International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Comparing ai detectors: evaluating performance and efficiency

Jeremie Busio Legaspi [1, *], Roan Joyce Ohoy Licuben [2], Emmanuel Alegado Legaspi [2] and Joven Aguinaldo Tolentino [2]

[1] *College of Education, Tarlac Agricultural University, Philippines*
[2] *College of Engineering and Technology, Tarlac Agricultural University, Philippines*

International Journal of Science and Research Archive, 2024, 12(02), 833–838

Publication history: Received on 02 June 2024; revised on 15 July 2024; accepted on 18 July 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1276

Abstract

The widespread utilization of AI tools such as ChatGPT has become increasingly prevalent among learners, posing a threat to academic integrity. This study seeks to evaluate capability and efficiency of AI detection tools in distinguishing between human-authored and AI-generated works.

Three-paragraph works on "AutoCAD and Architecture" were generated through ChatGPT, and three human-written works were subjected to evaluation. AI detection tools such as GPTZero, Copyleaks and Writer AI were used to evaluate these paragraphs. Parameters such as "Human/Human Text/Human Generated Text" and "AI/AI Content Detected" were used to evaluate the performance of the three AI detection tools in evaluating outputs. Findings indicate that GPT Zero and Copyleaks have higher reliability in determining human-authored work and AI generated work while Writer AI showed a notable content classification of "Human Generated Content" on all tested outputs showing less sensitivity on determining human-authored work and AI generated work.

Findings indicate that the use of Artificial Intelligence as an AI detection tool should be accompanied with thorough validation and cross-referencing of results.

Keywords: ChatGPT; GPTZero; Copyleaks; Writer AI; Artificial Intelligence

1. Introduction

The availability and usage of Artificial Intelligences in various web browsers had brought about significant changes in the educational system, resulting to issues on academic integrity and plagiarism. It has been very easy among learners to resort to copying and posting information from the internet to their schoolwork and claim it as their own work. Plagiarism means taking ideas or words from a source, such as book or journal, without giving credit (acknowledgment) to the author (Bailey, 2015, p. 31). Academic integrity on the other hand encompasses the values, behavior and conduct of academics in all aspects of their practice (Mcfarlane, 2012). Academic plagiarism has been a concern on the educational system as it creates breach on academic integrity resulting to intellectual dishonesty, especially among learners. As a result, implementing strategies to reduce plagiarism is vital for preserving academic integrity and preventing such dishonest practices in students' future scholarly and professional endeavors (Alsallal et al. 2013; Elkhatat 2022; Foltýnek et al. 2020).

As ChatGPT has been made accessible and free, it has been considered as a potential risk for cheating and plagiarism. ChatGPT is a sophisticated computer program that uses a type or artificial intelligence called natural language processing to understand and respond to human language. It has been trained on a vast amount of text data and uses that knowledge to generate responses to a wide range of questions and topics (McGeorge, 2023, p.7).

* Corresponding author: Jeremie B. Legaspi

Nowadays, AI detection tools are used by educators to assess originality and plagiarism on students' work. For instance, GPTZero, Copyleaks and Writers are AI detection tools which can easily distinguish human-authored work from AI-generated works. GPTZero is an artificial intelligence detection software developed to identify artificially generated text, such as that produced by large language models. While GPTZero has received positive coverage for its efforts to prevent academic dishonesty, its reported outputs of false positives have been source of criticism (Tian, et. al., 2023). Copyleaks on the other hand, scans through a vast database of websites, articles, documents, and previous submissions to compare with your students' submitted content. It will then provide a "similarity score" which is the percentage of how much of the student's submitted content matches the sources in the database. In addition to the similarity score, Copyleaks provides a report with more granular information such as which exact phrases and sentences match the database sources. It also provides a curated list of all the different database sources that it found a match within the student's work for you to be able to compare and contrast (Yamin, et al., 2015), while Writer AI is an AI detection tool which assess what percentage of your content is seen as human generated.

Artificial intelligence-based tools updates overtime, increasing the risk of highly undetectable, plagiarized work in the educational system. The usage of AI detection tool facilitates an avenue to minimized and irradicate the practice of plagiarism and inculcate learners the value of learning ethics and the cultivation of human dignity as a crucial factor for personal development and well-being. Furthermore, this study aims to investigate capability and efficiency of AI detection tools in distinguishing between human-authored and AI-generated works.

2. Review of related literature

AI has been increasingly propagated as having strategic value for education and could be an effective learning tool that lessens the burdens of both teachers and students and offers effective learning experiences for students. Coupled with current education reforms such as the digitalization of educational resources, gamification, and personalized learning experiences, there are many opportunities for the development of AI applications in education. (Zhai, 2021).

However, the problem of students using AI to cheat on schoolwork became palpable. While many existing AI content detectors can detect AI-generated texts, such as GPT-2 Content Detector and GPTZero, Turnitin, Copyleaks, and Writer AI, the accuracy of an AI content detector in detecting generated essays that have been post-edited by humans is unknown (Wu, H., & Flanagan, T. (2023). Since most of the AI content detector tools are new, not much research has been conducted to evaluate their efficacy, accuracy, and reliability in terms of distinguishing between works which are generated either by AI or written by humans. GPTZero, Copyleaks and Writer AI are some of the various AI detection that are widely studied for their efficiency, accuracy, and reliability in distinguishing between AI-generated and human-written works.

GPTZero as its name indicates, is intended to detect whether a text generated by ChatGPT is AI-generated or human-written (Chaka, 2023a; Ofgang, 2023; Tech Desk, 2023; Tyrrell, 2023). In this sense, it has a wider application beyond the ChatGPT-generated text or ordinary human-written responses or essays that have nothing to do with AI generation. GPTZero identifies whether the text is AI-generated or human-produced by using the two measures: perplexity and burstiness. Perplexity measures a text's randomness. The understanding here is that a handwritten text displays randomness or chaoticness and, thus, is likely to perplex or be unfamiliar with a language model such as GPTZero. The higher the perplexity of the text, the higher the likelihood that it is human written. The converse is true: the lower the text's perplexity, the lower the likelihood that it is human written. This lower perplexity index signals that a text is AI-generated. Burstiness measures the complexity of sentences or how highly varied sentence usage is in a text. The belief here is that humans are prone to varying the types and the length of their sentences when they write, while machines are not. So, burstiness relates to sentence variability or sentence bursting (Chaka, 2023a; Ofgang, 2023). Most importantly, GPTZero sometimes highlights or flags an AI-generated text in yellow in any given sample and allocates perplexity and burstiness scores to text samples. Higher scores for both measures indicate that a text is human-generated, while lower scores for both measures signal that a text is AI-generated. One of the drawbacks of this tool is that it sometimes misclassifies or misrecognizes portions of a text as either AI-generated or human-generated, even in instances where that is not the case (Tyrrell, 2023). So, it is not 100% per cent accurate (Chaka, 2023b).

However, according to Outlook Spotlight, 2023 Writer.com AI which is owned by Writer.com, unlike most of its peers, it is a no-sign-up or a no-create-an-account tool for usage. It evaluates a text and identifies (by calculating) how much of it is likely AI-generated through percentage scores. It has a 1,500-character limit per text/prompt where text can be added to this detector by pasting or writing it or by providing a URL of the intended text. The AI tool does not have a 100% accuracy rate, and sometimes, it can be tricked by certain texts (Help Center, 2023; see Lim, 2023). It can also be used for editing and generating text, and its parent company, Writer.com, has offerings such as products (e.g.,

Grammarly alternative, ChatGPT alternative, and Jasper alternative) and resources (e.g., Inclusive language and AI content generator) (Help Center, 2023; Outlook Spotlight, 2023).

Copyleaks is an AI content detection tool that is used to determine whether a text is generated by AI chatbots like ChatGPT and many other AI tools or whether a text is written by a human. Copyleaks is an AI content detector which can detect AI content written in multiple languages such as English, Spanish, Polish, Italian, and a few other languages, with more other languages being currently considered, verifying the authenticity of social media posts, online news articles, online reviews, etc. This tool has differentiating feature: A 99.12% detection accuracy rate In-depth, detailed analysis Detecting GPT-J, GPT-3, GPT-3.5, ChatGPT, GPT-4, and other related AI language models. (Chrome, 2023; Copyleaks, 2023).

According to Minitab 2023, Text with less than 20% AI content were classified as "very unlikely AI-generated" those with 20%-40% AI content was considered "unlikely AI-generated", those with 60-80% AI content were deemed "unclear if AI-generated", those 60-80% AI content were labeled "possibly AI-generated." Those with over 80% AI content were categorized as "likely AI-generated". However, Will Yeadon, 2024 states that Zero GPT achieved a 98% accuracy rate and proposes that text with ≤50% AI-generated content should be considered the upper limit for classification as human-authored.

2.1. Conceptual framework

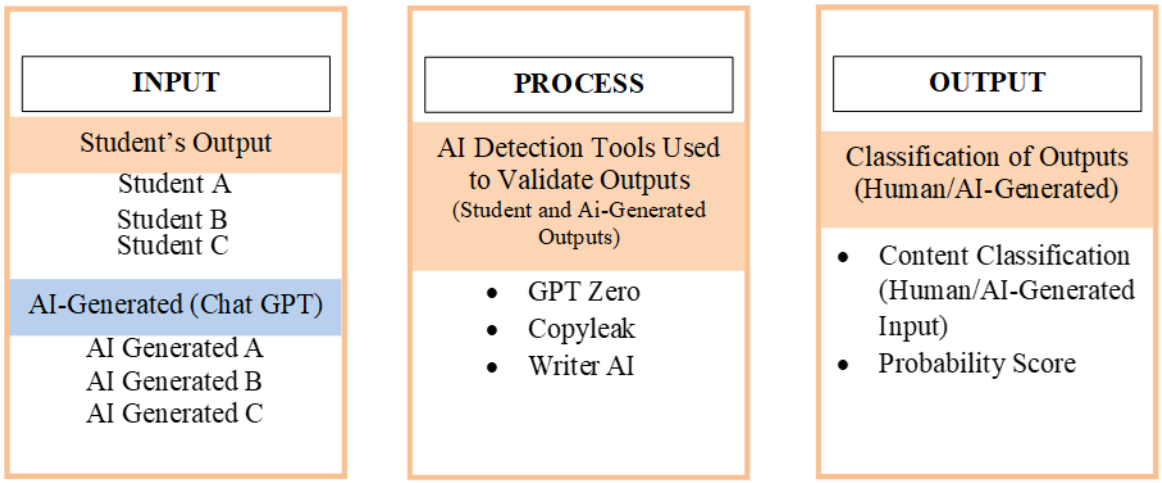

Figure 1 Conceptual framework

3. Methodology

The ChatGPT chatbot generated three essay works, each of which is composed of a three-paragraph response on "AutoCAD

and Architecture". The initial prompt was to "compose a three-paragraph essay work on AutoCAD and Architecture". Three human-authored work on the same topic will be used as control samples to evaluate false positive response of AI detectors.

Three AI content detectors were selected and evaluated for their capability and efficiency in determining human-written and AI generated works namely, GPT Zero, Copyleaks and Writer AI. The three AI detectors were selected based on in-depth research of existing references and recommendations of professionals at the time of the study.

AI detectors responses were classified as positive if the original content was human-authored work and the detector's classification was "human" and a negative response if the original content was AI generated and the detector's response was "human".

4. Result and discussion

This chapter presents the results and discussion of the study, which focuses on the performance and efficiency of AI detectors in determining human-authored work and AI generated works.

Table 1 Content classification of GPT Zero on AI generated works and human-authored works.

Samples	Content Classification	Probability AI-Generated
Student's Output A	Human	3%
Student's Output B	Human	7%
Student's Output C	Human	5%
AI Generated A	AI	100%
AI Generated B	AI	100%
AI Generated C	AI	100%

Table shows the outcome content detection using GPT Zero implemented on three human generated works and 3 AI-generated works. Result shows that student generated outputs were classified as human work with corresponding probability of three percent probability as AI-generated work on Student's Output A, seven percent probability on Student's Output B, and a five percent probability as AI-generated work on Students' Output C. On the other hand, all AI generated outputs showed a notable content classification of 100% AI-generated.

Table 2 Content classification of Copyleaks on AI generated works and human authored works**.**

Samples	Content Classification	Probability AI-Generated
Student's Output A	Human Text	
Student's Output B	Human Text	
Student's Output C	Human Text	
AI Generated A	AI Content Detected	
AI Generated B	AI Content Detected	
AI Generated C	AI Content Detected	

Table 2 shows results of the content classification of human-authored work and AI generated works using Copyleaks AI detector. Result shows that all Student's Outputs have a content classification of "Human Text" while AI-Generated samples A, B and C were classified as AI-Generated. However, the percentage of AI generated probabilities among all samples were not identified. Response as "This is human text" and "AI content detected" were the parameters given by the AI content detector as an indicator.

Table 3 Content classification using Writer AI on AI generated works and human authored works.

Samples	Content Classification	Probability AI-Generated
Student's Output A	Human Generated Content	1%
Student's Output B	Human Generated Content	1%
Student's Output C	Human Generated Content	0%
AI Generated A	Human Generated Content	17%
AI Generated B	Human Generated Content	27%
AI Generated C	Human Generated Content	30%

Table 3 outlines the outcome of Writer AI as an AI detection tool implemented on three human-authored work samples and three AI generated work samples. Result shows that content of Student's Output A, B and C were classified as "Human Generated Content" with a probability of 99 % (Student Output A), 99% (Student Output B), and 100% (Student Output C), respectively. However, result of AI-Generated A shows a notable percentage of 83 with a probability description of "Human Generated Content", 73% as "Human Generated Content" on AI-Generated B sample and a 70% probability result as "Human Generated Content" on AI-Generated C sample were shown.

As an implication to the study, performance of the Writer AI as an AI detection tool was notably less consistent, while based on the result of the tested samples, GPT Zero and Copyleaks showed a more remarkable reliability in detecting AI generated work and human-authored works.

5. Conclusion

This study sought to investigate the performance and efficiency of three AI detection tools namely, GPT Zero, Copyleaks, and Writer AI. The result of this study indicates the considerable reliability of the three AI detection tools to correctly identify and categorize works as "AI-Generated" and "Human-Authored works".

Results of the study showed that Writer AI has the least sensitivity in categorizing AI-generated work and human-authored work. Based on the results of this study, GPT Zero and Copyleaks on the other hand, showed significant and accurate performances on categorizing AI-generated work and human-authored work.

With the advancement of technologies and digitalized intelligences, considerable attention should be given to the reliability and usage of AI detection tools. AI detection tools can give significant insights and contribution in resolving academic integrity issues which should be supplemented with cross-referencing information across credible sources.

Compliance with ethical standards,

Disclosure of conflict of interest

Authors have no conflict of interest to disclose.

References

[1] BAILEY, S (2015). *The Essentials of Academic Writing for International Students.* https://books.google.com.ph

[2] MACFARLANE, B *et al* (2012). *Academic Integrity: A Review of the Literature* eSchoolNews (2024). *Impact of Artificial Intelligence in Education*

[3] TIAN, E et al (2023). *GPTZero: The Trusted AI Detector*

[4] YAMIN, A et al (2015). *Copyleaks: AI-Based Plagiarism and Ai Content* Detector. https://copyleaks.com

[5] MCGEORGE, D (2023). *The ChatGPT Revolution: How to Simplify Your Work and Life.* https://books.google.com.ph

[6] ALSALLAL M. et al (2013). *Intrinsic Plagiarism Detection Using Latent Semantic Indexing and Stylometry. 2013 Sixth International Conference on Development in eSystems Engineering*

[7] BAILEY, S (2015). *The Essentials of Academic Writing for International Students.* https://books.google.com.ph

[8] CINGILLIOGLU, I. (2023). *Detecting AI-generated essays: the ChatGPT challenge. The International Journal of Information and Learning Technology, 40(3), 259-268*.

[9] CHAKA C. (2023). *Generative AI chatbots - ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. International Journal of Learning, Teaching and Educational Research, 22(6), 1-19. https://doi.org/10.26803/ijlter.22.6.1*

[10] CHROME (2023). *AI content detector – Copyleaks. https:// chrome.google.com/webstore/detail/ai-content-detectorcopyl/gplcmncpklkdjiccbknjjkoidpgkcakd*

[11] COPYLEAKS (2023). *ChatGPT and AI content detection. https:// copyleaks.com/blog/chatgpt-and-ai-content-detection*

[12] MACFARLANE, B *et al* (2012). *Academic Integrity: A Review of the Literature*

[13] MCGEORGE, D (2023). *The ChatGPT Revolution: How to Simplify Your Work and Life.*https://books.google.com.pheSchoolNews (2024). *Impact of Artificial Intelligence in Education*

[14] TIAN, E et al (2023). *GPTZero: The Trusted AI Detector*

[15] WALTERS, W. H. (2023). The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. *Open Information Science*, *7*(1), 20220158

[16] WU, H., & FLANAGAN, T. (2023). The Limits of AI Content Detectors. *Journal of Student Research*, *12*(3).

[17] YAMIN, A et al (2015). *Copyleaks: AI-Based Plagiarism and Ai Content* Detector. https://copyleaks.com

[18] ZHAI, X., CHU, X., CHAI, C. S., JONG, M. S. Y., ISTENIC, A., SPECTOR, M., ... & LI, Y. (2021). *A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021, 1-18.*