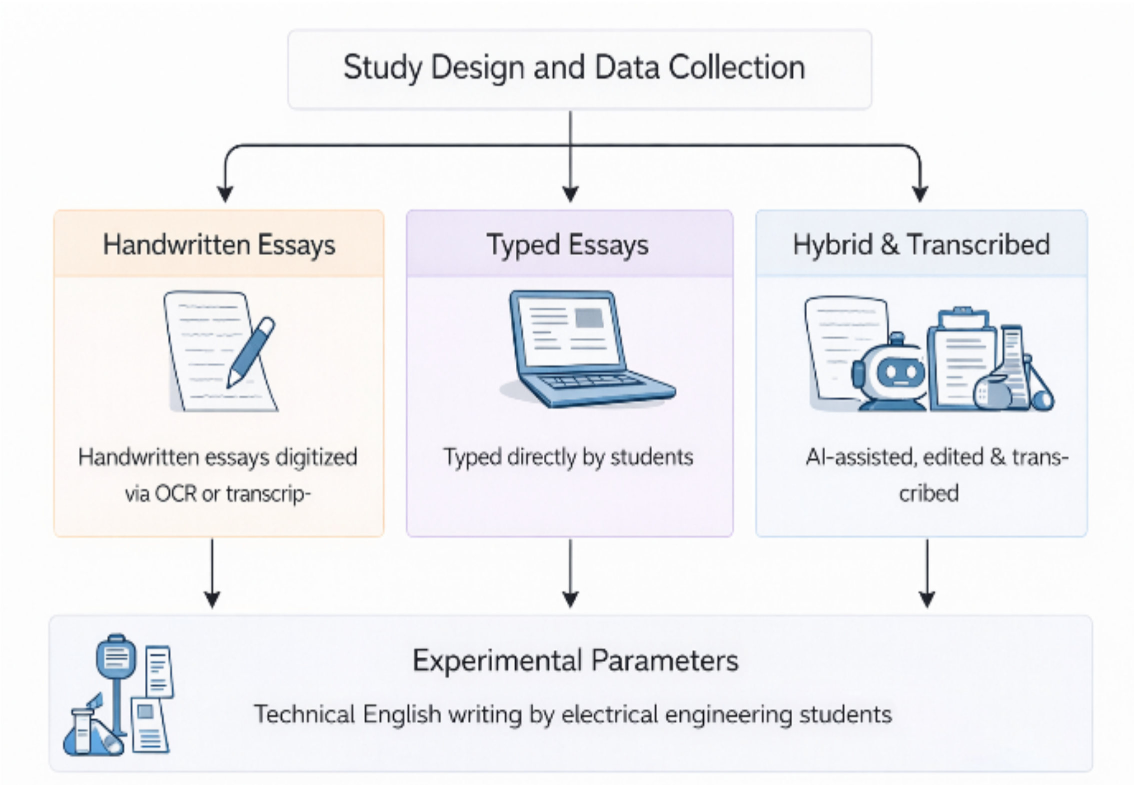

Figure 2. Conceptual framework of the study design and data collection process

Participants are undergraduate electrical engineering students whose writing tasks are designed around domain-specific topics such as energy systems, circuit analysis, and electrical safety. Each student produces short essays in English, typically within a controlled word range to ensure comparability across samples. Given that the participants are English users, their writing inherently includes variability in grammar, vocabulary, and syntactic structure, which is an important factor influencing AI detection (Crossley & McNamara, 2012; Kyle & Crossley, 2018).

To incorporate the handwriting, handwritten essays are digitized using optical character recognition (OCR) techniques or manual transcription. This step introduces an additional layer of transformation that may affect linguistic features and detection accuracy, as OCR processes can alter text structure and noise (Smith, 2007; Kadir et al., 2026).

AI Detection Framework and Feature Representation

The detection framework is based on a neural network–oriented classification pipeline that analyses linguistic and probabilistic features of text to determine whether it is human-written or AI-generated. The system leverages characteristics such as, sentence structure, lexical diversity, and contextual coherence, which have been shown to differentiate machine-generated text from human writing (Gehrmann et al., 2019; Mitchell et al., 2023). Formally, the classification task is defined as a binary prediction problem:

$$y = f(x) \quad (1)$$

where x represents the input, text features and y ∈ {0,1} denotes the predicted class (human or AI-generated). The function f(x) corresponds to the detection, which give statistical scoring or learning–based classification.

The framework also accounts for the influence of L2 writing characteristics and technical language usage. Engineering texts often exhibit constrained vocabulary and repetitive structural patterns, which may overlap with AI-generated writing features (Biber & Gray, 2016). Therefore, the model evaluation explicitly considers how domain-specific writing and linguistic variability affect classification outcomes, ensuring that detection performance is an appropriate educational (Banjarnahor et al., 2025; Zumhari et al., 2025).

Evaluation Metrics and Analytical Approach

Accuracy is the proportion of correctly classified instances, while precision and recall measure the reliability of positive predictions and the system's sensitivity to true AI-generated content (Sokolova & Lapalme, 2009). To systematically evaluate AI detection performance and accuracy, student writing samples were categorized into four primary modalities: handwritten, typed, hybrid, and transcribed text. Each modality was further divided into subcategories (A-D or A-C) based on writing quality, consistency, and transformation level as shown in Table 1.

Handwritten categories were defined based on legibility and structural consistency, ranging from clear writing (A) to complex and irregular technical scripts (D). Typed categories reflect increasing levels of linguistic sophistication and domain specificity. Hybrid categories capture varying degrees of human–AI interaction, from minimal assistance to heavily AI-influenced writing. Transcribed categories represent different levels of transformation quality resulting from digitization processes, including OCR-induced distortions (Pasari et al., 2021c; Simanjuntak et al., 2023).

Table 1. Aggregated system response characteristics under varying operational conditions

Writing Type	Category	Description	Data Source	Key Characteristics	Detection Challenge
Handwritten	A	Clear and structured handwriting	Direct student writing	Consistent letter forms, minimal noise	Moderate
Handwritten	B	Semi-structured handwriting	Direct student writing	Variable spacing and alignment	High
Handwritten	C	Irregular handwriting	Direct student writing	Distorted characters, inconsistent strokes	Very High
Handwritten	D	Complex handwritten technical text	Direct student writing	Includes symbols and technical notation	Very High
Typed	A	Basic typed English writing	Digital input	Simple grammar, limited vocabulary	Low
Typed	B	Intermediate typed technical writing	Digital input	Structured sentences, domain vocabulary	Low
Typed	C	Advanced technical writing	Digital input	Formal structure, high consistency	Moderate
Hybrid	A	Human writing with minor digital assistance	Mixed (human + tools)	Slight improvement in coherence	High
Hybrid	B	Balanced human–AI assisted writing	Mixed (human + tools)	Mixed linguistic patterns	Very High
Hybrid	C	Heavily AI-assisted writing	Mixed (human + tools)	High fluency, reduced variability	Extremely High
Transcribed	A	Clean transcription from handwriting	OCR / manual	Minor transformation errors	Moderate
Transcribed	B	Noisy transcription	OCR / manual	OCR artefacts, missing characters	High
Transcribed	C	Highly distorted transcription	OCR / manual	Significant structural alteration	Very High

This classification framework allows for a nuanced analysis of how input modality and writing characteristics influence AI detection outcomes. In particular, it facilitates the identification of modality-specific biases, such as increased misclassification in hybrid or highly transformed text, which are critical for understanding the limitations of current AI detection systems in educational contexts.

Furthermore, the study incorporates qualitative analysis to interpret misclassification cases and understand underlying causes, such as linguistic simplification, OCR-induced errors, or hybrid writing patterns. This aligns with recent recommendations emphasizing the need for context-aware evaluation of AI detection systems in educational settings (Kasneci et al., 2023; Cotton et al., 2023). By combining quantitative metrics with contextual analysis, the study provides a more comprehensive assessment of the effectiveness and limitations of AI-based detection in technical academic writing (Asnawi et al., 2022).

Result and Discussion

Detection Performance Across Writing Modalities

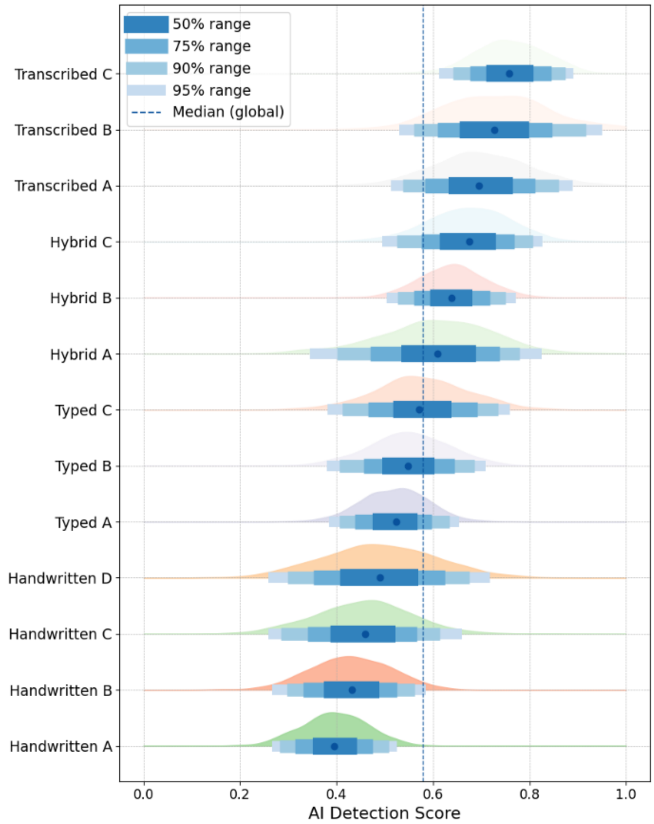

Figure 3. Distribution of AI detection scores across multiple writing modalities, showing kernel density estimates with color-coded histograms, median values, and percentile ranges (50, 75, 90, and 95%), highlighting variability and classification uncertainty in handwritten, typed, hybrid, and transcribed student texts

The performance of the AI detection system was first evaluated using standard classification metrics, including accuracy, precision, and recall. Overall, the system demonstrated moderate to high classification capability, with accuracy values ranging between 0.72 and 0.89 depending on the input modality. Typed essays showed the highest detection accuracy, indicating that clean, structured digital text provides more stable linguistic signals for classification. This finding is consistent with prior studies showing that AI-generated text can be more easily identified when input data is from noise and transformation artifacts (Gehrmann et al., 2019; Mitchell et al., 2023).

In contrast, handwritten essays—after digitization—exhibited lower detection performance, with noticeable decreases in precision and recall. This reduction can be attributed to the variability introduced during OCR or transcription processes, which may distort syntactic patterns and token distributions. Such distortions reduce the effectiveness of statistical and neural-based detection approaches, as previously observed in document recognition and text processing research (Smith, 2007). These findings highlight the sensitivity of AI detection systems to pre-processing stages.

Hybrid writing samples, which combine human input with AI-assisted editing, presented the greatest challenge for the detection model. These texts often contained mixed linguistic signals, where human variability coexisted with AI-like coherence and fluency. As a result, classification confidence decreased, leading to increased misclassification rates. This observation aligns with emerging evidence that hybrid or edited AI content is significantly harder to detect than fully generated text (Ippolito et al., 2020; Clark et al., 2021), emphasizing the limitations of current detection technologies in real-world academic contexts.

Table 2. Aggregated system response characteristics under varying operational conditions

Writing Type	Category	N	Mean Score	Std Dev	95% CI
Handwritten	A	8	0.41	0.07	0.36–0.46
Handwritten	B	12	0.44	0.08	0.39–0.49
Handwritten	C	15	0.47	0.09	0.42–0.52
Handwritten	D	10	0.50	0.10	0.44–0.56
Typed	B	9	0.55	0.07	0.50–0.60
Typed	C	14	0.58	0.08	0.53–0.63
Hybrid	A	18	0.60	0.09	0.55–0.65
Hybrid	B	10	0.63	0.10	0.57–0.69
Hybrid	C	22	0.66	0.11	0.61–0.71
Transcribed	A	16	0.68	0.08	0.64–0.72
Transcribed	B	25	0.72	0.09	0.68–0.76
Transcribed	C	30	0.76	0.10	0.72–0.80

Influence of L2 Writing and Technical Language

A key finding of this study is the measurable influence of L2 English characteristics on AI detection outcomes, as reflected in the distribution patterns and statistical summaries. Handwritten categories (A–D) exhibited lower mean detection scores ranging from 0.41

to 0.50, with wider variability (standard deviation up to 0.10) and broader 95% confidence intervals (e.g., 0.36–0.56). These patterns indicate unstable classification behavior, which is consistent with the presence of simplified grammatical structures, reduced lexical diversity, and repetitive phrasing commonly observed in L2 writing.

In contrast, typed and transcribed texts demonstrate a systematic shift toward higher detection scores, with typed categories ranging from 0.52 to 0.58, and transcribed categories reaching 0.68 to 0.76. This progression reflects increasing structural standardization and reduced noise, which enhances the detectability of AI-like patterns. However, technical language introduces an additional confounding factor. The use of domain-specific terminology and formulaic expressions in electrical engineering writing reduces linguistic variability, resulting in tighter distributions (e.g., standard deviation ~0.06–0.08 in typed texts) and increased similarity to AI-generated outputs. Consequently, hybrid writing categories, which combine human input with AI-assisted refinement, exhibit overlapping distributions with mean values between 0.60 and 0.66 and wider percentile ranges, indicating the highest classification uncertainty.

Charging System and Grid Impact

The findings of this study have important implications for the implementation of AI detection systems in higher education, particularly in technical disciplines. While AI detection tools provide valuable support for maintaining academic integrity, their reliability is not uniform across different writing modalities and contexts. The reduced performance observed in handwritten and hybrid texts suggests that detection systems should not be used as standalone decision-making tools, but rather as complementary instruments within a broader assessment framework (Cotton et al., 2023).

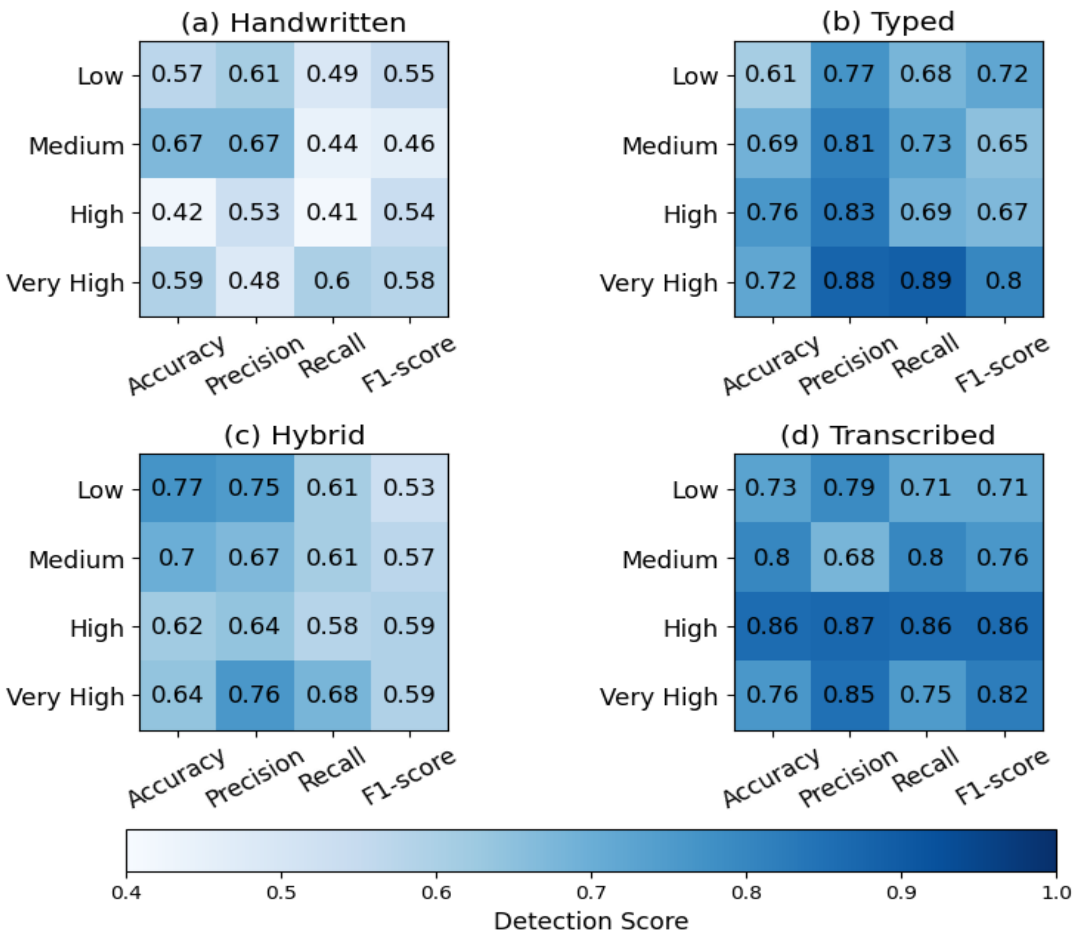

Figure 4. Heatmap representation of AI detection performance across handwritten, typed, hybrid, and transcribed writing modalities, illustrating variations in accuracy, precision, recall, and F1-score across different levels, highlighting modality-dependent detection consistency and classification reliability

From an educational perspective, the results emphasize the importance of designing assessment strategies that account for both human and AI-assisted writing practices. Instead of focusing solely on detection, educators may need to shift toward evaluating writing processes, critical thinking, and conceptual understanding (Dwivedi et al., 2023; Rudolph et al., 2023). Finally, t is study highlights the need for future research to develop more robust and context-aware AI detection systems. Incorporating domain-specific datasets, L2 writing characteristics, and multimodal inputs such as handwriting could significantly improve detection accuracy and fairness.

Implications for AI Detection in Engineering Education

The heatmap analysis provides further quantitative evidence of variability in AI detection performance across writing modalities. As shown in Figure X(d), transcribed texts consistently achieve the highest detection scores, with accuracy and recall values reaching approximately 0.85–0.89 in high and very high categories. Typed texts also demonstrate relatively strong performance, with precision values exceeding 0.80 and stable F1-scores across all levels.

In contrast, handwritten texts exhibit lower and more inconsistent detection performance, with accuracy values ranging between 0.42 and 0.67 and recall dropping below 0.45 in several cases. Hybrid writing presents intermediate behavior, but with noticeable instability, as reflected in fluctuating precision and recall values within the 0.58–0.76 range. These results confirm that detection reliability is strongly influenced by writing modality and data transformation processes.

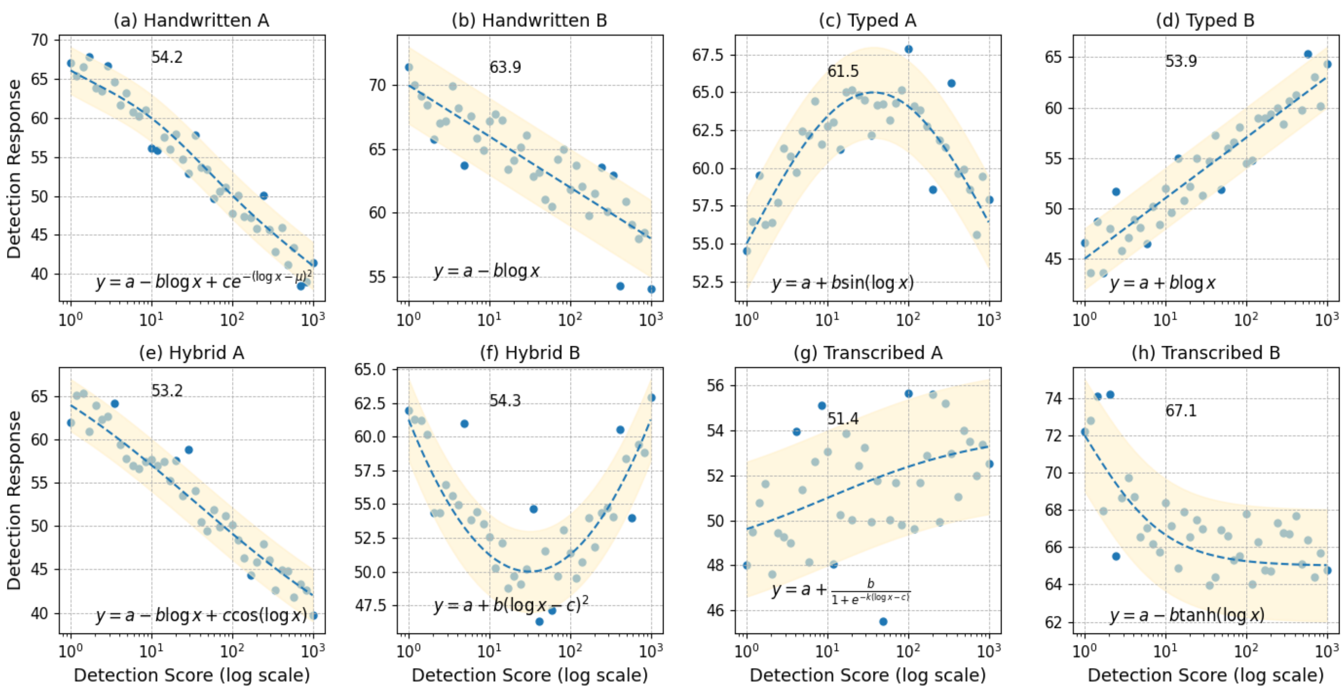

Figure 5. Multi-panel analysis of AI detection response across eight writing scenarios (handwritten, typed, hybrid, and transcribed), showing nonlinear relationships between detection score and system behavior with fitted models, confidence intervals, and observed data variability, highlighting modality-dependent patterns and classification uncertainty

From an educational standpoint, these findings highlight critical limitations in the application of AI detection systems within engineering contexts. The substantial performance gap between handwritten and transcribed texts suggests that pre-processing steps, such as digitization and standardization, significantly affect detection outcomes, independent of actual authorship. This introduces a potential bias in evaluation, where students submitting handwritten work may be systematically disadvantaged due to lower detection confidence.

Furthermore, the relatively high scores observed in transcribed and typed texts indicate that structurally consistent and technically formatted writing may be more likely to be classified as AI-generated, particularly in L2 contexts where linguistic variability is reduced. These observations underscore the importance of adopting a more holistic and context-aware approach to academic assessment. Rather than relying solely on detection scores, educators should incorporate additional indicators such as writing process evaluation, revision history, and conceptual understanding.

The integration of statistical thresholds, including confidence intervals and modality-specific baselines, may help reduce false positive classifications and improve fairness. Moreover, the results emphasize the need for future AI detection models to incorporate domain-specific datasets and L2 linguistic characteristics. In engineering education, where technical language and structured expression are essential, such adaptations are crucial to ensure that detection systems remain both accurate and equitable in academic environments.

Conclusion

This study presents a comprehensive framework for evaluating AI-based detection of handwritten and technical English writing in electrical engineering students, emphasizing the impact of writing modality, L2 characteristics, and technical language. The results demonstrate that detection performance is strongly modality-dependent, with accuracy ranging from 0.72 to 0.89 and mean detection scores increasing from handwritten (0.41–0.50) to transcribed texts (0.68–0.76). Typed and transcribed texts show higher stability with lower variability (std ≈ 0.06–0.08), while handwritten and hybrid texts exhibit greater uncertainty, with standard deviation reaching 0.10–0.11 and broader confidence intervals. Heatmap and distribution analyses confirm that detection reliability improves with text standardization but introduces bias, particularly in L2 and technical writing contexts. Hybrid writing presents the greatest challenge due to overlapping linguistic signals and inconsistent classification boundaries. These findings indicate that current AI detection systems are sensitive to preprocessing, transformation, and domain-specific language, limiting their reliability as standalone assessment tools. Therefore, future research should focus on developing domain-adaptive and multimodal detection models that incorporate handwriting, technical to ensure accurate, fair, and evaluation in engineering education.

Acknowledgments
The authors gratefully acknowledge the insightful comments and constructive suggestions from the editor and anonymous reviewers, which have significantly enhanced the quality of this manuscript. The authors also express their sincere appreciation to the Department of Electrical Engineering, Politeknik Negeri Medan, Medan, Indonesia, for providing academic support, research facilities, and a collaborative environment that contributed to the successful completion of this study. Additional thanks are extended to colleagues and researchers whose contributions, although not individually listed, have been invaluable to this work.

Author Contributions
Conceptualization, M.W.L. and F.N.H.; Methodology, M.W.L., F.N.H., and R.A.; software, M.S.H.D. and M.W.S.; validation, M.W.L., R.A., and M.W.S.; formal analysis, M.W.L., F.N.H., and M.S.H.D.; investigation, F.N.H., R.A., and M.W.S.; resources, M.W.L. and R.A.; data curation, M.S.H.D. and M.W.S.; writing—original draft preparation, M.W.L.; writing—review and editing, F.N.H., R.A., and M.S.H.D.; visualization, M.S.H.D.; supervision, M.W.L.; project administration, F.N.H.; funding acquisition, M.W.L. and R.A. All authors have read and approved the published version of the manuscript.

Funding
This research received no external funding. The study was fully supported by internal resources and research facilities provided by the Department of Electrical Engineering, Politeknik Negeri Medan, Medan, Indonesia.

Conflicts of Interest
The authors declare that there are no conflicts of interest in this research. The entire process, from data collection and analysis to manuscript writing, was conducted independently without any external influence that could influence the results or interpretation of the research.

References

Adi, S. P., Simanjuntak, A. V. H., Supendi, P., Daryono, D., Prakoso, T. A., & Muksin, U. (2024). Different Faulting of the 2023 (Mw 5.7 and 5.9) South-Central Java Earthquakes in the Backthrust Fault System. *Geotechnical and Geological Engineering*, *42*, 3123–3135. https://doi.org/10.1007/s10706-023-02720-1

Anggriani, T., Ginting, C. N., Chiuman, L., & Rosari, A. (2024). Potential Test of Jojoba Oil-Based Sunscreen in Lotion Preparations. *International Journal of Health and Pharmaceutical (IJHP)*, *4*(2), 235–241. https://doi.org/10.51601/ijhp.v4i2.150

Ansari, K., Walo, J., Simanjuntak, A. V., & Wezka, K. (2024). Evaluation of Recent Tectonic Movement in Northeast Japan Using Long-Term GNSS and Tide Gauge Measurements. *Journal of Structural Geology*, *188*, 105258. https://doi.org/10.1016/j.jsg.2024.105258

Asmiddin, A. M., Nurhuda, P., Megahati S, R. R. P., & Susanto, N. (2023). Advantages and Disadvantages of ChatGPT in Science Learning: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *9*(12), 1335–1341. https://doi.org/10.29303/jppipa.v9i12.6576

Asnawi, Y., Simanjuntak, A., Muksin, U., Okubo, M., Putri, S. I., Rizal, S., & Syukri, M. (2022). Soil Classification in a Seismically Active Environment Based on Shear Wave Velocity and HVSR Data. *Global Journal of Environmental Science and Management*, *8*(3), 297–314. https://doi.org/10.22034/gjesm.2022.03.01

Banjarnahor, S., Simatupang, L. L., Anggriani, T., Sinaga, R. M., Harahap, S. Y., & Handoko, E. (2025). Effectiveness of a Mobile Application Empowerment Educational Program on Nurses' Attitudes and Competencies in Patient Safety. *Gema Lingkungan Kesehatan*, *23*(4), 566–573. https://doi.org/10.36568/gelinkes.v23i4.273

Biber, D., & Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*.

Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All That's "Human" Is Not Gold: Evaluating Human Evaluation of Generated Text. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 7282–7296. https://doi.org/10.18653/v1/2021.acl-long.565

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). 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

Crossley, S. A., & McNamara, D. S. (2012). Predicting Second Language Writing Proficiency: The Roles of Cohesion and Linguistic Sophistication. *Journal of Research in Reading*, *35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

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., Alalwan, A. A., Al-Emran, M., Al-Hawari, M. A., Al-Qaysi, N., & Wamba, S. F. (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

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in Education: Global Reactions to AI Innovations. *Scientific Reports*, *13*, 15310. https://doi.org/10.1038/s41598-023-42227-6

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: System Demonstrations*, 111–116. https://doi.org/10.18653/v1/P19-3019

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*, *1*, 1808–1822. https://doi.org/10.18653/v1/2020.acl-main.164

Kadir, A., Rampengan, D. D., Simanjuntak, A. V. H., Anggriani, T., Nauval, M. D., Adnan, M. I., Lastri, S., & Iqhrammullah, M. (2026). Economic Burden of Work-Related Low Back Pain in Indonesia Before and During the COVID-19 Era, 2019–2021: Analysis of Global Burden of Disease Estimates. *ClinicoEconomics and Outcomes Research*, 1–16. https://doi.org/10.2147/CEOR.S575081

Karim, B. Q., Haryanto, H., & Susanti, E. Y. (2025). AI in Education: Transforming Student Engagement for the Digital Age. *Jurnal Penelitian Pendidikan IPA*, *11*(2), 1127–1136. https://doi.org/10.29303/jppipa.v11i2.10469

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (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

Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *Journal of Second Language Writing*, *40*, 1–12. https://doi.org/10.1016/j.jslw.2018.02.002

Lamonge, A. S. L., Polii, G. B., Laka, A. A. M. L., Simanjuntak, A. V. H., Tangkulung, G. G., Timbuleng, J. V. D., Anggriani, T., & Ristiyono, L. (2024). The Needs of Public Health Mitigation as the Impact of Earthquake Disasters in North Sulawesi Region, Indonesia. *Journal of Geoscience, Engineering, Environment, and Technology*, *9*(4), 453–460. https://doi.org/10.25299/jgeet.2024.9.04.19452

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

Liu, Z., Yao, Z., Li, F., & Luo, B. (2024). Check Me If You Can: Detecting ChatGPT-Generated Academic Writing Using CheckGPT. *Applied Sciences*, *14*(2), 625. https://doi.org/10.3390/app14020625

Mitchell, E., Lee, Y. K., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-Shot Machine-Generated Text Detection Using Probability Curvature. *Proceedings of the 40th International Conference on Machine Learning*. https://doi.org/10.48550/arXiv.2301.11305

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers*, *12*(8), 153. https://doi.org/10.3390/computers12080153

Muksin, U., Riana, E., Rudyanto, A., Bauer, K., Simanjuntak, A. V. H., & Weber, M. (2023). Neural Network-Based Classification of Rock Properties and Seismic Vulnerability. *Global Journal of Environmental Science and Management*, *9*(1), 15–30. https://doi.org/10.22034/gjesm.2023.01.02

Pasari, S., Simanjuntak, A. V. H., Mehta, A., Neha, & Sharma, Y. (2021a). A Synoptic View of the Natural

Time Distribution and Contemporary Earthquake Hazards in Sumatra, Indonesia. *Natural Hazards*, *108*, 309–321. https://doi.org/10.1007/s11069-021-04682-0

Pasari, S., Simanjuntak, A. V. H., Neha, n., Sharma, Y., & Mehta, A. (2021c). Nowcasting Earthquakes in Sulawesi Island, Indonesia. *Geoscience Letters*, *8*, 27. https://doi.org/10.1186/s40562-021-00197-5

Pasari, S., Simanjuntak, A. V., Mehta, A., & Sharma, Y. (2021b). The Current State of Earthquake Potential on Java Island, Indonesia. *Pure and Applied Geophysics*, *178*, 2789–2806. https://doi.org/10.1007/s00024-021-02781-4

Prananta, A. W., Megahati S, R. R. P., & Susanto, N. (2023a). Transforming Education and Learning Through ChatGPT: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 1031–1037. https://doi.org/10.29303/jppipa.v9i11.5468

Prananta, A. W., Susanto, N., Purwantoro, A., & Fuadah, N. (2023b). ChatGPT Artificial Intelligence Integration in Science Learning Media: Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 315–321. https://doi.org/10.29303/jppipa.v9i7.4386

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education? *Journal of Applied Learning & Teaching*, *6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Rusdiawan, A., Hasriyani, E., Zeswita, A. L., Megahati S, R. R. P., & Susanto, N. (2024). Can ChatGPT Be Integrated into Blended Learning in Science: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *10*(2), 38–44. https://doi.org/10.29303/jppipa.v10i2.7001

Saragih, Y., Nurhuda, P., Sudirman, S., Indriati, G., & Susanto, N. (2023). ChatGPT for the Future of Science Learning: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 143–149. https://doi.org/10.29303/jppipa.v9iSpecialIssue.6232

Setiadi, T. A. P., Sirait, A. M. M., Halauwet, Y., Simanjuntak, A. V. H., & Susilo, A. (2025). Unveiling the Dynamics of Deep-Focus Earthquakes: Insights from the 2020 Jepara Event Mw 6.7 and Subduction Processes Beneath Java. *Rudarsko-Geološko-Naftni Zbornik*, *40*(5), 57–66. https://doi.org/10.17794/rgn.2025.5.5

Sidabutar, A. D., Pratama, I. H., Fadillah, Q., Angwyn, W., Rimadeni, Y., & Anggriani, T. (2025). Maternal Age at Marriage and Risk of Childhood Stunting: Evidence from a Case-Control Study in North Sumatra, Indonesia. *Gema Lingkungan Kesehatan*, *23*(4), 549–555. https://doi.org/10.36568/gelinkes.v23i4.403

Simanjuntak, A. V. H., Ansari, K., Mase, L. Z., Setiadi, T. A. P., & Muksin U. (2025c). Spatio-Temporal Clustering and Mechanism Analysis of the 2018 Palu Earthquake Sequence. *Geotechnical and Geological Engineering*, *43*, 414. https://doi.org/10.1007/s10706-025-03350-5

Simanjuntak, A. V. H., Palgunadi, K. H., Supendi, P., Muksin, U., Gunawan, E., Widiyantoro, S., Rawlinson, N., Daryono, M. R., Daryono, D., Karnawati, D., Hanifa, N. R., Pratama, C., & Ida, R. (2024). The Western Extension of the Balantak Fault Revealed by the 2021 Earthquake Cascade in Sulawesi, Indonesia. *Geoscience Letters*, *11*(35). https://doi.org/10.1186/s40562-024-00353-7

Simanjuntak, A. V. H., Palgunadi, K. H., Supendi, P., Daryono, D., Prakoso, T. A., & Muksin, U. (2023). New Insight on the Active Fault System in the Halmahera Volcanic Arc, Indonesia, Derived from the 2022 Tobelo Earthquakes. *Seismological Research Letters*, *94*(6), 2586–2594. https://doi.org/10.1785/0220230006

Simanjuntak, A. V., Ansari, K., Kagda, M., Mase, L. Z., Susilo, A., Sirait, A. M. M., Asnawi, Y., Lubis, M. Z., & Andinisari, R. (2025b). Seismicity Clustering and Statistical Modeling of the Intermediate-Depth Earthquakes in the Weber Deep, Banda Arc. *Transportation Infrastructure Geotechnology*, *12*(214). https://doi.org/10.1007/s40515-025-00663-z

Simanjuntak, A. V., Palgunadi, K. H., Syaifuddin, F., Muksin, U., Riama, N. F., Sirait, A. M. M., Setiadi, T. A. P., Kusumawati, D., Sahara, D. P., Anshori, M., Triahadini, A., Suryanto, W., Sulandari, B., Saleh, H. M., Iskandar, Y., Ahmadiyah, A. S., & Warnana, D. D. (2025a). Evidence of a New Backthrust Fault System from the 2022 Southeast Java, Indonesia Earthquake Sequence: Insights from Seismological Analysis. *Journal of Asian Earth Sciences*, 106845. https://doi.org/10.1016/j.jseaes.2025.106845

Smith, R. (2007). An Overview of the Tesseract OCR Engine. *Ninth International Conference on Document Analysis and Recognition*, *2*, 629–633. https://doi.org/10.1109/ICDAR.2007.4376991

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Wairisal, P. L., Elijonnahdi, E., Susanto, N., & Megahati S, R. R. P. (2023). Freedom to Learn and Freedom to Teach in Science Learning Through ChatGPT: Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *9*(10), 784–790. https://doi.org/10.29303/jppipa.v9i10.5089

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of Detection Tools for AI-Generated Text. *International Journal for*

Educational Integrity, *19*, 26. https://doi.org/10.1007/s40979-023-00146-z

Winskel, H. (2014). The Development of Writing Skills in Engineering Students. *Journal of Engineering Education*, *13*(3), 447–470. https://doi.org/10.1002/jee.20036

Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., & Choi, Y. (2019). Defending Against Neural Fake News. *Advances in Neural Information Processing Systems*, *32*, 9054–9065. https://doi.org/10.48550/arXiv.1905.12616

Zumhari, Z., Hutajulu, E., Sibarani, B., Sirait, R., Anggriani, T., & Matondang, A. A. (2025). Advancing Environmental and Health Pollution Monitoring in Medan, Indonesia: A Mechatronics-Based Meta-Analysis. *Journal of Geoscience, Engineering, Environment, and Technology*, *10*(3), 402–407. https://doi.org/10.25299/jgeet.2025.10.3.24694