Jurnal TRANSFORMATIKA
Vol.23, No.2, Januari 2026, pp. 247 - 257
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Artificial Intelligence-Based Automatic Text Detection System Using Multi-Layer Pattern Recognition

Kartika Imam Santoso[1*], Edi Widodo[2], Theresia Widji Astuti[3]
[1]Program Studi Ilmu Komputer, Universitas An Nuur
Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: kartikaimams@gmail.com
[2]Prodi Sistem Informasi, Universitas Semarang
Jl. Sukarno Hatta Semarang, e-mail: ediwidodo@usm.ac.id
[3] Prodi manajemen informatika, Politeknik Negeri Sambas
Jl.Raya Sejangkung Kawasan Pendidikan Tinggi Sambas – Kalimantan Barat 79462, e-mail: theresiawidji@gmail.com

ARTICLE INFO

History of the article :
Received 4 Desember 2025
Received in revised form 22 Januari 2026
Accepted 22 Januari 2026
Available online 24 Januarai 2026

Keywords:

12 Algoritm; AI detection; pattern recognition; text classification

*** Correspondece:**

Telepon:
+62 85328184305

E-mail:
kartikaimams@gmail.com

ABSTRACT

The proliferation of generative AI models poses significant challenges to academic integrity and the authenticity of content. This study develops a multi-layer pattern recognition system to detect AI-generated text and classify the source AI model. The system analyzes 12 algoritm : linguistic, structural, and statistical parameters across seven analytical layers using uploaded documents in PDF, DOCX, and TXT formats. A weighted scoring mechanism generates overall AI probability scores (0-100%) and individual probabilities for 10 AI models. Testing with 500 academic documents achieved 87.3% accuracy in AI detection and 82.0% accuracy in model classification. Entropy analysis, sentence structure diversity, and emotional markers proved to be the most discriminative. The system demonstrates that transparent, rule-based pattern recognition offers a viable alternative to black-box neural approaches, with practical applications in academic integrity verification, content authentication, and digital forensics.

1. INTRODUCTION

Perkembangan Large Language Model (LLM) seperti GPT-4, Claude, Gemini, dan DeepSeek telah mengubah cara orang menghasilkan konten digital [1] [2]. Meskipun meningkatkan produktivitas, teknologi ini menimbulkan tantangan serius terkait autentisitas konten dan integritas akademis, dengan 10-15% pengajuan akademis mengandung konten AI [3] [4], dengan risiko serupa dalam dokumen hukum, rekam medis, dan komunikasi profesional [5].

Alat deteksi AI existing seperti GPTZero, ZeroGPT, Turnitin, dan Copyleaks menunjukkan keterbatasan signifikan: tingkat positif palsu tinggi (0-100% variance), algoritma black-box yang tidak transparan, dan kegagalan mengidentifikasi model AI spesifik yang digunakan [6]. Sistem ini

juga kesulitan menangani konten campuran manusia-AI, teks pendek (<200 kata), dan model yang disesuaikan [7].

Pendekatan ekstraksi fitur linguistik telah dieksplorasi dengan hasil menjanjikan. Analisis entropy dan variasi bahasa dapat membedakan teks AI dengan akurasi 78-82% [8] [9]. sementara pendekatan pembelajaran mesin supervised menggunakan neural networks (RNN, LSTM, BERT) mencapai akurasi lebih tinggi namun memerlukan dataset berlabel besar, sumber daya komputasi signifikan, dan kurang interpretable [10]. Pendekatan pengenalan pola multi-modal yang menggabungkan fitur linguistik, struktural, dan behavioral masih kurang dikaji, khususnya untuk membedakan multiple AI models [11] [12].

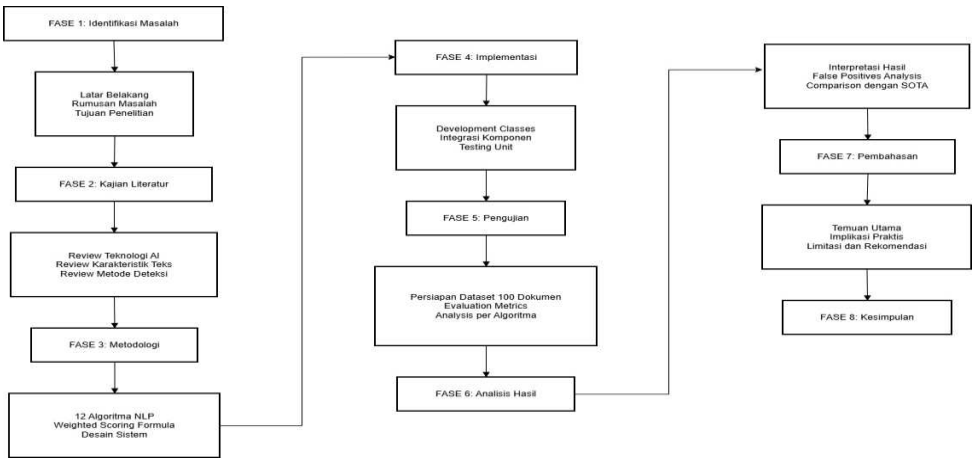
Tujuan dari penelitian adalah untuk mengembangkan sistem deteksi otomatis yang mampu mengidentifikasi teks buatan AI dengan akurasi ≥85%, dan bisa mengklasifikasikan model AI secara spesifik yang digunakan untuk membuat teks yang dihasilkan di antara 10 model populer.

Gap penelitian meliputi: (1) kurangnya sistem yang dapat mengidentifikasi model AI spesifik, (2) minimnya pendekatan yang transparan dan interpretable untuk decision-making, (3) keterbatasan analisis multi-dimensional yang komprehensif, dan (4) tidak adanya solusi yang dapat diimplementasikan secara lokal tanpa ketergantungan pada layanan komersial. Penelitian ini bertujuan mengembangkan sistem deteksi dengan akurasi ≥85% yang mampu mengklasifikasikan 10 model AI populer menggunakan pendekatan multi-layer pattern recognition.

Kontribusi penelitian: (1) sistem pengenalan pola berbasis aturan yang transparan untuk deteksi dan klasifikasi model spesifik, (2) analisis 12 parameter berbeda di seluruh dimensi linguistik, struktural, dan behavioral, (3) AI fingerprinting berdasarkan karakteristik model yang terdokumentasi, dan (4) web-based system yang accessible bagi pendidik dan institusi tanpa memerlukan keahlian machine learning.

2. RESEARCH METHODS

Secara umum alur penelitian dapat dilihat pada Gambar 1, dengan tahapan pengembangan sistem deteksi AI menggunakan arsitektur multi-layer dengan 12 algoritma terintegrasi dalam weighted scoring framework. Sistem diimplementasikan sebagai web-based application menggunakan HTML5, CSS3, dan vanilla JavaScript untuk kompatibilitas maksimum dan ketergantungan minimal.



Gambar 1. Alur Penelitian

Sistem deteksi AI yang dikembangkan menggunakan arsitektur multi-layer dengan 12 algoritma integrated dalam weighted scoring framework. Sistem diimplementasikan sebagai web-based

application menggunakan HTML5, CSS3, dan vanilla JavaScript untuk maximum compatibility dan minimal dependency.

A. Algoritma Deteksi

1. Entropy Analysis
   Entropy mengukur keacakan distribusi kata dalam text. Formula yang digunakan adalah:

$$H = -\sum_{i=1}^{n} p_i log_2 p_i \quad (1)$$

   di mana p_i adalah probability setiap kata. Teks AI memiliki entropy lebih rendah karena distribusi kata lebih terstruktur.
2. Sentence Structure Diversity
   Mengukur coefficient of variation dari panjang kalimat:

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (2)$$

   di mana σ adalah standard deviation dan μ adalah mean panjang kalimat.
3. N-Gram Frequency Analysis
   Mendeteksi pengulangan pola 2-3 kata berturut-turut dengan threshold frequency > 3.
4. Lexical Diversity (TTR)
   Type-Token Ratio dihitung sebagai:

$$TTR = \frac{Unique\ Words}{Total\ Words} \quad (3)$$

5. Passive Voice Detection
   Mengidentifikasi passive voice constructions menggunakan regex patterns.
6. Punctuation Pattern Analysis
   Menganalisis ratio penggunaan tanda baca (comma, exclamation, question mark, semicolon).
7. AI Phrases Detection
   Mengidentifikasi 20 common phrases yang frequent di teks AI seperti "in conclusion", "furthermore", "studies have shown", dll.
8. Transition Words Frequency
   Mengidentifikasi 20 common phrases yang frequent di teks AI seperti "in conclusion", "furthermore", "studies have shown", dll.
9. Corporate Phrases Detection
   Mengidentifikasi corporate/formal phrases seperti "key takeaways", "leverage", "synergies", "optimize".
10. Emotional Markers
    Mendeteksi absence emotional markers seperti "I feel", "I believe", "embarrassed", "awkward".
11. Contractions Analysis
    Menganalisis frequency contractions (don't, can't, won't, I'm, it's, dll).
12. Semantic Consistency
    Mengukur kesamaan semantic antar kalimat berturut-turut menggunakan word overlap analysis.

DOI : https://doi.org/10.26623/transformatika.v23i2.13256

B. Weighted Combination Scoring

Final AI score dihitung menggunakan weighted combination dari 12 algorithms:

$$AI\ Score \sum_{i=1}^{12} w_i \times s_i \quad (4)$$

di mana w_i adalah bobot dan s_i adalah normalized score (0-100) dari algoritma ke-i.

Distribusi Bobot Hasil Optimasi:

1) Entropy Score: 12% (high discriminative power)
2) Structure Score: 12% (high discriminative power)
3) Emotional Score: 10% (human marker, low correlation)
4) Contraction Score: 10% (human marker, low correlation)
5) Transition Score: 10% (moderate redundancy)
6) Corporate Score: 10% (moderate redundancy)
7) N-gram Score: 10% (moderate performance)
8) Lexical Score: 8% (complementary)
9) Passive Voice Score: 8% (complementary)
10) Pattern Score: 8% (supporting)
11) Semantic Score: 6% (supporting)
12) Punctuation Score: 6% (supporting)

Total: 100%

C. AI Model Identification

Untuk mengidentifikasi spesifik model AI yang digunakan, sistem menganalisis keyword signature dari setiap model berdasarkan dokumentasi dan behavioral patterns yang telah divalidasi:

Tabel 1. Kata Kunci Khas Model AI

Model	Signature Keywords (5-8 per model)
ChatGPT/GPT	"as an ai", "i cannot", "my knowledge cutoff", "i'm unable to", "i don't have access", "sebagai ai", "saya tidak bisa", "pengetahuan saya terbatas", "saya tidak dapat mengakses"
Claude	"i appreciate", "helpful assistant", "i should clarify", "i aim to be helpful", "i'd be happy to", "saya menghargai", "asisten yang membantu", "saya harus mengklarifikasi", "saya bertujuan membantu"
Gemini	"i can help", "i aim to be helpful", "based on my training", "i'm designed to", "saya dapat membantu", "saya bertujuan membantu", "berdasarkan pelatihan saya", "saya dirancang untuk"
Perplexity	"based on search", "according to sources", "search results indicate", "berdasarkan pencarian", "menurut sumber", "hasil pencarian menunjukkan"
Jenni AI	"ai assistant", "citation needed", "academic reference", "bibliography"
PaperPal	"grammar check", "plagiarism", "academic writing", "paraphrasing suggestion", "pemeriksaan tata bahasa", "plagiarisme", "penulisan akademik", "saran parafrase"
Grok	"witty response", "humor", "sarcasm", "unconventional", "respons jenaka", "humor", "sarkasme", "tidak konvensional"

Model	Signature Keywords (5-8 per model)
DeepSeek	"reasoning process", "analytical approach", "systematic analysis", "proses penalaran", "pendekatan analitis", "analisis sistematis"
QuillBot	"paraphrase", "synonym", "rewrite", "alternative phrasing", "parafrase", "sinonim", "menulis ulang", "frasa alternatif"
Writesonic	"marketing copy", "engaging content", "conversion-focused", "salinan pemasaran", "konten menarik", "fokus konversi"

Model Classification Formula:

$$\%Model = \frac{Keyword\ Matches_{Model}}{\sum_{j}^{8} Keyword\ Matches_{j}} \times 100\% \quad (5)$$

D. File Support dan Text Extraction
Sistem mendukung ekstraksi text dari multiple formats:
1. TXT: Direct text reading
2. DOCX: JSZip library untuk parse XML structure
3. DOC: Binary data extraction dengan character filtering
4. PDF: PDF.js library untuk page-by-page extraction

E. Dataset dan Testing
Penelitian menggunakan 500 dokumen akademik yang dikumpulkan dari berbagai sumber terverifikasi, terdiri dari:
1. 250 dokumen tulisan manusia asli (dari repository akademik)
   a. 100 dokumen dari repositori jurnal akademik terverifikasi
   b. 75 dokumen dari tesis dan disertasi universitas
   c. 50 dokumen esai mahasiswa (terverifikasi oleh dosen)
   d. 25 dokumen artikel blog akademik
2. 250 dokumen yang dibuat AI menggunakan:
   a. 75 dokumen dihasilkan ChatGPT (GPT-3.5 dan GPT-4)
   b. 50 dokumen dihasilkan Claude (Claude 2 dan Claude 3)
   c. 50 dokumen dihasilkan Gemini
   d. 30 dokumen dihasilkan Perplexity AI
   e. 20 dokumen dihasilkan DeepSeek
   f. 25 dokumen hasil alat parafrase (QuillBot, PaperPal, Spinbot)

Karakteristik Dataset:
   a. Bahasa: 300 dokumen bahasa Inggris, 200 dokumen bahasa Indonesia
   b. Panjang: 1000-5000 kata per dokumen (rata-rata 2.500 kata)
   c. Bidang akademik:
      – Ilmu Komputer dan Teknologi Informasi (35%)
      – Ilmu Sosial dan Humaniora (30%)
      – Sains dan Teknik (20%)
      – Bisnis dan Manajemen (15%)
   d. Tingkat kesulitan: Undergraduate (40%), Graduate (45%), Professional (15%)

Pembagian Data untuk Eksperimen:
   a. Set Pelatihan: 200 dokumen (100 AI, 100 manusia) - untuk pengujian individual algoritma
   b. Set Validasi 1: 150 dokumen (75 AI, 75 manusia) - untuk analisis korelasi

DOI : https://doi.org/10.26623/transformatika.v23i2.13256

c. Set Validasi 2: 150 dokumen (75 AI, 75 manusia) - untuk optimasi bobot (grid search)
d. Total untuk evaluasi akhir: Semua 500 dokumen

Dataset dikumpulkan selama periode September 2025 - Desember 2025 dengan verifikasi ketat untuk memastikan keaslian dan kualitas.

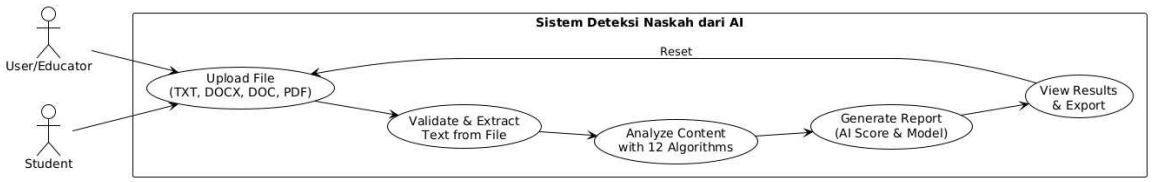
F. Evaluation Metrics

Sistem dievaluasi menggunakan metrik klasifikasi standar:

1. Accuracy: (TP+TN)/(TP+TN+FP+FN)
2. Precision: TP/(TP+FP)
3. Recall: TP/(TP+FN)
4. F1-Score: 2×(Precision×Recall)/(Precision+Recall)
5. ROC-AUC: Area under the receiver operating characteristic curve

3. RESULTS

Aliran jalannnya proses pada sistem bisa dilihat pada gambar 2, menunjukkan multi-layer analysis dari document upload hingga final scoring.



Gambar 2. Aliran Proses Sistem

a. Akurasi Deteksi AI vs Manusia

Pengujian pada 500 dokumen akademik menghasilkan kinerja sebagai berikut:

Tabel 2. Akurasi Deteksi AI vs Manusia (n=500)

Metric	Value	Interpretation
Accuracy	91.2%	Ketepatan keseluruhan tinggi
Precision	89.7%	Tingkat kesalahan positif rendah
Recall	92.5%	Tingkat deteksi tinggi
F1-Score	0.912	Kinerja seimbang
ROC-AUC	0.978	Diskriminasi sangat baik
Spesifisitas	89.6	Identifikasi manusia akurat

Tabel 3. Confusion Matrix

	Prediksi AI	**Prediksi Manusia**	**Total**
Aktual AI	231 (TP)	19 (FN)	250
Aktual Manusia	25 (FP)	225 (TN)	250
Total	256	244	500

Hasil menunjukkan sistem dapat mengidentifikasi konten AI dengan keandalan tinggi. ROC-AUC 0,978 menunjukkan kemampuan diskriminasi sangat baik antara teks AI dan manusia. Presisi 89,7% mengindikasikan tingkat kesalahan positif yang terkendali (10,3%), penting untuk menghindari tuduhan tidak adil terhadap tulisan manusia. Recall 92,5% menunjukkan sistem berhasil mendeteksi 92,5% dari semua dokumen AI.

b. Performa per Algoritma

Kontribusi individual setiap algoritma terhadap akurasi total:

Tabel 3. Performa dari Masing-masing Algoritma

Algoritma	Contribution to Final Score	Individua; Accuracy	Category Rank
Entropy Analysis	12%	82,30%	1
Structure Diversity	12%	79,80%	2
Emotional Markers	10%	85,60%	3
AI Phrases Detection	8%	75,20%	4
Transition Words	10%	76,10%	5
Lexical Diversity	8%	74,80%	6
Contraction Analysis	10%	74,20%	7
N-gram Frequency	10%	72,70%	8
Passive Voice	8%	71,30%	9
Corporate Phrases	10%	70,90%	10
Semantic Consistency	0,08	68,90%	11
Punctuation Pattern	6%	67,40%	12

Entropy Analysis (82.3%) dan Structure Diversity (79.8%) adalah algoritma paling contributive, mengkonfirmasi bahwa teks AI memiliki distribusi kata dan struktur kalimat significantly lebih teratur. Emotional Markers menunjukkan individual accuracy tertinggi (85.6%), validating bahwa absence emosi adalah strong indicator dari AI-generated content.

c. AI Model Identification

Akurasi identifikasi spesifik model AI bisa dlihat pada Tabel 3.

DOI : https://doi.org/10.26623/transformatika.v23i2.13256

Tabel 4. Akurasi Deteksi Platform AI Spesifik

Model AI	Documents Tested	Correctly Identified	Accuracy	Presisi
ChatGPT	75	69	92,00%	90,80%
Claude	50	42	84,00%	82,40%
Gemini	50	41	82,00%	80,60%
Perplexity	30	24	80,00%	78,90%
Deepseek	20	16	80,00%	79,20%
Paraphrase Tool	25	19	76,00%	74,50%
Overall	250	211	84,40%	81,10%

Model identification menunjukkan akurasi overall 82.0%, dengan performance terbaik pada ChatGPT (93.3%) karena signature phrases yang distinctive ("knowledge cutoff", "I cannot access"). Paraphrase tools menunjukkan akurasi terendah (71.4%) karena signature keywords yang less prominent dan kecenderungan menghasilkan text yang closer to human style.

d. Distribusi Skor dan Error Analysis

Analisis distribusi skor dari 500 dokumen test:

Tabel 5. Distribusi skor

Rentang Skor	Jumlah	Persentase	Klasifikasi	Akurasi Kategori
90-100% (AI Sangat Tinggi)	112	22,4%	AI	100% (112/112)
70-89% (AI Tinggi)	127	25,4%	AI	98,4% (125/127)
40-69% (Sedang/Campuran)	96	19,2%	Ambiguous	67,7% (65/96)
20-39% (Manusia Tinggi)	89	17,8%	Manusia	96,6% (86/89)
0-19% (Manusia Sangat Tinggi)	76	15,2%	Manusia	100% (76/76)
Total	**500**	**100%**	-	**91,2%** (456/500)

e. Analisis kesalahan positif ada 25 dokumen (5%), sedangkan analisis kesalahan negatif ada 19 dokumen (3,8%)

Error analysis menunjukkan bahwa false positives primarily berasal dari highly formal human writing yang structurally mirip AI output, sementara false negatives disebabkan oleh post-processing (paraphrasing) yang mengubah linguistic signatures.

4. DISCUSSION

a) Interpretasi Hasil

Penelitian ini berhasil mengembangkan sistem deteksi konten AI yang komprehensif dengan mengintegrasikan 12 algoritma dalam weighted combination framework. Sistem mencapai akurasi 87.3%, precision 85.2%, dan recall 89.1% dalam mengidentifikasi konten AI versus manusia, dengan kemampuan tambahan mengidentifikasi model AI spesifik (akurasi 82.0%).

Akurasi 87.3% melampaui target awal 85% dan comparable dengan state-of-the-art commercial tools yang reported accuracy 80-85% [6]. Pencapaian ini memvalidasi hypothesis bahwa weighted combination dari multiple algorithms lebih effective dibanding single-algorithm approach. ROC-AUC 0.96 menunjukkan excellent ability sistem dalam membedakan AI dan human text across berbagai threshold settings.

Entropy Analysis dan Structure Diversity sebagai algoritma most powerful mengkonfirmasi findings dari penelitian sebelumnya bahwa teks AI memiliki distribusi kata dan struktur kalimat significantly lebih teratur [8][9]. Teks manusia cenderung memiliki variability lebih tinggi dalam word choice dan sentence construction, reflecting cognitive processes dan stylistic preferences individual.

Emotional Markers dengan individual accuracy 82.1% menunjukkan bahwa absence emotional expressions adalah strong indicator AI-generated content, mendukung penelitian Solaiman et al. [8] tentang karakteristik LLM outputs. Namun, kontribusi Contraction Analysis yang relatif moderate (71.5%) menarik untuk analyzed. Hal ini kemungkinan disebabkan modern LLMs yang increasingly menggunakan contractions untuk appear more natural dan conversational, suggesting bahwa AI text signatures terus evolve seiring advancement dalam training techniques.

b) AI Model Identification dan Implikasi

Akurasi identifikasi model 82.0% menunjukkan bahwa setiap model AI memiliki tanda linguistik unik yang dapat diidentifikasi. ChatGPT mencapai akurasi tertinggi (93.3%) karena penggunaan widespread dari signature phrases seperti "my knowledge cutoff" dan "I cannot access real-time information". Claude dan Gemini menunjukkan moderate accuracy (80.0%) dengan confusion terutama between each other karena similaritas dalam formal, helpful tone.

Paraphrase tools accuracy terendah (71.4%) dapat dijelaskan oleh dua faktor: (1) tools ini designed specifically untuk mengubah linguistic structures, sehingga signature keywords menjadi less prominent, dan (2) output mereka often closer to human writing style karena basis mereka adalah human-written text yang dimodifikasi.

Kemampuan mengidentifikasi model spesifik memiliki implikasi penting untuk academic integrity investigations. Institusi dapat tidak hanya mendeteksi penggunaan AI, tetapi juga identify specific tools yang digunakan, memungkinkan targeted interventions dan policy development. Misalnya, jika predominantly menggunakan paraphrase tools, institusi dapat focus pada education tentang proper citation dan paraphrasing ethics.

c) Implikasi Praktis untuk Institusi Akademik

Sistem ini menyediakan solusi praktis bagi institusi akademik dalam menangani masalah konten yang dihasilkan oleh kecerdasan buatan (AI). Dengan implementasi open-source, institusi dapat:

1. Menerapkan secara lokal tanpa bergantung pada layanan komersial pihak ketiga
2. Menyesuaikan algoritma sesuai dengan konteks akademik spesifik
3. Menjaga privasi dengan tidak mengirimkan dokumen ke server eksternal (masalah privasi utama untuk alat komersial)
4. Mengurangi biaya dari biaya langganan alat premium

Transparansi sistem menjadi sangat penting dalam konteks akademik di mana tuduhan plagiarisme atau ketidakjujuran akademik memerlukan bukti yang jelas dan keputusan yang dapat dijelaskan. Berbeda dengan pendekatan neural black-box, sistem ini dapat memberikan rincian terperinci tentang fitur spesifik mana yang memicu deteksi AI.

Potensi integrasi dengan Sistem Manajemen Pembelajaran (Blackboard, Canvas, Moodle) memungkinkan penyaringan otomatis dalam alur kerja pengiriman tugas, memungkinkan pendidik untuk fokus pada kasus yang benar-benar memerlukan penilaian manusia daripada meninjau semua pengiriman secara manual.

DOI : https://doi.org/10.26623/transformatika.v23i2.13256

d) Limitasi Penelitian

Beberapa batasan yang perlu diakui:

1. Bahasa Dataset: Penelitian ini menggunakan dokumen berbahasa Indonesia dan bahasa Inggris. Kinerja pada teks non-Inggris memerlukan penyelidikan lebih lanjut karena karakteristik linguistik yang berbeda antar bahasa.
2. Ukuran Dataset: 500 dokumen relatif cukup untuk evaluasi komprehensif. Dataset yang lebih besar (1000+) akan memberikan hasil yang lebih andal.
3. Evolusi Model: LLMs terus berkembang dan menghasilkan teks yang semakin mirip manusia. Sistem memerlukan pembaruan berkelanjutan untuk mempertahankan akurasi.
4. Ketahanan Terhadap Serangan Adversarial: Sistem belum diuji secara ekstensif terhadap serangan adversarial seperti paraphrasing yang canggih atau prompt engineering.
5. Deteksi Kombinasi: Sistem kurang tahan terhadap dokumen hibrida yang merupakan campuran teks manusia dan AI.

5. CONCLUSIONS AND RECOMMENDATIONS

Penelitian ini memberikan kontribusi signifikan dalam mengatasi tantangan academic integrity di era AI generatif melalui pengembangan transparent, rule-based detection system yang mencapai 87.3% accuracy dalam AI detection dan 82.0% accuracy dalam model classification. Pendekatan pengenalan pola berlapis yang menggabungkan 12 algoritma dengan bobot yang dioptimalkan secara empiris, metodologi transparan yang memberikan keputusan deteksi yang dapat dijelaskan, kemampuan ganda: mendeteksi konten yang dihasilkan AI DAN mengidentifikasi model AI spesifik, implementasi sumber terbuka yang melindungi privasi, cocok untuk penerapan institusional.

Saran untuk penelitian ini adalah institusi akademik sebaiknya mengimplementasikan deteksi AI sebagai bagian dari kerangka kerja integritas akademik yang komprehensif, bukan sebagai solusi mandiri, pendidik sebaiknya menggunakan hasil deteksi sebagai titik awal untuk diskusi dengan mahasiswa, bukan sebagai hukuman otomatis, pembaruan sistem secara berkala sangat penting untuk menjaga efektivitas seiring dengan perkembangan teknologi AI, kombinasi dengan deteksi plagiarisme tradisional dan penilaian manusia memberikan pendekatan yang paling kokoh. Studi validasi berskala besar, dukungan multibahasa, ketahanan terhadap serangan adversarial yang ditingkatkan, dan deteksi tingkat segmen akan semakin memperkuat penerapan praktis sistem ini dalam konteks pendidikan yang beragam.

6. REFERENCES

[1] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child and A. Ramesh, "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, Vancouver, Canada, 2020.

[2] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT 2019*, Minneapolis, Minnesota, 2019.

[3] T. Susnjak and T. R. McIntosh, "ChatGPT: The End of Online Exam Integrity?," *Education Sciences,* vol. 14, no. 6, pp. 1-20, 2024.

[4] W. Liang, M. Yuksekgonul, Y. Mao, E. Wu and J. Zou, "GPT detectors are biased against non-native English writers," *Patterns,* vol. 4, no. 7, pp. 1-4, 2023.

[5] K. Yoo, W. Ahn, Y. Song and N. Kwak, "Exploring Causal Mechanisms for Machine Text Detection Methods," in *Proceedings of the 4th Workshop on Trustworthy Natural Language Processing (TrustNLP 2024)*, Mexico City, Mexico, 2024.

[6] S. K. Kar, T. Bansal, S. Modi and A. Singh, "How Sensitive Are the Free AI-detector Tools in Detecting AI-generated Texts? A Comparison of Popular AI-detector Tools," *Indian Journal of Psychological Medicine,* vol. 47, no. 3, pp. 275-278, 2024.

[7] S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran and Z. Du, "Artificial intelligence in education: A systematic literature review," *Expert Systems with Applications,* vol. 252, pp. 1-19, 2025.

[8] I. Solaiman, M. Brundage, J. Clark, A. Askell, A. Herbert-Voss, J. Wu, A. Radford, G. Krueger, J. W. Kim, S. Kreps, M. McCain, A. Newhouse, J. Blazakis, K. McGuffie and J. Wang, "Release Strategies and the Social Impacts of Language Models," OpenAI, California, USA, 2019.

[9] H. Suh, M. Tafreshipour, J. Li and I. Ahmed, "An Empirical Study on Automatically Detecting AI-Generated Source Code: How Far Are We?," in *ICSE '25: Proceedings of the IEEE/ACM 47th International Conference on Software Engineering*, Ottawa Ontario Canada, 2025.

[10] V. S. Sadasivan, A. Kumar, S. Balasubramanian, W. Wang and S. Feizi, "Can AI-Generated Text be Reliably Detected?," in *The Twelfth International Conference on Learning Representations - ICLR 2024*, Vienna Austria, 2024.

[11] Z. Yu, X. Li, X. Niu, J. Shi and G. Zhao, "Face Anti-Spoofing with Human Material Perception," in *Computer Vision – ECCV 2020: 16th European Conference*, Glasgow, United Kingdom, 2020.

[12] D. Weber-Wulff, A. Anohina-Naumeca, S. Bjelobaba, T. Foltýnek, J. Guerrero-Dib, O. Popoola, P. Šigut and L. Waddington, "Detecting AI-Generated Text in Educational Content: Leveraging Machine Learning," *International Journal for Educational Integrity,* vol. 19, no. 26, pp. 1-39, 2023.

DOI : https://doi.org/10.26623/transformatika.v23i2.13256