

RESEARCH ARTICLE



AI-Generated Content Detectors: Boon or Bane for Scientific Writing

Namrata Ladha^{1,2*}, Kaushal Yadav¹, Priya Rathore³

¹ Senior Research Fellow, Devi Ahilya Vishawavidyalaya, Devi Ahilya Vishawavidyalaya, Indore, M.P., India

² Executive Programme Student, Indian Institute of Management, Kozhikode, Kerala, India

³ Junior Research Fellow, Devi Ahilya Vishawavidyalaya, Indore, M.P., India

 OPEN ACCESS

Received: 02-07-2023

Accepted: 12-09-2023

Published: 25-10-2023

Citation: Ladha N, Yadav K, Rathore P (2023) AI-Generated Content Detectors: Boon or Bane for Scientific Writing. Indian Journal of Science and Technology 16(39): 3435-3439. <https://doi.org/10.17485/IJST/v16i39.1632>

* **Corresponding author.**

namratladha777@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2023 Ladha et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Objectives: The purpose of the study is to help publishers identify AI-generated text in scientific research, academic works, and assignments as a critical step toward the regulation and promotion of the ethical usage of AI in academia.

Method: Recently developed literature on Generative AI suggests that human reviewers may fail to distinguish between human and AI-generated articles. Therefore, the present study evaluates AI-powered software as a potential solution for AI-generated content detection. We performed an experiment to see whether AI detector tools are capable of identifying and distinguishing between human- and AI-generated texts. To determine the accuracy of the AI-detector, we created and submitted four research articles to AI-detector tools for a pre- and post- manipulation test. **Findings:** The study shows that it is quite impossible for any AI detector to identify all AI-generated content accurately, thus, human-AI collaboration strategies can be employed to achieve the maximum accuracy. This paper demonstrates in a novel manner how AI detector tools can be manipulated to provide false results. **Novelty:** To the best of the authors' knowledge, this paper is the first to acknowledge growing AI-literacy among students and scholars. This study edifies academia by providing scientifically verified human-AI collaboration strategies to capitalize on these tools and thwart academic misconduct.

Keywords: Artificial Intelligence; Chatbot; ChatGPT; Generative AI; Scientific Research

1 Introduction

In the new era of human-AI collaboration, researchers can now save time, reduce errors, and automate most of the exhausting tasks. By leveraging generative AI, researchers can revolutionize academic research by processing large amounts of data, uncovering discoveries, generating hypotheses, and conducting literature reviews faster than traditional methods⁽¹⁻³⁾.

Generative AI chatbots, such as ChatGPT, Grammarly, and Copy.ai, generate text based on patterns and structures found in the vast amount of data they have been trained on. The literature has extensively discussed the challenges of AI-generated content⁽¹⁻⁴⁾,

including: AI writing tools failing to produce unique and novel content, delivering less value to readers, and identical, false, and fabricated content harming search engine optimization and the publisher's credibility^(4,5). While ChatGPT can generate original and unique content, there is a possibility that it will match with other existing content. Using AI-generated text without acknowledgement or citation of the original source can be considered as plagiarism^(6,7).

Since AI-generated content started to pose a threat to creativity and credibility, publishers, educators, and leading organizations are advising to refrain from the usage of generative AI^(8–10). With the growing AI literacy, it is becoming difficult for a human to identify AI-generated content in research articles. One helpful solution is AI-generated content detector software (further mentioned as AI-detector). This AI-powered technology utilizes enormous datasets to identify whether the content is truly human-written or created by a chatbot. The AI-detector can be useful for writers, researchers, educators, reviewers, and publishers to check the authenticity of research papers and student assignments.

The rapidly growing literature on generative AI is still in its early stages. Most articles provide a perspective rather than empirically and theoretically grounded research^(1,2,5–7). We used search terms like generative AI, Chatbot, ChatGPT, AI detection, academic writing, etc., to include relevant and incremental research studies from various databases. The literature review points out that the most recent investigators share their experiences of using ChatGPT, with an emphasis on academic integrity, capabilities of AI, and relative ethical concerns^(7,11–13).

The significant concern is the likelihood of ChatGPT producing false or fabricated information^(2,5). Beginners could fail to detect these falsifications and fabrications. As a result, the leading science journals recently prohibited all ChatGPT-generated text, figures, photos, or graphics^(2,5,8). It is discovered that ChatGPT could produce highly realistic writing with minimum input, thereby compromising the integrity of online exams^(12,13). Researchers compared scientific abstracts generated by ChatGPT. Blinded human reviewers correctly recognized ChatGPT as the source of 68% of the abstracts. However, human reviewers misclassified 14% of real abstracts as being AI-generated, indicating they were highly skeptical when reviewing the abstracts⁽¹⁴⁾.

Consequently, when human reviewers fail to detect AI-generated content, AI-detection tools can be a promising solution. AI-text output detectors may be useful in tackling this issue but are not infallible^(11,15–17). Previous researchers evaluated a set of two academic papers using ChatGPT and found it difficult to distinguish between AI-generated and original abstracts⁽¹¹⁾. However, one freely available AI-content detector performed a comparison between human and AI-generated abstracts, limiting the utility of the study. Similarly, several researchers evaluated the usability of AI-based detectors^(15–17). However, the input was simple essays with no complexity or hybrid writing, which is easily detectable by AI.

1.1 Research Gap

Due to the novelty of the field, existing researches have several shortcomings:

1. The current literature used simple, non-complex input following a single pattern, which makes the work of AI-detectors easier.
2. Previous studies were not able to evaluate true potential of AI-detectors by failing to recognize real-world scenarios.
3. Existing researches do not provide an alternative solution to the problem.

This paper creates new knowledge by keeping in view the growing AI-literacy among students and scholars and therefore, uses complex AI and Human written texts as inputs. This is a significant attempt that addresses the call for the research made by prior research in the domain^(1,4,5). This study aims to fill the research gap, examine the predictive accuracy of AI-detection tools, and recommend human-AI collaborated strategies to capitalize on these tools and achieve maximum accuracy. The novelty lies in raising awareness about the unprecedented threat that generative AI and its detectors pose to academic integrity.

The study makes several significant contributions. First, it is one of the early scientific studies in the emerging field of AI regulation in academic and scientific research. Second, it presents the result of one of the most rigorous tests conducted so far, based on a pre- and post-manipulation test design. Third, this study edifies academia by providing scientifically verified human-AI collaboration strategies to capitalize on AI-detector tools and thwart academic misconduct.

2 Methodology

2.1 Research design

For the purpose of determining the accuracy of the AI-detector, we created four research articles. These articles were submitted to AI-detector tools. The AI-detector, in response, provides a confidence percentage of whether an AI wrote that text. It is an experiment to see whether AI detector tools are capable of identifying and distinguishing between human and AI-generated text.

2.2 Stimuli

Article 1 was written by a human author. Its word length is 220 words. Article 2 was solely written by ChatGPT in response to the prompt “Write a research paper on the topic ‘AI usage in scientific writing’ using interviews of 15 content writers. Include sections like Literature review, Research Methodology, and Results and Discussion.” ChatGPT answered with an article of 254 words with requested sections. We did not request ChatGPT to include in-text citations to the claims because of its drawback of creating false citations^(2,5). We further inserted in-text citations to Articles 1 and 2 manually to create Articles 3 and 4, respectively. Article 4 was paraphrased using an AI-based paraphraser, Quillbot.

2.3 AI-detector tools

Three AI-detectors were employed to evaluate the above four articles. These AI-detector tools are developed by Copyleaks, Writer.com, and Content at scale to help combat AI-generated plagiarism. These three websites are top-ranked in the Google search engine. These AI tools are three of the most accurate content detectors⁽¹⁸⁾ and are trusted by leading organizations across the world^(19–21). They detect AI-generated text created across ChatGPT, GPT-4, GPT-3, Jasper, and others.

3 Results and Discussion

The confidence (probability) percentage reported in Table 1 presents whether AI-detector believes the article is human- or AI-generated.

Table 1. Table of AI-generated content percentage reported by AI detector tools for human-generated text and AI-generated text, pre and post paraphrasing and in-text citation (dated May 14, 2023)

Author	AI detector tool	Without paraphrasing and In-text citations	With paraphrasing and In-text citations	Change
AI	Copyleaks	91.4%	62.4%	↓
	Writer	99%	16%	↓
	Content at Scale	48%	16%	↓
Human	Copyleaks	32.5%	24.4%	↓
	Writer	0%	0%	=
	Content at Scale	17%	4%	↓

Findings show that AI-detector tools are able to identify AI-generated content with high accuracy. However, they falsely identify human-generated texts as AI with up to 32.5% probability (Table 1), as supported by the literature^(15–17). Thus, reviewers’ and educators’ reliance on these tools can turn problematic for the education system. Researchers and students may have to face penalties for plagiarism, which they did not commit^(22,23). The probable reason behind these false-positive results is the way AI-detectors work. An AI-detector employs AI to analyze patterns in the text and predicts the most likely word choices, resulting in an inflated AI detection probability⁽¹²⁾. It goes deep to detect robotic-sounding content. Thus, it is inferred that human writing, especially scientific writing, can still trigger AI probability if it follows a predictable pattern.

As per the results, after rewriting the text and inserting in-text citations, the confidence percentage goes down drastically (from 91.4% to 62.4% for Copy leaks, from 99% to 16% for Writer, and from 48% to 16% for Content at scale). By including in-text citations in the human-generated text, the AI-generated content probability is reduced by 8.1 to 13% (Table 1). We used an AI-based paraphrasing tool, Quillbot, to re-write the AI-generated text in different words and structures by maintaining the meaning and essence of the text. It is inferred that paraphrasing and in-text citations alter the pattern of text, thus defying the algorithm of AI detection tools.

4 Conclusion

Originality has a new threat in the name of generative AI. While this technology is arguably one of the greatest advancements, the advanced AI generation tools have exposed potential issues. Generative AI poses a threat to art, science, and academic integrity^(4,10,11). Moreover, ever since AI writing tools started to be advertised, the scientific community is debating over its usage in academic writing. Publication houses, regulators, and government are reckoning with the rapid pace of AI progression and its implications for society. ChatGPT and similar natural language processors are a concern for teachers and publishers since, when prompted, they can generate human-like content, making it easier for anyone to commit AI-driven plagiarism

without being noticed.

Therefore, various publishers are exploring how to regulate AI; AI-generated content detectors seem to be the only solution so far. Thus, this study contributes to academics and practice by evaluating the capabilities of AI-generated content detectors. The study tests AI-detectors' accuracy in four articles. Results show that the AI-detectors are able to identify AI-generated text with high accuracy, but the confidence percentage decreases when citations are included in the text. The AI detectors with current potential are able to recognize human-written text. Nevertheless, it has been discovered that human writing, especially scientific writing, can still trigger AI-generated content probability if it follows a predictable pattern.

The study shows that it is quite impossible for any AI-detector to accurately identify all AI-generated content alone; thus, human judgment is needed to achieve the maximum accuracy. The study concludes that human-AI collaboration is needed to safeguard academic integrity.

4.1 Implications for academics and practice

AI-generated content detectors can be of great help to researchers, educators, publishers, and copywriters by incorporating the findings of the study.

1. The accuracy of an AI detector depends on several factors, including the complexity of the text. The in-text citation breaks the pattern of text, thus making it difficult for AI detection. Thus, the study suggests removing all the citations from the text before proceeding to the AI detection test. It is suggested that journal editors can revise their paper editing guidelines to insert in-text citations numerically at the end of the sentence.
2. There may be cases where the AI detector produces false positives or false negatives. This can occur when the algorithm is not trained on enough diverse data or if the original article is hidden behind paywalls.
3. The precision of an AI detector should always be evaluated in the context of the task and application it is used for. Since research papers adhere to a scientific writing style, they may appear robotic to an AI-detector.
4. At this point, an individual can beat AI detection by inserting human-like inputs, informal and casual phrases, citations, typos, and slang. As a result, reviewers must pay special attention.
5. There are various methods for determining whether a piece of text was generated by AI. AI-generated text is characterized by the repetition of words and phrases, erroneous and outdated information, a generic tone, and a lack of originality^(12,24).
6. To avoid the risk of plagiarism, researchers must verify that the content is properly attributed or cited. Additionally, the study recommends using human experts and AI-powered plagiarism checkers as just one part of a larger strategy to identify and avoid AI-generated plagiarism.

References

- 1) Buriak JM, Akinwande D, Artzi N, Brinker CJ, Burrows C, Chan WC, et al. Best Practices for Using AI When Writing Scientific Manuscripts: Caution, Care, and Consideration: Creative Science Depends on It. *ACS nano*. 2023;17(5):4091–4093. Available from: <https://doi.org/10.1021/acsnano.3c01544>.
- 2) Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns*. 2023;4(3):1–4. Available from: <https://doi.org/10.1016/j.patter.2023.100706>.
- 3) Cao Y, Li S, Liu Y, Yan Z, Dai Y, Yu PS, et al. A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. 2023. Available from: <https://doi.org/10.48550/arXiv.2303.04226>.
- 4) Saxena A, Chang AYY, Kulathuramaiyer N, Bhatt A. The Conceptual Framework of Articulated Human Intelligence in Perspective of Industry 4.0. *Indian Journal of Science and Technology*. 2019;12(6):1–5. Available from: <https://doi.org/10.17485/ijst/2019/v12i6/141147>.
- 5) Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion Paper: “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*. 2023;71:1–63. Available from: <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- 6) Zheng H, Zhan H. ChatGPT in Scientific Writing: A Cautionary Tale. *The American Journal of Medicine*. 2023;136(8):725–726.E6. Available from: <https://doi.org/10.1016/j.amjmed.2023.02.011>.
- 7) Eke DO. ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*. 2023;13:1–4. Available from: <https://doi.org/10.1016/j.jrt.2023.100060>.
- 8) Indian Society for Education and Environment. Publication Policies. 2023. Available from: <https://sciresol.s3.us-east-2.amazonaws.com/IJST/Policy/policy-and-conditions.pdf>.
- 9) ChatGPT and Artificial Intelligence in higher education. 2023. Available from: https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf.
- 10) English Google SEO office-hours from April 1, 2022. 2023. Available from: <https://youtu.be/Yx997SsiYlw>. Accessed.
- 11) Anderson N, Belavy DL, Perle SM, Hendricks S, Hespagnol L, Verhagen E, et al. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sport & Exercise Medicine*. 2023;9(1):1–4. Available from: <http://dx.doi.org/10.1136/bmjsem-2023-001568>.

- 12) Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* . 2023;p. 1–12. Available from: <https://doi.org/10.1080/14703297.2023.2190148>.
- 13) Susnjak T. ChatGPT: The End of Online Exam Integrity?. 2022. Available from: <https://doi.org/10.48550/arXiv.2212.09292>.
- 14) Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo YS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine*. 2023;6(1):1–5. Available from: <https://doi.org/10.1038/s41746-023-00819-6>.
- 15) Cingillioglu I. Detecting AI-generated essays: the ChatGPT challenge. *International Journal of Information and Learning Technology*. 2023;40(3):259–268. Available from: <https://doi.org/10.1108/IJILT-03-2023-0043>.
- 16) Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, et al. Testing of Detection Tools for AI-Generated Text. 2023;p. 1–46. Available from: <https://arxiv.org/abs/2306.15666>.
- 17) Walters WH. The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors. *Open Information Science*. 2023;p. 1–27. Available from: https://www.researchgate.net/publication/373948577_The_effectiveness_of_software_designed_to_detect_AI-generated_writing_A_comparison_of_16_AI_text_detectors.
- 18) 8 Most Accurate and Reliable AI Content Detectors in 2023. 2023. Available from: <https://10web.io/blog/ai-detectors/Accessed>.
- 19) Copyleaks. Trusted by Leading Organizations. . Available from: <https://copyleaks.com/>.
- 20) The AI writing Chat platform of top teams. 2023. Available from: <https://writer.com/customers/>.
- 21) About us. 2023. Available from: <https://contentatscale.ai/about-us/>.
- 22) Professors are using ChatGPT detector tools to accuse students of cheating. But what if the software is wrong?. 2023. Available from: <https://www.usatoday.com/story/news/education/2023/04/12/how-ai-detection-tool-spawned-false-cheating-case-uc-davis/11600777002/>. Accessed May 10.
- 23) Misinformation machines? AI chatbots can spew falsehoods, even accuse people of crimes they never committed. 2023. Available from: <https://www.foxnews.com/media/misinformation-machines-ai-chatbots-spew-falsehoods-even-accuse-people-crimes-never-committed>. Accessed.
- 24) Ai. New AI classifier for indicating AI-written text. 2023. Available from: <https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>.