Acta Informatica Pragensia
2026, Volume 15, Issue 1, 126–134
https://doi.org/10.18267/j.aip.293

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Article Open Access

Evaluating AI Text Detection Tools for Distinguishing Human-Written from AI-Generated Abstracts in Persian-Language Journals of Library and Information Science

Amrollah Shamsi [1], **Ting Wang** [2], **Maryam Amraei** [3], **Narayanaswamy Vasantha Raju** [4]

[1] Clinical Research Development Center, The Persian Gulf Martyrs Hospital, Bushehr University of Medical Sciences, Bushehr, Iran
[2] School of Library and Information Management, Emporia State University, Emporia, Kansas, USA
[3] Department of Information Science and Epistemology, Shahid Chamran University, Ahvaz, Iran
[4] Department of Library and Information Science, Government First Grade College, Talakadu, India

Corresponding author: Ting Wang (twang2@g.emporia.edu)

Editorial Record

First submission received:
August 18, 2025

Revision received:
October 16, 2025

Accepted for publication:
October 17, 2025

Academic Editor:
Adela Jarolimkova
Charles University, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Shamsi, A., Wang, T., Amraei, M., & Raju, N. V. (2026). Evaluating AI Text Detection Tools for Distinguishing Human-Written from AI-Generated Abstracts in Persian-Language Journals of Library and Information Science. *Acta Informatica Pragensia*, 15(1), 126–134. https://doi.org/10.18267/j.aip.293

Abstract

Background: Researchers are using artificial intelligence (AI) tools in academic writing. However, their use may compromise the integrity and originality of the work. Hence, AI text detection tools have come to increase transparency.

Objective: This study aims to evaluate the accuracy of AI text detection tools in recognizing human-written and AI-written abstracts in library and information science (LIS).

Methods: Seven Persian academic journals in LIS were selected. ZeroGPT and GPTZero as AI text detectors were used. AI-generated abstracts were produced by AI chatbots (ChatGPT 4.0, DeepSeek and Qwen).

Results: Despite performing strongly in detecting AI-generated text, especially from models such as DeepSeek and Qwen, ZeroGPT and GPTZero struggle to accurately identify human-written content, resulting in high false positive rates and raising concerns about their reliability.

Conclusion: The findings highlight the need for culturally and linguistically inclusive AI detection tools, as current systems such as ZeroGPT and GPTZero show limitations in diverse language contexts, underscoring the importance of improved algorithms and human-involved evaluation to ensure fairness and reliability in academic settings.

Index Terms

Artificial intelligence; AI; AI detection; ZeroGPT; GPTZero; ChatGPT; DeepSeek; Qwen.

1 INTRODUCTION

With the birth of ChatGPT in November 2022, the use of large language models in scientific writing has rapidly expanded (Liang et al., 2024). This use can be for reasons such as improving grammar, structure, summarization, research productivity, improving the efficiency and quality of academic writing, idea generation, grant applications and peer review (Golan et al., 2023; Khalifa and Albadawy, 2024; Naddaf, 2025; Kobak et al., 2025). Also, the pressure of the "publish or perish" culture in scientific production (Siegel et al., 2018) can be a reason for the widespread acceptance of AI as it improves the speed of writing and the quality of articles (Huang and Tan, 2023).

Despite its significant benefits, the use of AI in research faces serious ethical and practical challenges. Data bias, lack of transparency, potential for plagiarism, ethical, moral, legal and academic integrity are among these concerns (Khalifa and Albadawy, 2024; Shah, 2024; Nguyen et al., 2023; Malik and Amjad, 2025). In addition, pressures such as the "publish or perish" culture can themselves lead to unethical practices in research (Else, 2023). These concerns arise at a time when many researchers are using AI without disclosure in their publications (Kwon, 2025) and surveys show that the quality and authenticity of AI-generated texts remain questionable (Wiley, 2025).

Such problems have led to the development of AI content detectors as an option to deal with these problems. These detectors analyse text to determine whether a human or an AI system wrote it. These tools use algorithms to analyse word usage patterns, sentence structure and meaning of the text, and compare the content with existing data from human and AI-generated texts (Chen, 2024). These tools are designed to maintain scientific integrity, prevent AI abuse and ensure the authenticity of the content (Rafiq and Qurat-ul-Ain, 2025). The importance of using detectors is highlighted when AI may produce unsupported content (Chelli et al., 2024), which compromises the integrity of science (Walters, 2023).

The importance of this issue has led researchers to address this area and test the performance of AI recognition tools. Kar et al. (2024) found that the sensitivity of free AI recognition tools in recognizing ChatGPT-3.5 texts varies between 0% and 100%. However, some tools, such as Sapling and Undetectable AI, were able to identify software-modified texts with 100% accuracy. Such accuracy has also been seen in other studies (Walters, 2023). On the other hand, some studies have shown that existing tools for recognizing texts generated by AI do not have the necessary accuracy and reliability and misclassify texts (Gotoman et al., 2025; Weber-Wulff et al., 2023). Subramaniam (2023), examining the performance of OpenAI's AI recognition tool in six languages, showed that this tool has linguistic biases and has low accuracy in recognizing non-English texts (such as Arabic and Hindi).

However, many of these studies have been conducted in the English language and the performance of AI recognition tools in other languages has not been fully tested. Some studies have shown that these tools have low accuracy in recognizing non-English texts and have linguistic biases (Weber-Wulff et al., 2023; Subramaniam, 2023), which is especially important in languages with different structures such as Persian.

Given that Iran is the second most active country in scientific production in the Middle East (SJR, 2025) and has almost two thousand Persian journals, and given the importance of the Persian language in achieving scientific authority (Amirarjmandi et al., 2022) and the lack of similar studies in the Persian language context, the need to evaluate the performance of identification tools as supportive strategies in maintaining scientific integrity and research ethics seems essential.

Therefore, this study aims to comparatively evaluate the accuracy, sensitivity and validity of AI text recognition tools in identifying Persian abstracts generated by AI as opposed to human abstracts. The findings of this study can play an important role in enhancing our understanding of the challenges and opportunities of using AI text recognition tools in Persian-speaking academic environments.

2 LITERATURE REVIEW

Recently, there has been considerable interest in research into AI detection tools and their performance in detecting and differentiating between AI-generated content and human-generated content in the educational sector and in the AI domain to experiment with various LLMs and their suitability in detecting AI-generated content in English and non-English languages.

Elkhatat et al. (2023) examined the capabilities of five AI content detectors, namely, OpenAI, Writer, Copyleaks, GPTZero and CrossPlag, in distinguishing between AI-generated content and human-written text. The study found that content generated using GPT-3.5 was accurately identified by many of the AI detection tools compared to the content generated by ChatGPT-4. In a similar study, Chaka (2024a) examined whether student essays had been authored by a generative artificial intelligence (GAI) or human-written employing 30 AI detection tools. The study found that of the 30 AI detection tools, only two (Copyleaks and Undetectable AI) accurately identified human-written text. In their comparative analysis, Kar et al. (2025) examined 10 AI detection applications for their accuracy in identifying AI-generated text. The study results show that Copyleaks, Quillbot, Sapling, Undetectable AI and Wordtune identified AI-generated content with 100% accuracy. The remaining five applications were between 50-100% accurate in detecting AI-generated content. In yet another comparative study, Gao et al. (2023) made an attempt to compare ChatGPT-generated abstract with an original abstract to find out that the AI-generated abstract could be easily detectable. The result shows that the "GPT-2 Output Detector" was highly effective in identifying AI-generated abstracts. Blinded human reviewers were also relatively successful in distinguishing between the AI-generated abstract and the original abstract. In a very recent study, Elek et al. (2025) carried out an evaluative study to examine the efficacy of the perplexity metric score and assess the effectiveness of AI detection tools in differentiating between human-authored and AI-generated abstracts in the field of radiology. There was no consistency in identifying AI-generated content by the AI detection tools employed in the study.

Akram (2024) conducted a quantitative analysis to quantify the presence of AI-generated content in preprint texts submitted to the arXiv preprint server. Thousands of preprints submitted to arXiv from various fields of study including computer science, physics and mathematics were examined for quantifying the AI-generated and human-written text. The study found that there was a considerable increase in AI-generated content after the public release of ChatGPT generative AI applications in 2022. In a similar study, Howard et al. (2024) investigated the accuracy of three AI content detectors (GPTZero 4, Originality.ai and Sapling, estimating the change in AI utilization and characteristics associated with AI content in scientific abstracts. The study findings revealed that abstracts submitted in 2023 had more AI-generated content compared to the previous year's submissions. Furthermore, the study also found that more mixed AI/human content was present in the abstracts submitted in 2023. Mese (2024) also found a growing presence of AI-generated content in Scopus Q1 Journals scientific article abstracts since 2022. Odri & Ji Yun Yoon (2023) made an attempt to examine the effectiveness of 11 AI detection tools (ZeroGPT, Originality, Writer, Copyleaks, Crossplag, GPTZero, Sapling, Contentatscale, Corrector, Writefull and Quill) in accurately identifying AI-generated content. The study found that many of the AI detection tools can be easily evaded with simple adversarial techniques.

Malik & Amjad (2025) examined the effectiveness of four AI detection tools (Turnitin, ZeroGPT, GPTZero and Writer AI) on text generated by three different LLMs (ChatGPT 3.5, Gemini and Perplexity). The study also employed different adversarial techniques (editing through Grammarly, paraphrasing through Quillbot and 10-20% editing by a human expert) to check the efficiency and accuracy in detecting AI-generated text. The study found that there were inconsistencies across all four AI detection tools in identifying AI-generated content. Apart from the above studies, there are a few other studies which have carried out a systematic review on AI detection tools and their effectiveness and accuracy in differentiating AI-generated content and human-written text. These studies provide a bird's-eye view of recent studies on AI detection tools and their effectiveness (Canyakan, 2025; Chaka, 2024b; Foltýnek et al., 2020). The studies that have been discussed in the review section largely used English language datasets; very few AI detection application studies have been reported on non-English language AI-generated content. For instance, Subramaniam (2023) examined the performance and effectiveness of AI detection tools in identifying non-English AI-generated text compared to English AI-generated content. The study found that the OpenAI Text Classifier AI detection tool has a low rate of accuracy in detecting text written in non-English languages. Alshammari et al. (2024) introduced a new AI text classifier to detect AI-generated content in the Arabic language. The study mainly focused on developing a specific Arabic AI text classifier to accurately identifying Arabic language human-generated text. The study found that AraELECTRA and XML-R models outperformed GPTZero and OpenAI Text Classifier in detecting human-written and AI-generated text. In a very recent paper, Sani et al. (2025) aimed at developing an AI detector tool for the Hausa language to differentiate human-generated and AI-generated content.

In this study, an attempt is made to examine the AI detection tools and their performance in identifying Persian language abstracts published in LIS journals. This study is also significant in terms of detecting the AI-generated content in LIS journals, which has not been explored much in scholarly literature.

3 METHODOLOGY

The purpose of this study is to investigate the ability of AI text detectors to distinguish human-generated abstracts from abstracts generated by artificial intelligence in Persian-language library and information science journals. They were identified and evaluated through previous AI text detection studies (Weber-Wulff et al., 2023). Finally, two tools, ZeroGPT (www.zerogpt.com) and GPTZero (www.gptzero.me), were selected for this study. There were three main reasons for this choice. Firstly, these tools are available for free. Due to international sanctions, Iranian researchers are often unable to purchase external services; therefore, using free tools is a more likely choice for them. Secondly, these tools have performed well in previous studies (Walters, 2023), which helped examine the most robust tools. Of course, those studies was conducted on English texts. Thirdly, both tools claimed to support the Persian language. However, it turned out that one of them, GPTZero, is still in the training and optimization phase for recognizing Persian texts, as announced on its website. However, the existence of even such a capability in a tool that can identify Persian texts well in the future was an incentive to choose these two tools. Both AI detectors report the probability that the text was written by artificial intelligence or a human as a percentage. In the GPTZero tool, in addition to these two items, it shows a mixed percentage, which indicates the percentage of simultaneous cooperation of artificial intelligence and humans.

The list of Persian journals in library and information science in Iran was identified from the Scientific Journals Portal (www.journals.msrt.ir) and limited to academic journals. From each journal, 20 articles from the period 2016—2017 were randomly selected. In recent years, due to the lack of artificial intelligence (such as ChatGPT) at that time, it was guaranteed that the articles were written by humans. Given that the journal *Academic Librarianship and Information Research* (online ISSN: 2783-4638) had no articles in the time period considered, it was excluded from the study, and finally, seven journals were examined (Table 1). To create abstracts, 10 random articles from each journal were selected as human-written abstracts. The other ten random articles were used to produce structured abstracts by AI chatbots (ChatGPT 4.0, DeepSeek and QWEN). The full text of the articles (without abstracts) was provided to these tools and 250-word structured abstracts were generated, including the purpose, method, findings and conclusion in adherence to the journal guidelines. Since the two AI tools DeepSeek and Qwen are available in Iran without any restrictions and have had a huge wave of use (especially DeepSeek), they were included in this study.

Table 1. *List of selected Persian LIS journals and sampling of articles from 2016–2017.*

Journal names	Online ISSN	Number of articles
Journal of Knowledge Retrieval and Semantic Systems	1795-2783	52
Digital and Smart Libraries Researches	2538-5356	57
Scientometrics Research Journal	2423-5563	34
Library and Information Science Research	3092-6130	71
Human Information Interaction	2423-7418	55
Sciences and Techniques of Information Management	2476-6534	53
Journal of Studies in Library and Information Science	2717-4093	33

This study adapts the classification accuracy framework from Weber-Wulff et al. (2023) to evaluate human-written and AI-generated texts (Table 2). Abstracts written by humans identified by ZeroGPT and GPTZero, with over 80% generated by AI, are classified as false positive (FP). On the contrary, AI-generated abstracts that are correctly identified with over 80% are classified as true positive (TP).

Table 2. *Classification accuracy scales for human-written and AI-generated texts.*

Human-written text, classified by the tool as:		
[100—80%) AI	False positive	FP
[80—60%) AI	Partially false positive	PFP
[60—40%) AI	Unclear	UNC
[40—20%) AI	Partially true negative	PTN
[20—0%] AI	True negative	TN
AI-written text, classified by the tool as:		
[100—80%) AI	True positive	TP
[80—60%) AI	Partially true positive	PTP
[60—40%) AI	Unclear	UNC
[40—20%) AI	Partially false negative	PFN
[20—0%] AI	False negative	FN

When partial classifications are excluded, accuracy is defined as the proportion of correctly classified cases (true positives and true negatives) out of the total number of cases.

Strict Accuracy (ACC_strict) = (TN+TP)/(TN+TP+FN+FP)

When a partially correct classification is regarded as incorrect, its calculation accuracy is as follows:

Accuracy – Partially Incorrect (ACC_part_incorrect) = (TN+TP)/(TN+PTN+TP+PTP+FN+PFN+FP+PFP+UNC)

In the second approach, partially correct evaluations were included and counted as correct ones. The formula accuracy computation is as follows:

Accuracy – Partially correct (ACC_part_correct) = (TN+TP+PTN+PTP)/(TN+PTN+TP+PTP+FN+PFN+FP+PFP+UNC)

A semi-binary scoring approach was used to capture the nuances of detection outcomes better. This method differentiates partially accurate results (such as PTN or PTP) from fully accurate or incorrect ones. Under this system, partial classifications receive 0.5 points, whereas fully correct results (TP or TN) are assigned the full value of 1.0. The following formula was applied to compute the semi-binary accuracy:

*Accuracy – Semi-binary (ACC_semi) = (TN+TP+0.5*PTN+0.5*PTP)/(TN+PTN+TP+PTP+FN+PFN+FP+PFP+UNC)*

4 RESULTS

Although the detection tools indicate a strong general performance (80.33% accuracy for ZeroGPT and 73.91% for GPTZero), their ability to make clear-cut decisions is more limited, with less than half of all cases classified with full confidence under a strict binary framework (Table 3). Even though the detection tools are highly sensitive to AI-generated text, especially from DeepSeek (ACC = 100%) and Qwen (ACC = 100%), they struggle to distinguish human writing (ZeroGPT ACC = 2.08%, GPTZero ACC = 0%), leading to a high rate of false positives. The misclassification poses concerns in contexts such as academic integrity and authorship verification. In addition, ZeroGPT provides higher confidence and more accurate classifications than GPTZero.

Table 3. *Accuracy of detection tools for ZeroGPT and GPTZero.*

Source	ZeroGPT ACC	ZeroGPT ACC_bin	GPTZero ACC	GPTZero ACC_bin
Overall	80.33%	47.31%	73.91%	6.14%
Human-written	2.08%	1.43%	0%	0%
ChatGPT 4.0	100%	81.43%	100%	7.14%
DeepSeek	100%	98.57%	100%	2.86%
Qwen	100%	92.86%	100%	14.29%

ZeroGPT shows strong performance in detecting AI-generated content, especially from DeepSeek and Qwen, and is more consistent overall. However, ZeroGPT and GPTZero exhibit severe limitations in accurately identifying human-written text, even under a scoring system that rewards partial matches (Table 4).

Table 4. *Accuracy of ZeroGPT and GPTZero (binary inclusive approach).*

Source	ACC_semi_bin ZeroGPT	ACC_semi_bin GPTZero
Overall	75.00%	50.54%
Human-written	1.43%	4.29%
ChatGPT 4.0	81.43%	68.57%
DeepSeek	98.57%	40.71%
Qwen	92.86%	67.14%

Under the semi-binary scoring approach, ZeroGPT outperforms GPTZero in identifying AI-generated content from ChatGPT, DeepSeek and Qwen. However, both tools fail to accurately identify human-written text, even when partial credit is given. These results reinforce concerns about the risk of false positives (Table 5).

Table 5. *Accuracy of ZeroGPT and GPTZero (semi-binary approach).*

Source	ACC_semi_bin ZeroGPT	ACC_semi_bin GPTZero
Overall	53.89%	28.34%
Human-written	1.43%	2.14%
ChatGPT 4.0	81.43%	35.71%
DeepSeek	98.57%	35.71%
Qwen	92.86%	40.71%

5 DISCUSSION

Since ChatGPT was released to the public at the end of 2022, its ability to perform various tasks, from text generation to problem-solving, has sparked intense social interest and attention (Dwivedi et al., 2023; Hua et al., 2024; Mannuru et al., 2023). The ease of use and accessibility of various generative AI platforms that emerged subsequently make them a valuable tool and a potential threat to traditional educational practices (Batta, 2024; Lindell & Utterberg Moden, 2025). As various institutions actively explore policies and test tools to guide the ethical application of AI and the academic community seeks originality and fairness, the demand for detection tools is becoming increasingly urgent, resulting in the emergence of systems such as Turnitin AI detector, ZeroGPT, GPTZero and Sapling AI detector (Ardito, 2024).

However, many studies discussing the use of multiple detection tools to test the accuracy of AI-generated and human-written content have highlighted that the accuracy of these detection tools is not reliable (Chaka, 2024b; Elkhatat et al., 2023; Legaspi et al., 2024). Building on the line of inquiry, the current study did not aim to evaluate a wide range of detection tools. Instead, it focused on two widely used platforms, ZeroGPT and GPTZero, to assess their ability to accurately detect content produced by multiple AI systems alongside human-written texts in Persian. The results affirmed and expanded on previous research findings that AI detectors demonstrated high accuracy in detecting content generated by AI (Elkhatat et al., 2023; Kar et al., 2024) and have shown improvements in detection performance over the last two years. However, the improvement has not extended to detecting human-written texts. On the contrary, the tools frequently misclassified human-written content as AI-generated.

Fraser et al. (2025) noted that the training data of many AI detection systems are predominantly in English, thereby limiting their effectiveness in cross-linguistic contexts. The current findings support the observation and point to a deeper issue. The accuracy rates for human-written Persian texts approached zero under strict classification, indicating structural limitations in the models. Persian academic writing, characterized by standardized structures, formulaic phrasing, and technical terminology, may exhibit statistical similarities to AI-generated patterns. Detectors trained largely on English corpora thus fail to capture these linguistic and stylistic conventions, resulting in systematic false positives.

The misclassifications weaken confidence in detection technologies and risk causing unfair suspicion, harm to reputation, and rejection of work by non-English speaking scholars. Relying too heavily on detection tools, especially in editorial or institutional settings, could push scholars further to the margins if their languages or academic fields differ from those used in the AI's training data, and could perpetuate an English-focused notion of how humans write. This can lead to epistemic injustice, where valid research from diverse linguistic and cultural backgrounds is deemed less valuable simply because it does not align with the writing patterns found in the AI's training data. The consequences may exacerbate inequalities in academic publishing and deter scholars from writing in their native languages. To maintain fairness, automated results should be viewed as just one piece of evidence, supported by human review, subject-matter expertise, and an understanding of the context.

Future detection systems need to incorporate more language training corpora, including a diverse range of languages beyond English. Also, collaboration among AI developers, linguists, and regional education and publishing institutions can help establish databases that are culturally and linguistically representative. Such initiatives should be accompanied by transparent policies in academic publishing that acknowledge the limitations of detection technologies and prevent their overuse as decisive evidence. Integrating technical improvements and policy safeguards into publishing workflows could help promote equity and accuracy in global academic evaluation.

6 CONCLUSION

This research highlights the challenges of utilizing AI detection tools in academic settings that encompass multiple languages. Although AI content detection tools like ZeroGPT and GPTZero offer practical solutions to ethical concerns, their limitations highlight the need for approaches that are more inclusive of diverse cultures and languages. As AI's role in academic writing continues to evolve, developing detection systems that are fair, reliable, and transparent will necessitate the use of more sophisticated algorithms and institutional policies that effectively combine human judgment with machine analysis.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: A.S.: Conceptualization, Writing – Original draft preparation. T.W.: Methodology, Data analysis, Writing – Reviewing and Editing, Supervision. M.A.: Data Curation, Writing – Original draft preparation. N.V.R.: Investigation, Writing – Original draft preparation.

Statement on the Use of Artificial Intelligence Tools: The authors used DeepSeek (DeepSeek-V3.2-Exp version) to assist with translation and linguistic structuring of this article.

Data Availability: The list of selected articles, the AI-generated abstracts, and the detection tool outputs constitute the processed dataset and are available from the corresponding author upon reasonable request.

REFERENCES

Akram, A. (2024). Quantitative Analysis of AI-Generated Texts in Academic Research: A Study of AI Presence in Arxiv Submissions using AI Detection Tool. arXiv:2403.13812. https://doi.org/10.48550/ARXIV.2403.13812

Alshammari, H., El-Sayed, A., & Elleithy, K. (2024). AI-Generated Text Detector for Arabic Language Using Encoder-Based Transformer Architecture. *Big Data and Cognitive Computing*, 8(3), Article 32. https://doi.org/10.3390/bdcc8030032

Amirarjmandi, Z., Navabakhsh, M., & Sarookhani, B. (2022). Scientific authority and eco-centric values in the comprehensive scientific map of the country. *Applied Issues in Islamic Education*, 7(3), 33–64. https://doi.org/10.52547/qaiie.7.3.33

Ardito, C. G. (2024). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025(182), 11–28. https://doi.org/10.1002/tl.20624

Batta, A. (2024). Transforming higher education through generative AI: Opportunity and challenges. *Paradigm*, 28(2), 241–243. https://doi.org/10.1177/09718907241286221

Bhattacharjee, A., & Liu, H. (2024). Fighting fire with fire: can ChatGPT detect AI-generated text?. *ACM SIGKDD Explorations Newsletter*, 25(2), 14–21. https://doi.org/10.1145/3655103.3655106

Canyakan, S. (2025). Comparative accuracy of AI-based plagiarism detection tools: An enhanced systematic review. *Journal of AI, Humanities, and New Ethics*, 1(1), 5–18.

Chaka, C. (2024a). Accuracy pecking order – How 30 AI detectors stack up in detecting generative artificial intelligence content in university English L1 and English L2 student essays. *Journal of Applied Learning & Teaching*, 7(1), 127–139. https://doi.org/10.37074/jalt.2024.7.1.33

Chaka, C. (2024b). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching*, 7(1), 115–126. https://doi.org/10.37074/jalt.2024.7.1.14

Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination Rates and reference Accuracy of CHATGPT and BARD for Systematic Reviews: Comparative analysis. *Journal of Medical Internet Research*, 26, e53164. https://doi.org/10.2196/53164

Chen, V. (2024). How AI content detectors work to spot AI. https://gptzero.me/news/how-ai-detectors-work

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. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). 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*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elek, A., Yildiz, H. S., Akca, B., Oren, N. C., & Gundogdu, B. (2025). Evaluating the Efficacy of Perplexity Scores in Distinguishing AI-Generated and Human-Written Abstracts. *Academic Radiology*, 32(4), 1785–1790. https://doi.org/10.1016/j.acra.2025.01.017

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1), Article 17. https://doi.org/10.1007/s40979-023-00140-5

Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, 613(7944), Article 423. https://doi.org/10.1038/d41586-023-00056-7

Foltýnek, T., Meuschke, N., & Gipp, B. (2020). Academic Plagiarism Detection: A Systematic Literature Review. *ACM Computing Surveys*, 52(6), 1–42. https://doi.org/10.1145/3345317

Fraser, K. C., Dawkins, H., & Kiritchenko, S. (2025). Detecting AI-generated text: Factors influencing detectability with current methods. *Journal of Artificial Intelligence Research*, 82, 2233–2278. https://doi.org/10.1613/jair.1.16665

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine*, 6(1), Article 75. https://doi.org/10.1038/s41746-023-00819-6

Golan, R., Reddy, R., Muthigi, A., & Ramasamy, R. (2023). Artificial intelligence in academic writing: a paradigm-shifting technological advance. *Nature reviews urology*, 20(6), 327–328. https://doi.org/10.1038/s41585-023-00746-x

Gotoman, J. E. J., Luna, H. L. T., Sangria, J. C. S., Santiago Jr., C. S., & Barbuco, D. D. (2025). Accuracy and reliability of AI-generated text detection tools: A literature review. *American Journal of IR 4.0 and Beyond*, 4(1), 1–9. https://doi.org/10.54536/ajirb.v4i1.3795

Howard, F. M., Li, A., Riffon, M. F., Garrett-Mayer, E., & Pearson, A. T. (2024). Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023. *JCO Clinical Cancer Informatics*, 8, e2400077. https://doi.org/10.1200/CCI.24.00077

Hua, S., Jin, S., & Jiang, S. (2024). The limitations and ethical considerations of ChatGPT. *Data intelligence*, 6(1), 201–239. https://doi.org/10.1162/dint_a_00243

Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research*, 13(4), 1148–1154.

Kar, S. K., Bansal, T., Modi, S., & Singh, A. (2025). 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*, 47(3), 275–278. https://doi.org/10.1177/02537176241247934

Khalifa, M. & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27), eadt3813. https://doi.org/10.1126/sciadv.adt3813

Kwon, D. (2025). Is it OK for AI to write science papers? Nature survey shows researchers are split. *Nature*, 641(8063), 574–578. https://doi.org/10.1038/d41586-025-01463-8

Legaspi, J. B., Licuben, R. J. O., Legaspi, E. A., & Tolentino, J. A. (2024). Comparing AI detectors: evaluating performance and efficiency. *International Journal of Science and Research Archive*, 12(2), 833–838. https://doi.org/10.30574/ijsra.2024.12.2.1276

Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., ... & Zou, J. Y. (2024). Mapping the increasing use of LLMs in scientific papers. arXiv:2404.01268. https://doi.org/10.48550/arXiv.2404.01268

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

Lindell, T. L., & Utterberg Modén, M. (2025). Conflicting motives: challenges of generative AI in education. *Learning, Media and Technology*, 50, 324–337. https://doi.org/10.1080/17439884.2024.2438933

Malik, M. A., & Amjad, A. I. (2025). AI vs AI: How effective are Turnitin, ZeroGPT, GPTZero, and Writer AI in detecting text generated by ChatGPT, Perplexity, and Gemini? *Journal of Applied Learning & Teaching*, 8(1), 91–101. https://doi.org/10.37074/jalt.2025.8.1.9

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*, *41*(3), 1036–1054. https://doi.org/10.1177/02666669231200628

Mese, I. (2024). Tracing the Footprints of AI in Radiology Literature: A Detailed Analysis of Journal Abstracts. *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, 196(8), 843–849. https://doi.org/10.1055/a-2224-9230

Naddaf, M. (2025). How are researchers using AI? Survey reveals pros and cons for science. *Nature*. https://doi.org/10.1038/d41586-025-00343-5

Nguyen, T. T., Hatua, A., & Sung, A. H. (2023). How to detect AI-generated texts?. In *2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, (pp. 464–471). IEEE. https://doi.org/10.1109/UEMCON59035.2023.10316132

Odri, G.-A., & Ji Yun Yoon, D. (2023). Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity. *Orthopaedics & Traumatology: Surgery & Research*, 109(8), 103706. https://doi.org/10.1016/j.otsr.2023.103706

Rafiq, S., & Qurat-ul-Ain, D. A. A. (2025). The role of AI detection tools in upholding academic integrity: An evaluation of their effectiveness. *Contemporary Journal of Social Science Review*, 3(1), 901–915.

Sani, B., Soy, A., Imam, S. H., Mustapha, A., Aliyu, L. J., Abdulmumin, I., Ahmad, I. S., & Muhammad, S. H. (2025). Who wrote this? Identifying machine vs human-generated text in Hausa. arXiv:2503.13101. https://doi.org/10.48550/arXiv.2503.13101

Shah, S. (2024). The role of artificial intelligence in research writing: A critical analysis. Journal of Universal College of Medical Sciences, 12(3), 1–2. https://doi.org/10.3126/jucms.v12i03.73724

Siegel, M. G., Brond, J. C., Rossi, M. J., & Lubowitz, J. H. (2018). “Publish or perish” promotes medical literature quantity over quality. *Arthroscopy*, 34(11), 2941–2942. https://doi.org/10.1016/j.arthro.2018.08.029

SJR. (2025). Scimago Journal & Country Rank. https://www.scimagojr.com/countryrank.php?region=Middle%20East

Subramaniam, R. (2023). Identifying Text Classification Failures in Multilingual AI-Generated Content. *International Journal of Artificial Intelligence & Applications*, 14(5), 57–63. https://doi.org/10.5121/ijaia.2023.14505

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. https://doi.org/10.1515/opis-2022-0158

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., ... & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), 1–39. https://doi.org/10.1007/s40979-023-00146-z

Wiley. (2025). An artificial intelligence study by Wiley. https://www.wiley.com/en-us/ai-study

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951