

How Sensitive Are the Free AI-detector Tools in Detecting AI-generated Texts? A Comparison of Popular AI-detector Tools

Sujita Kumar Kar¹ , Teena Bansal¹, Sumit Modi¹ and Amit Singh¹ 

ABSTRACT

Background: Recently, Artificial intelligence (AI) has significantly influenced academic writing. We aimed to investigate the sensitivity of the free versions of popular AI-detection software programs in detecting AI-generated text.

Methods: We searched for AI-content-detection software on Google and selected the first 10 free versions that allowed a minimum of 500 words for text analysis. Then, we gave ChatGPT 3.5 version a command to generate a scientific article on the "Role of Electroconvulsive Therapy (ECT) in Treatment-resistant Depression" under 500 words. After generating the primary text, we rephrased it using three different software tools. We then used AI-detection software to analyse the original and paraphrase texts.

Results: 10 AI-detector tools were tested on their ability to detect AI-generated text. The sensitivity ranged from 0% to 100%. 5 out of 10 tools detected AI-generated content with 100% accuracy. For paraphrased texts, Sapling and Undetectable AI detected all three software-generated contents with 100% accuracy. Meanwhile, Copyleaks, QuillBot, and Wordtune identified content

generated by two software programs with 100% accuracy.

Conclusion: The integration of AI technology in academic writing is becoming more prevalent. Nonetheless, relying solely on AI-generated content can diminish the author's credibility, leading most academic journals to suggest limiting its use. AI-content-detection software programs have been developed to detect AI-generated or AI-assisted texts. Currently, some of the platforms are equally sensitive. However, future upgrades may enhance their ability to detect AI-generated text more accurately.

Keywords: Artificial intelligence, AI-detector tool, ChatGPT, paraphrasing tools

Key Messages:

1. There is wide variation in the sensitivity of the AI-detector tools in detecting AI-generated texts.
2. Copyleaks, QuillBot, Sapling, Undetectable AI and Wordtune precisely detect the AI-generated content with 100% accuracy.
3. Sapling and Undetectable AI detected the AI-assisted paraphrased contents with 100% accuracy.

Artificial intelligence (AI) has significantly influenced scientific writing in recent years. AI is predicted to mark a critical point in the turnabout of mankind's technological evolution. It is helping people achieve academic and publishing goals, and its role will increase substantially soon. Lately, driven by the AI tool ChatGPT, which uses natural language processing (NLP) techniques to help create customized content through prompts, there has been a massive surge in AI-generated literary content.¹ However, with AI assisting people in writing content related to diverse topics, there is a valid concern regarding people needing to catch up on the tedious process of searching, organizing, writing, and revising literature content. Thus, with the help of AI, achieving specific academic goals has become more accessible, but the overall learning process is likely to suffer.² Also, there is increasing concern over academic plagiarism, which uses concepts, literary works, or organizational frameworks without crediting the original author.³ It has risen in educational settings and has been detected in numerous

¹Dept. of Psychiatry, King George's Medical University, Lucknow, Uttar Pradesh, India.

HOW TO CITE THIS ARTICLE: Kar SK, Bansal T, Modi S and Singh A. How Sensitive Are the Free AI-detector Tools in Detecting AI-generated Texts? A Comparison of Popular AI-detector Tools. *Indian J Psychol Med.* 2024;XX:1-4.

Address for correspondence: Sujita Kumar Kar, Dept. of Psychiatry, King George's Medical University, Lucknow, Uttar Pradesh 226033, India.
E-mail: drsujita@gmail.com

Submitted: 01 Mar 2024
Accepted: 03 Apr. 2024
Published Online: xxxx

 Sage



Copyright © The Author(s) 2024

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Sage and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

ACCESS THIS ARTICLE ONLINE
Website: journals.sagepub.com/home/szj
DOI: 10.1177/02537176241247934

student works, assignments, projects, and beyond. The AI-generated content may not be original, and there is a potential risk of plagiarism, which could have severe academic and legal consequences.⁴ Thus, implementing strategies for reducing plagiarism is essential to preventing plagiarism and protecting academic integrity in research and educational endeavors.^{4,5}

The development and implementation of AI-content detection tools highlight the increasing importance and need to differentiate between human-written and AI-generated content in various fields, including education. Many AI-detection tools divide the text into tokens, words, or other frequently occurring character sequences and estimate the likelihood that a given token will be followed by the next in the sequence. Texts with high predictability and low perplexity, with comparatively few random elements and human-generated idiosyncrasies, are the most likely to be recognized as AI-generated texts. Some AI text detectors use different techniques.⁶ Nevertheless, most publicly accessible detectors employ techniques based on perplexity and its associated notions.

AI-detector tools provide qualitative and quantitative assessments of the likelihood that a document is AI generated. A limited number of studies have thoroughly examined the ability of different AI-content detectors to distinguish between human- and artificial-generated content.

This study evaluated the sensitivity of free versions of popular AI-detection software programs in detecting AI-generated text.

Methods

We searched the Google database using the term “AI detector free.” A list of AI-detection software programs was displayed on the opened page. We selected the first 10 AI-detector software programs with a free version; those allowed a minimum of 500 words for text analysis in a single run. The investigators randomly selected a topic related to psychiatry and gave a specific command on the ChatGPT 3.5 version to generate text. The command was to write a brief scientific article on the “Role of Electroconvulsive Therapy (ECT) in Treatment-resistant Depression” within 500 words.” The text generated by the ChatGPT 3.5 version was then rephrased using three different software programs (the free version of QuillBot, the premium version of Grammarly, and the ChatGPT 3.5 version). The prompt for rephrasing the text content was “Kindly rephrase the following content.” After this prompt, the text generated initially by ChatGPT 3.5 was pasted. While paraphrasing in Grammarly premium generative AI, the “improve it” command was given, and in the QuillBot free version, the “standard” setting was kept. The identified AI-detection software programs analysed the primary AI-generated text and the

AI-paraphrased texts (**Supplementary file 1**). These software programs provide a probability percentage of the text generated by AI. The command was given on January 11, 2024, and the AI-generated content was analysed on January 12, 2024. 100% of the text was AI generated, as the investigator did it exclusively using AI. Analysing the content using AI-detection software programs revealed the degree to which these software programs could detect the AI content.

Ethical Approval

Formal ethical approval was not obtained as the investigators used free, publicly available software programs to analyse the AI-generated text.

Statistical Analysis

The data was presented in percentages by using descriptive statistics.

Results

10 popular AI-detection tools (free versions), including ZeroGPT, Copyleaks, QuillBot, GPTZero, Sapling, Undetectable AI, Content Detector AI, Wordtune, Dupli Checker, and Scispace, were compared for their sensitivity in detecting exclusively AI-generated text, and it was found that their sensitivity in detecting AI-generated content ranges from 0% to 100%. Five software programs detected AI-generated contents with 100% accuracy (**Table 1**),

TABLE 1.
Comparison of AI-detection Software.

Parameter	Copyleaks	QuillBot	ZeroGPT	GPTZero	Sapling	Undetectable AI	Content Detector AI	Wordtune	Dupli Checker	Scispace
• AI content	100%	100%	95.03%	97%	100%	100%	78.26%	100%	0%	61%
• Human content	0%	0%	4.97%	3%	0%	0%	21.74%	0%	100%	39%
% of AI content in paraphrased text										
• QuillBot	0	90	32.31	29	100	100	64.58	0	0.1	29
• Grammarly	100	100	96.74	38	100	100	75	100	0	42
• ChatGPT	100	100	90.61	97	100	100	60.87	100	0	26

Copyleaks: <https://copyleaks.com/ai-content-detector>
QuillBot: <https://quillbot.com/ai-content-detector>
ZeroGPT: <https://www.zerogpt.com/>
GPTZero: <https://gptzero.me/>
Sapling: <https://sapling.ai/ai-content-detector>
Undetectable AI: <https://undetectable.ai/>
Content Detector AI: <https://contentdetector.ai/>
Wordtune: <https://www.wordtune.com/ai-content-detector>
Dupli Checker: <https://www.duplichecker.com/ai-content-detector.php>
Scispace: <https://typeset.io/ai-detector>

whereas four detected contents with 50%–100% accuracy.

When the AI-assisted paraphrased texts were analysed, the detection accuracy of *QuillBot*-, *Grammarly*-, and *ChatGPT*-generated texts ranged from 0% to 100%. The *Sapling* and *Undetectable AI* software detected the paraphrased contents developed by all three software programs (*QuillBot*, *Grammarly*, and *ChatGPT*) with 100% accuracy. In contrast, *Copyleaks*, *QuillBot*, and *Wordtune* identified paraphrased content generated by two software programs (*Grammarly* and *ChatGPT*) with 100% accuracy.

Discussion

Our analysis utilized 10 different AI-content detectors and compared their performance effectively and correctly, pinning down the percentage of AI and human content. As discussed above, the results were contrasting, with *Copyleaks*, *QuillBot*, *Sapling*, *Undetectable AI*, and *Wordtune* precisely detecting content to be 100% AI generated and *Dupli Checker* inaccurately detecting content to be 100% human generated.

There are a few studies that have assessed the sensitivity of AI-content-detection tools. One such study, by Walters et al., evaluated the reliability of 16 publicly accessible AI content detectors to discriminate AI-generated from human-generated writing.⁷ They used 126 documents and divided them into three sets (42 generated using *ChatGPT-3.5* and *ChatGPT-4* each in the first week of April 2023 and 42 written without AI by first-year students for composition courses in 2014–2015). Their results indicate that three AI-content detectors, namely, *Copyleaks*, *Originality*, and *Turnitin*, could precisely differentiate among these three sets of documents. This finding is similar to our results, as *Copyleaks*, the AI detector common in both studies, accurately detected the AI content. Also, in line with our analysis, other common tools such as *ZeroGPT* and *GPTZero* could also discern the AI and human content, but with less accuracy than *Copyleaks*. The tool with contrasting findings was *Sapling*. It could identify the AI content with 100% precision in our analysis but had low overall accuracy in their study and appeared in the bottom two. Their study also suggested that the accuracy or precision of the detection tool is only moderately associated with its free or paid status.

Another perturbing fact in the context of AI-generated text the authors have discovered is that paraphrasing tools make it more challenging to discern between AI-generated and original content. For example, in a study, two essays were written by *ChatGPT* and then paraphrased using AI; the “real” percentage of applying the *GPT-2 Output Detector* markedly changed from 0.02% to 99.5% in essay one and from 61.96% to 99.8% in essay two.⁷ Similarly, the results of our analysis on rephrasing the *ChatGPT 3.5* version generated text using three different software programs (free *QuillBot*, premium version of *Grammarly*, and *ChatGPT 3.5* version). Only three AI-content detectors (*QuillBot*, *Sapling*, and *Undetectable AI*) could precisely detect the AI content even after rephrasing the content. The *Dupli Checker* still inaccurately measured the content as 0% AI generated. The text was rephrased using *QuillBot* and successfully tricked three tools (*Copyleaks*, *ZeroGPT*, and *Wordtune*).

Another study by Elkhatat et al. aimed to compare the potential of various AI-content-detection tools in differentiating human- and AI-authored content.⁸ They compared 15 *ChatGPT 3.5*- and *ChatGPT 4*-generated paragraphs and five control responses (human-written) on *OpenAI*, *Writer*, *Copyleaks*, *GPTZero*, and *CrossPlag*. Their results indicate that these tools more successfully identified *ChatGPT 3.5*-generated content than *ChatGPT 4*-generated content while inconsistently analysing human-written content. They suggested that these differences between the *ChatGPT 3.5* and *ChatGPT 4* results indicated that the detector performance could remarkably vary based on the AI model’s sophistication.

All the studies mentioned above, including our analysis, highlight the need for novel AI-detector tools that can produce homogenous results even after rephrasing and can be relied on for AI detection universally. Also, the fact that AI-text-generation capabilities are rapidly evolving calls our attention to the need for novel advancements in AI detection tools to keep pace with them.

These inaccuracies and inconsistencies in the results also raise concerns about the reliability of these AI-detection tools, especially in academic integrity investigations and other crucial contexts. While these tools serve as a practical and time-saving method

in discerning AI-generated content from the original, they should not be considered the sole determinant in academic integrity cases. Instead, the fairer approach in the evaluation process could be more holistic, including manual reviews and consideration of contextual factors.^{8,9} This study also gives insights to the researchers, editors, and other concerned individuals involved in checking the credibility of academic writings about the more sensitive and reliable free AI-detection software programs available.

Our study’s limitation is that it compares only 10 online software programs with free versions. It also compares a limited number of AI-generated and paraphrased contents, which can be another limitation of this study. The subscribed versions may be more sensitive in detecting AI-generated text, and future research may compare the subscription versions to elicit more insights into the AI-content-detection abilities of these software programs.

Conclusion

The use of AI in academic writing is increasing. However, it is essential to understand the limitations of using AI and its negative impacts on the author’s credibility. Scientific journals do not accept fully AI-generated content. The AI-content-detection software programs can successfully distinguish between AI- and human-generated texts. However, the predictive value of free AI-content-detection tools is highly variable. It is worth keeping in mind that the sensitivity of these tools may change over time with software upgrades. There is a need for periodic research in this area to give researchers and academicians insights into choosing appropriate software programs for writing assistance.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

Formal ethical approval was not obtained as the investigators used free, publicly available software programs to analyse the AI-generated text.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

ORCID iDs

Sujita Kumar Kar  <https://orcid.org/0000-0003-1107-3021>

Amit Singh  <https://orcid.org/0000-0001-8877-8869>

References

1. Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training. 2018. *Proceedings of 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*. https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aea-daeob7d230fa35c8afo4c041-Paper-Conference.pdf (accessed 14 February 2024)
2. Dweck CS. Motivational processes affecting learning. *Am Psychol* 1986; 41(10): 1040–1048. DOI: 10.1037/0003-066X.41.10.1040.
3. Fishman T. “We know it when we see it” is not good enough: toward a standard definition of plagiarism that transcends theft, fraud, and copyright. 2009. *4th Asia Pacific Conference on Educational Integrity (4APCEI)*, 28–30 September 2009. University of Wollongong NSW Australia. <https://www.bmartin.cc/pubs/09-4apcei/4apcei-Fishman.pdf> (accessed on 14 February 2024)
4. Foltýnek T, Meuschke N and Gipp B. Academic plagiarism detection: a systematic literature review. *ACM Comput Surv (CSUR)* 2019; 52(6): 1–42.
5. Alsallal M, Iqbal R, Amin S, et al. Intrinsic plagiarism detection using latent semantic indexing and stylometry. In: 2013 sixth international conference on developments in systems engineering, 2013, pp.145–150. DOI: 10.1109/DeSE.2013.34.
6. Crothers EN, Japkowicz N and Viktor HL. Machine-generated text: a comprehensive survey of threat models and detection methods. *IEEE Access* 2023; 11: 70977–71002. DOI: 10.1109/ACCESS.2023.3294090.
7. Anderson N, Belavy DL, Perle SM, 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 Exerc Med* 2023; 9(1): e001568. DOI: 10.1136/bmjsem-2023-001568.
8. Elkhatat AM, Elsaid K and Almeer S. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *Int J Educ Integr* 2023; 19(1): 1–16. DOI: 10.1007/s40979-023-00140-5.
9. Walters WH. The effectiveness of software designed to detect AI-generated writing: a comparison of 16 AI text detectors. *Open Inf Sc* 2023; 7(1). DOI: 10.1515/opis-2022-0158.