SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 1639-1652
Open Access. Online & Print

DOI: 10.5281/zenodo. 122.126132

AN INVESTIGATION OF TURNITIN AI'S CAPABILITY TO UNCOVER ACADEMIC MISCONDUCT

Amani Saif Aljabri[1], Fatma Khalifa Alghafri[2], Asia Sultana. Hasheenaa Bagem[3], Faisal Alsaidi[4]

Received: 20/11/2025
Accepted: 25/12/2025

ABSTRACT

This quantitative study examines Turnitin AI in detecting AI generated texts. It also examined how different popular AI tools (ChatGPT, Gemini, and Copilot) and text types (Original vs. Paraphrased) influence Turnitin's AI detection scores. The accuracy result was just over two thirds of the cases. Moreover, paraphrasing texts has a significant effect on the detection rate. The results raise concerns about academic dependency on Turnitin AI in academic institutions and recommend limiting it to the initial stage of evaluation if needed. The findings of this study corroborate and extend prior work (e.g., Weber-Wulff et al., 2023) with inferential statistics showing paraphrasing is the dominant factor reducing detection, while the originating AI model is not. It also recommends revisiting assessment methods to incorporate methods that are less vulnerable to AI misuse and reflect students' originality and critical thinking.

KEYWORDS: Turnitin AI, Artificial Intelligence, Academic misconduct, ChatGPT, Gemini, Copilot.

1. INTRODUCTION

With the rapid spread of artificial intelligence (AI) writing tools like OpenAI Chat GPT, Google Gemini and others, academic integrity has become a pressing concern of academic institutions, as these tools challenge traditional plagiarism detection methods. Weber-Wulff et al. (2023) emphasize that the advanced capabilities of large language models (LLMs), including Open AI Chat GPT, Google Gemini and others, demand urgent development for new verification tools as the conventional methods are outdated. In response, Turnitin, which is the leading plagiarism detection tool in academia, has introduced Turnitin AI which claims to detect AI generated text with "98% confidence" (Turnitin, 2023). Turnitin AI tool is currently utilized at the University of Technology and Applied Science - Al Musannah (UTASA) as the primary means of plagiarism detection. However, recent research has identified that there are certain significant limitations in its ability to detect AI-generated content that has been altered using AI paraphrasing tools. Cotton et al. (2023) notes that "current detection systems fail to account for AI-generated text that has been processed through secondary algorithms, creating a substantial loophole for academic misconduct". Additionally, as demonstrated in the following section, many studies assess detection of original AI prose, but systematic investigations of detectors' accuracy on AI-paraphrased versions are still scarce.

Therefore, this research aims to investigate Turnitin AI's capability at UTAS, to identify AI-produced content particularly focusing on AI paraphrased writing. The research seeks to fill existing research gaps as highlighted by Khalil and Er (2023) who found that while Turnitin demonstrates moderate success in flagging unmodified AI output, its accuracy varies significantly across disciplines and writing. Additionally, Dawson et al. (2023) "AI-paraphrased text evades detection in 62% of cases, raising serious concerns about the tool's reliability. Finally, Rogerson and McCarthy (2023) indicated a linear decrease in detection accuracy corresponding to the number of paraphrasing iterations applied to AI-generated content.

2. RESEARCH QUESTIONS

This research aims to evaluate the effectiveness of Turnitin AI in detecting various forms of AI-generated text, which is essential for upholding academic integrity. The following research questions will guide this study

1. How accurately does Turnitin AI detect text generated solely by AI writing tools?
2. How accurately does Turnitin AI detect text generated by AI writing tools and paraphrased by another AI?
3. What is the detection success rate when AI-generated text is processed through AI paraphrasing tools?

According to the pre-research survey, the top three AI writing tools used among UTASA PSC students are Open AI ChatGPT, Google Gemini, and Microsoft Copilot. This research aims to address the research questions utilizing these tools. The significance of this research extends beyond the evaluation of Turnitin as a technological tool. It highlights the fundamental principles of academic integrity, which may be at risk if AI generated texts are not accurately identified as pointed out by Lancaster (2023). The research findings will provide educators with data that can tell the development of policies related to AI and academic integrity, as well as provide insights for creating AI detection algorithms

3. LITERATURE REVIEW

3.1. Introduction

Academic integrity is increasingly threatened by the growing use of AI in academic writing, presenting significant challenges in maintaining ethical standards. This literature review seeks to investigate research on the effectiveness of Turnitin AI for plagiarism detection, particularly concerning AI-generated and paraphrased content. It focuses on studies related to AI-based plagiarism detection, the rising popularity of AI in content creation, and Turnitin AI's role in ensuring academic integrity. The review specifically examines both traditional and contemporary forms of academic misconduct, the difficulties in detecting AI-generated text, and the emergence of plagiarism detection tools. Additionally, it will discuss Turnitin AI's capabilities in differentiating between AI-generated and human-produced writing. The review highlights the broader implications of AI-facilitated academic misconduct, emphasizing the need for more advanced methods to identify and detect such issues while maintaining academic integrity.

3.2. Defining Academic Misconduct in The AI Era

Artificial intelligence (AI) is increasingly revolutionizing different fields, including education (Vasile, 2023). While AI has revolutionized several elements of education and research, it has also created significant challenges, such as students

creating assignments utilizing AI that are undetectable by traditional plagiarism detectors (Perkins, 2023; Tirumala & Mayhew, 2024).

Recent studies highlight the growing implementation of AI in education, particularly tools like ChatGPT. ChatGPT offers multiple benefits, including enhanced student engagement, collaboration, and accessibility. However, it has also raised new challenges, particularly regarding plagiarism and academic integrity (Wang & Cornely, 2023). Automated paraphrasing and AI-generated essays have begun to substitute conventional forms of plagiarism (Plata et al., 2023). Research shows that ChatGPT can generate complex, unique content that escapes traditional plagiarism detection software (Khalil & Er, 2023). According to Kronivets et al. (2024), while these tools may improve overall plagiarism detection, AI's dual nature raises significant concerns regarding academic integrity in higher education (Khatri & Karki, 2023).

Studies suggest that traditional plagiarism detection tools struggle to identify AI-driven text due to syntactic variations. However, AI detection tools can identify machine-generated content based on specific features (Santra & Majhi, 2023). As AI continues to develop, institutions will need to adapt their approaches to maintain academic integrity, emphasizing the distinction between AI-generated and human-generated content in plagiarism detection (Santra & Majhi, 2023; Khalil & Er, 2023).

3.3. The Evolution of Plagiarism Detection Tools

Plagiarism detection techniques have advanced significantly to address the rising issues related to academic misconduct, especially with easy access to resources on the internet (Naik et al., 2015). Previous approaches primarily emphasized text matching and the detection of directly copied materials (Patil & Bomanwar, 2016). These methods combine various techniques, including linguistic, semantic, and syntactic components, to identify different kinds of plagiarism (Kaur et al., 2020).

Despite these advancements, challenges remain in detecting complex forms of plagiarism, particularly extensive paraphrased material and translation obfuscation (Kaur et al., 2020). To assess the effectiveness of these methodologies, datasets like PAN-PC are typically used to enhance accuracy and capabilities (Kaur et al., 2020).

With the widespread use of large language models like GPT, concerns regarding academic integrity and plagiarism detection have escalated (Patit Paban Santra & Debasis Majhi, 2023). Traditional plagiarism detection tools struggle to identify AI-generated text due to structural and syntactic variations (Patit Paban Santra & Debasis Majhi, 2023). Although AI detection tools can identify machine-generated content, their accuracy and reliability are often challenged, particularly when applying advanced algorithms like GPT-4 (Ahmed M. Elkhatat et al., 2023; Debora Weber-Wulff et al., 2023). When examining human-written texts, these methods frequently yield false positives and exhibit poor performance against content obfuscation strategies (Ahmed M. Elkhatat et al., 2023; Debora Weber-Wulff et al., 2023).

To overcome the shortcomings of traditional models, Retrieval-Augmented Generation (RAG) has evolved as an approach to enhance contextual relevance and accuracy in text generation (Fnu Neha et al., 2024). As AI-generated content becomes more complex, further advancements and improvements in AI content identification techniques will be necessary.

3.4. The Role of AI Tools In Content Creation and Paraphrasing

Recent studies highlight the growing significance of AI-based paraphrasing tools like ChatGPT and QuillBot in content creation and academic writing (Reza Putri Angga et al., 2024). These AI tools enhance productivity and writing quality by improving brainstorming, organizing content, reviewing literature, and conducting statistical analyses (Mohamed Khalifa et al., n.d.). According to Thaweesak Chanpradit et al. (2024), students primarily employ synonym substitution and sentence restructuring strategies to enhance their writing; however, their competencies vary. Furthermore, Siti Latifah et al. (2024) and PM Ntsobi (2024) note that these tools not only promote writing proficiency but also provide linguistic assistance for non-native speakers. The AI paraphrasing tool QuillBot substantially aids in identifying errors, rephrasing, and expanding vocabulary while also addressing plagiarism and semantic limitations. Similarly, the AI-powered chatbot ChatGPT significantly supports idea generation, summarizing, paraphrasing, and proofreading texts (Alqadi et al., 2023).

A significant amount of research has been conducted to investigate the impact of AI-driven content modification on plagiarism detection. Existing paraphrasing methods are limited, making it particularly challenging to detect AI-generated content (Chao Zhou et al., 2022). Recent studies have explored different approaches for paraphrase

detection, comparing human-authored and machine-generated paraphrases (Becker et al., 2023; Zhou et al., 2022). The results indicate that human-authored paraphrases are currently more complex than those generated by machines (Becker et al., 2023). A comparative study of AI detectors and human reviewers found that, despite this complexity, some AI detectors and experienced peer reviewers can accurately detect AI-generated medical writing even after paraphrasing (Liu et al., 2024). One study on paraphrasing detection emphasizes the importance of semantic understanding and similarity, which are applied in various natural language processing (NLP) applications, such as summary analysis and plagiarism identification. The key findings indicate that while existing detection techniques are promising, improvements are needed, particularly in developing a broader range of databases and better patterns to enhance detection features (Zhou et al., 2022).

3.5. Effectiveness Of Turnitin AI in Detecting AI-Generated Content

Recent studies have examined the effectiveness of AI-generated content detectors, including Turnitin. In some studies, Turnitin AI demonstrated notable accuracy, as reported by Walters (2023) and Gosling, Ybarra, & Angulo (2024), where it successfully detected all AI-generated content without any false positives. However, other studies indicate that its performance varies, particularly when applying advanced prompting techniques. According to Perkins et al. (2023), Turnitin AI could identify only 54.8% of AI-generated content. Furthermore, Weber-Wulff, Anohina-Naumeca, Bjelobaba, et al. found that existing AI detection tools, including Turnitin, are unreliable in distinguishing between AI-generated and human-written material. Perkins et al. (2023) also noted that both AI-generated essays and human-written essays received the same grades. Additionally, Kost (2024) suggests that Turnitin's claim of a 98% confidence level may be exaggerated.

3.6. Limitations Of Plagiarism Detection Tools

Extensive research has been conducted to investigate the limitations of plagiarism detection tools. Bidermann and Raff (2022) demonstrated that advanced models like GPT-J can bypass detection tools such as MOSS (Measure of Software Similarity), suggesting that newer large language models (LLMs) are increasingly difficult to detect. Users often seek ways to test or circumvent AI detection. It has been found that existing AI-generated text detection tools, including Turnitin, are inaccurate and unreliable, particularly when content is obfuscated (Weber-Wulff, Anohina-Naumeca, Bjelobaba, et al., 2023). Moreover, these tools have limitations in detecting complex forms of plagiarism, such as extensive paraphrasing and the use of technical tricks, including replacing original characters with similar-looking characters from foreign alphabets (Mozgovoy, Kakkonen, & Cosma, 2010). Another emerging threat is that AI-based plagiarism detection tools can inadvertently facilitate a new type of plagiarism called "Aigiarism," which assists in paraphrasing copied text (Lannoy, 2023).

A study conducted before the widespread use of AI-generated text examined the limitations of Turnitin in detecting plagiarism, and its findings remain relevant today. According to Patel, Bakhtiyari, and Taghavi (2011), anti-plagiarism tools often ignore very short sentences and single words when checking for similarities. One trick to confuse the system involves replacing spaces with white-colored dots, making each word appear as a separate sentence. Another method is replacing words in a sentence with their synonyms to decrease the similarity percentage. A third trick for bypassing AI detection is using translation software to translate text through multiple languages before translating it back to the desired language. For example, an English text could be translated into Spanish, Swahili, Persian, Malay, French, and Chinese, and then back to English. Additionally, adding fake references can lead to lower plagiarism detection scores, as the forged authors, dates, books, and journals are not recognized by the tools.

3.7. Academic Integrity Policies in The AI Era

As AI develops rapidly, the legalization of its use in academia is not keeping pace. Perkins and Roe (2024) examined how academic integrity policies in higher education institutions (HEIs) worldwide have addressed students' use of new technology by analyzing 142 institutions. The study revealed a significant gap in these policies concerning the role and use of AI tools in academia, as well as a lack of specificity and clarity regarding acceptable and unacceptable uses of such tools. It is recommended that institutions remain vigilant and proactive by regularly updating their policies, providing ongoing training for academic staff, and restructuring assessment methods.

Marcel and Kang (2024) conducted a similar study focused on Canadian universities, which yielded comparable results, highlighting the absence of clear

guidelines or official policies. The study suggests that as instructors currently decide how to incorporate AI into their courses, institutional support is essential to guide ethical usage, particularly in academic writing. Institutions should share best practices, develop ethical frameworks, and encourage inclusive policymaking, while also fostering discussions on AI literacy and ethical writing to promote academic integrity.

3.8. Research Gap

The introduction of AI presents significant opportunities for research. While there is a substantial body of work comparing AI and human detection of AI-generated images, a notable gap exists in the comparison of AI detection versus human detection of AI-generated texts. Additionally, there is a lack of studies focused on the effectiveness of AI detection tools concerning the newer paraphrasing tools that continue to evolve. Furthermore, there is a pressing need for research aimed at improving the algorithms of AI detection tools to address issues of false positives and false negatives. Such research would enhance these systems, leading to better usability and fostering greater academic integrity.

4. METHODOLOGY

This study aims to assess the effectiveness of Turnitin AI in detecting various types of AI-generated text, which is important for maintaining academic integrity. The research will focus on three key questions: (1) How accurately does Turnitin AI identify text produced solely by AI writing tools? (2) How accurately does Turnitin AI detect text generated by AI tools that has been paraphrased by another AI? and (3) What is the detection success rate for AI-generated text that has been processed through AI paraphrasing tools?

Adopting a quantitative research approach, the study will utilize three AI writing tools that are OpenAI ChatGPT, Google Gemini, and Microsoft Copilot (free version). These tools were chosen based on their popularity among UTASA-PSC as determined through a brief online survey distributed to over 100 students. It was decided in this study to limit content generation to these three widely used AI tools to ensure controlled comparison and ecological validity; however, it is understood that this sampling constrains external validity, and the results may not generalize to other current or future models

Twelve distinct topics will be chosen for generating original text, resulting in a total of 36 unique samples. These original texts will then be paraphrased using the other two AI tools, creating an additional 72 samples. In total, the research will analyze 108 samples, combining both the original and paraphrased texts.

To evaluate the effectiveness of Turnitin AI, all 108 samples will be submitted to the Turnitin system provided by UTAS. The accuracy of Turnitin AI in detecting text generated by AI tools that has been paraphrased by another AI will be assessed using this system. Data analysis will involve the use of a t-test to compare detection rates across the different forms of AI-generated text and their paraphrased versions.

Ethically, this study presents minimal concerns as it does not involve human participants. However, it will ensure compliance with the terms of service and ethical guidelines associated with the AI tools used. Potential limitations of the study include the specific reliance on selected AI tools, which may constrain the generalizability of the findings to other detection systems. Furthermore, the effectiveness of Turnitin AI may vary based on the complexity of the generated texts and the nature of their paraphrasing.

5. FINDINGS

The study uses AI tools to generate academic essays and explored the capabilities and efficiency of Turnitin Al detection system in identifying these essays. Specifically, the research involved generating a series of 250-word essays on four academic writing types which are pros and cons, cause and effect, opinion, and problem and solution. Each essay is created by one fixed prompt aligning with PSC Level 4 GFP standards which also equals CEFR B1 level.

The study examined Turnitin AI capabilities of detecting AI written and AI paraphrased essays. The study tested a total of 108 essays on 12 different topics. Each topic had 9 essays. Three originally generated essays by Open AI ChatGPT, Google Gemini, and Microsoft Copilot. These essays were consequently paraphrased by other AI tools, generating 6 essays for the same topic.

After submitting these essays through Turnitin AI, 68.52% of the 108 writings were detected with 100% AI generated texts. The remaining 34 submissions reveal different data. Interestingly, Turnitin AI displayed “No Information Found” for 26 documents, “Wrong Result” for 7 other essays and 1 “Incomplete Detection”. All these anomalies were run for a second detection attempt.

Table1: Detection of All Essays: AI Written and AI Paraphrased Essays.

Detection result	Frequency	Percentage
100% AI	74	68.52%
No Detection Result	26	24.07%
Wrong Result	7	6.48%
Incomplete Detection	1	0.93%
Total	108	100%

Essays generated by AI without any paraphrasing:

A total of 36 essays generated by 3 AI tools; ChatGPT, Copilot and Gemini without any paraphrasing were submitted to Turnitin AI to check its accuracy in AI detection.

The detection outcome was:

Table2: Detection Of Essays Generated by AI Without Any Paraphrasing.

Detection result	Frequency	Percentage
100% AI	27	75%
No Detection Result	8	22%
Incomplete Detection	1	3.7%
Total	36	100%

Turnitin successfully detected three out of every four as 100% AI detected (75%). However, 8 essays (22%) returned no results, which means the system was unable to evaluate them. The detection of one essay (3.7%) was incomplete.

The results show that while Turnitin AI is generally effective in detecting AI generated essays, it is not fully reliable as around a quarter of the essays did not return results.

Essays generated by AI and Consequently Paraphrased by other AI tools:

A total of 71 AI paraphrased were submitted to Turnitin AI to evaluate its AI detection effectiveness. All of these texts were generated by ChatGPT, Copilot or Gemini and then paraphrased using one of the other two tools using cross-paraphrasing design. These texts appear to be less detectable.

The detection outcome was:

Table3: Detection of Essays Generated by AI And Consequently Paraphrased by Other AI Tools.

Detection result	frequency	percentage
100% AI	45	62.5%
No Detection Result	19	26.38%
Wrong Result	7	9.7%
Total	72	100%

Among the 7 texts that were not flagged as fully AI-generated and paraphrased, the scores were; 75% AI, 72% AI, 59% AI, 33% AI and 0% AI (two cases).

This indicates that Turnitin AI results tend to drop after paraphrasing. While the majority 62.5% were still flagged as 100% AI, the tool sometimes shows false negative results either by partially flagging them and giving them a low AI detection score (9.7%) or not flagging them at all (26.38%).

The results suggest that paraphrasing texts using AI tools can reduce their detectability using Turnitin AI.

The false negative rate (approximately 36%) raises concerns for academic institutions relying only on Turnitin AI results in evaluating academic integrity.

Second Detection Attempt:

The first AI detection attempt resulted in a total of 27 anomalies revealing 26 essays with no information and 1 incomplete detections test. These irregularities were tested with a second submission through Turnitin AI.

The second test showed that 80.77% remained the same with no detection results. Only 5 documents presented new data.

Table 4: Second Detection Test for Essays with No Results.

Detection result	Frequency	Percentage
100% AI	3	11.5%
No Detection Result	21	80.77%
Incomplete Detection	1	3.8%
Wrong Result	1	3.8%
Total	26	100%

Among the five new results, 3 texts were fully detected as 100% AI generated; however, one writing was not flagged as fully AI created with a percentage of 65% AI and another text was labeled as incomplete detection.

Interestingly, one document, that was previously presented as incomplete detections in the first testing attempt, is now identified as 100% AI generated script.

Is there any difference in results between the three AI tools ChatGPT, Copilot, and Gemini?

A two-way ANOVA was conducted to examine the effects of AI tools (ChatGPT, Gemini, Copilot) and text type (Original vs. Paraphrased) on Turnitin AI results.

The results showed no significant effect of the original AI tool used, F (2) =.301, p = .741, indicating that the choice of the original tool had no statistically significant impact on AI detection rates. Similarly, the interaction between original tool and version type was not statistically significant, *F* (2) = .341, *p* = .741, suggesting that the effect of the AI tool on Turnitin AI detection scores did not depend on whether the text was original or paraphrased.

However, there was a significant effect of version type (original/ paraphrased), F (1) = 3.084, p = .83, suggesting that Turnitin AI scores differed between original and paraphrased texts. See (Table 5)

Table 5: Two-Way ANOVA Results for Effects of AI Tool And Version Type on Turnitin AI Detection Scores.

source	df	F	p-value
Original tool	2	.301	.741
Version type (original/ paraphrased)	1	3.084	.083
Original tool*version type	2	.341	.741

A profile plot (Figure 1) showed that

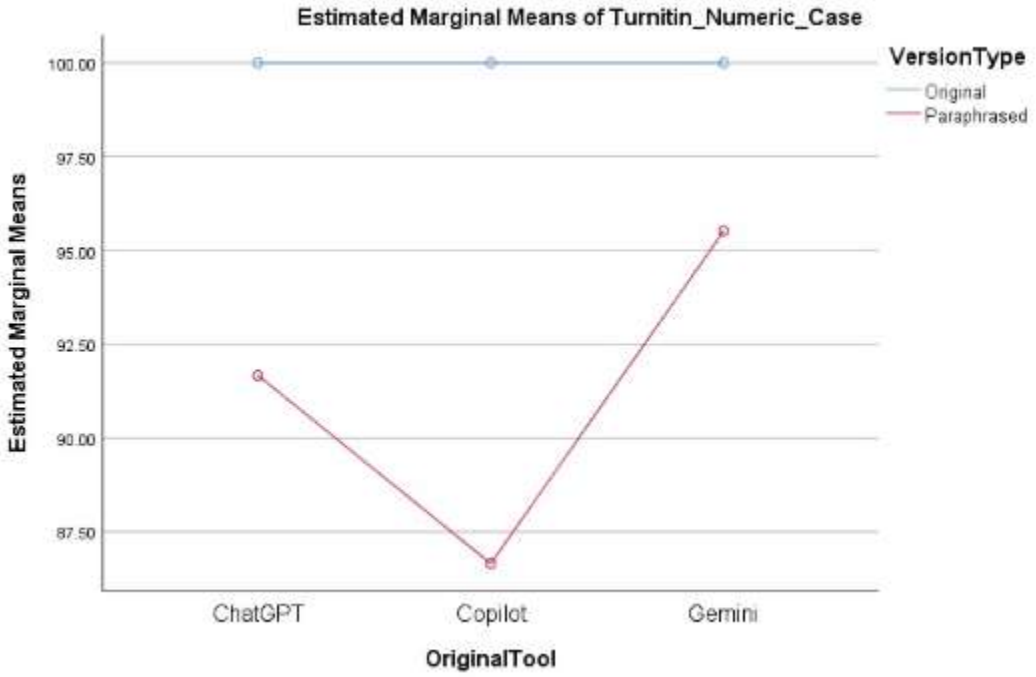


Figure 1: A Profile Plot of Estimated Marginal Means of Turnitin AI Results Based on AI Tools And Version Type.

There is an interaction between text type and AI tool. For original texts, all tools produced similarly high AI detection scores. However, for paraphrased texts, the detection scores varied substantially between the three tools. Copilot had the lowest detection, suggesting it was more effective at bypassing detection after paraphrasing, while Gemini maintained high detection levels. This visual pattern supports the statistically significant interaction effect found in the ANOVA results.

6. DISCUSSION

This study investigated the accuracy of Turnitin AI in detecting AI generated texts. It also examined how different AI tools (ChatGPT, Gemini, and Copilot) and text types (Original vs. Paraphrased) influence Turnitin's AI detection scores. The findings revealed that Turnitin AI was able to discover only

68.52% of the assignments as 100% AI while the rest of the assignments were given false negative results or no results at all. It was also shown that paraphrasing an AI generated text using AI lowers its chances of detection as the results dropped from 75% to 62.5%. Furthermore, Turnitin AI showed inability to generate AI detection report at 24.07% of assignments which persisted even after a second run of tests with a minor change which raises concerns about dependency on Turnitin AI to unleash academic misconduct. These findings align with previous research by Perkins et al. (2023) and Weber-Wulff, Anohina-Naumeca, Bjelobaba, et al. (2003)

The two-way ANOVA revealed that the choice of the original tool (Chat GPT, Gemini, Copilot) had no statistically significant impact on AI detection rates. Similarly, the interaction between the original AI tool and version type (AI generated, AI generated, and AI paraphrased) was not statistically significant. However, the two-way ANOVA confirmed there was a significant effect of version type, suggesting that Turnitin AI scores differed between original and paraphrased texts.

The current results suggest that Turnitin AI's detection capabilities are neither sufficient nor reliable to identify AI-generated content, especially when consequently paraphrased using AI tools. It also highlights the adaptability of generative AI tools in producing content that can bypass detection, which poses serious implications for academic integrity policies. Notably, the lack of significant differences across the AI tools (ChatGPT, Gemini, and Copilot) indicates that Turnitin AI does not differentiate between text generated by different generative AI tools. This raises the question of whether there is a unique stylistic or structural signature for each AI tool that can be detected?

To sum up, these findings highlight serious concerns about depending on Turnitin AI in schools and universities.

7. CONCLUSION

This study examined the accuracy of Turnitin AI in detecting AI generated texts using three tools; ChatGPT, Gemini, and Copilot. It also compared how Turnitin responds to original AI-generated text versus paraphrased versions. The results showed that Turnitin only identified over half of the AI-generated assignments as 100% AI, while the rest either produced wrong results or didn't produce any results at all. Paraphrasing the text with AI made it even harder for Turnitin AI to detect, lowering the detection rate. Further analysis showed that the type of AI tool used (ChatGPT, Gemini, or Copilot) didn't significantly affect Turnitin's detection results, and there was no meaningful interaction between the tool and the version of the text. However, there was a clear difference between original and paraphrased texts as paraphrased texts were less likely to be flagged.

In conclusion, the findings raise critical concerns about the reliance on Turnitin AI in educational settings. These tools can only provide a starting point for identifying academic misconduct. However, the limitation shown in detecting paraphrased AI content as well as the occasional failure to generate any report at all, highlights the need for using other approaches. While Turnitin AI is a time saving approach, sticking to the old way of using human judgment can complement the process by reviewing submissions more closely as trained educators can detect what detection tools miss. It is also advised that academic institutions revise their pedagogical practices by giving assignments that are hard to outsource to AI. This will include assignments that focus on critical thinking or reflections, multiple drafts written in class, assessments that track learning over time, and oral presentations. Moreover, there should be clear policies on the use of AI that define what is acceptable or unacceptable in using AI. It is also time to guide students on how to use AI responsibly and cite or disclose its use in their academic conduct.

8. RECOMMENDATIONS, LIMITATIONS AND FUTURE RESEARCH:

Given that Turnitin AI failed to detect a substantial amount of AI-written assignments, and was easily bypassed by AI paraphrasing, it might be more useful to serve only as a secondary, non-punitive check alongside human judgment. Furthermore, because the detection accuracy drops on paraphrased outputs, faculty could benefit from training to recognize AI-driven writing patterns and to adapt tasks, rubrics, and feedback practices accordingly.

This study has several limitations that should be acknowledged. First, it relied solely on quantitative data. While this approach provides measurable results, adding a qualitative or mixed method approach can provide deeper insights into both students and teachers' behavior in submitting/evaluating assignments. Second, the paraphrasing of texts was done entirely by AI while in reality, students often edit AI-generated content to make it appear more human produced. Including manual paraphrasing or other editing methods or tricks used by real students could improve

authenticity testing. Third, the study relied only on AI-generated content. Including a comparison with human written content can help determine how well Turnitin AI distinguishes between AI and genuine student writing. Fourth, the study focused on a limited set of AI tools, as well as one AI detection tool; Turnitin AI. Exploring less popular tools can enrich the results. Finally, the sample size of the topics and level of English language tested was limited. Expanding the number and variety of topics improves the findings and enables testing whether the effectiveness of detection can depend on the subject or on the complexity/simplicity of the language.

The above limitations can pave the way for future research, which can validate the results of this paper. It is worth noting that technology continues to advance, and the above results can change accordingly with the development of AI or AI detection tools.

REFERENCES

Alqadi, R., Alrbaiyan, A., Alrumayyan, N., Alqahtani, N., & Najjar, A. B. (2023, July). Exploring the user experience and the role of ChatGPT in the academic writing process. In 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE) (pp. 1082-1089). IEEE.

Angga, R. P., Jasmine, K., Agustine, S., Aryasatya, M., & Anggraini, N. D. (2024). Pemanfaatan Parafrase Berbasis Artificial Intelligence Sebagai Salah Satu Teknologi Digital Untuk Meningkatkan Efisiensi Penyelesaian Tugas Mahasiswa di Surabaya. Repeater: Publikasi Teknik Informatika dan Jaringan, 2(3), 35-42.

Becker, J., Wahle, J. P., Ruas, T., & Gipp, B. (2023). Paraphrase detection: Human vs. machine content. arXiv preprint arXiv:2303.13989.

Biderman, S., & Raff, E. (2022, October). Fooling MOSS detection with pretrained language models. In Proceedings of the 31st ACM international conference on information & knowledge management (pp. 2933-2943).

Chanpradit, T., Samran, P., Saengpinit, S., & Subkasin, P. (2024). English Paraphrasing Strategies and Levels of Proficiency of an AI-Generated QuillBot and Paraphrasing Tool: Case Study of Scientific Research Abstracts. Journal of English Teaching, 10(2), 110-126.

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z., & Chen, X. (2024). Research integrity in the era of artificial intelligence: Challenges and responses. Medicine, 103(27), e38811.

Chinengundu, T. (2023). Higher education assessment at crossroads on the use of ChatGPT. In Proceedings of the International Conference on Future of Education (Vol. 6, No. 1, pp. 87-98).

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in education and teaching international, 61(2), 228-239.

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), 17.

Foltynek, T., Bjelobaba, S., Glendinning, I., Khan, Z. R., Santos, R., Pavletic, P., & Kravjar, J. (2023). ENAI Recommendations on the ethical use of Artificial Intelligence in Education. International Journal for Educational Integrity, 19(1), 1-

Gosling, S. D., Ybarra, K., & Angulo, S. K. (2024). A widely used Generative-AI detector yields zero false positives. Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport, 42(2), 31-43.

Kalbhor, P., Patil, G., Agarwal, S., Rajput, A. S., & Dhamdhere, P. (2021). Research on paraphrase identification. International Research Journal of Engineering and Technology (IRJET), 8(5), 4050-4055.

Kaur, M., Gupta, V., & Kaur, R. (2021). Review of recent plagiarism detection techniques and their performance comparison. In Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2020 (pp. 157-170). Springer Singapore.

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. Computer Methods and Programs in Biomedicine Update, 100145.

Khalil, M., & Er, E. (2023, June). Will chatgpt g et you caught? rethinking of plagiarism detection. In International conference on human-computer interaction (pp. 475-487). Cham: Springer Nature Switzerland.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. Computers and Education: Artificial Intelligence, 5, 100156.

Kost, C. (2024). A comparative study of the accuracy of Turn-It-In's artificial intelligence detector in CTU doctoral assignments. The Pinnacle, 2(1), 1-12.
Kumar, R. (2023). Faculty members' use of artificial intelligence to grade student papers: A case of implications. International Journal for Educational Integrity, 19(1),
Lannoy, V. (2023). AI-based plagiarism detectors: Plagiarism fighters or makers?. Medical Writing, 32, 44-47.
Latifah, S., Muth'im, A., & Nasrullah, N. (2024). The Use of QuillBot in Academic Writing: A Systematic Literature Review. Journey: Journal of English Language and Pedagogy, 7(1), 110-121.
Marcel, F., & Kang, P. (2024). Examining AI Guidelines in Canadian Universities: Implications on Academic Integrity in Academic Writing. Discourse and Writing/Rédactologie, 34, 93-126.
Mozgovoy, M., Kakkonen, T., & Cosma, G. (2010). Automatic student plagiarism detection: future perspectives. Journal of Educational Computing Research, 43(4), 511-531.
Naik, R. R., Landge, M. B., & Mahender, C. N. (2015). A review on plagiarism detection tools. International Journal of Computer Applications, 125(11), 16-22.
Neha, F., Bhati, D., Shukla, D. K., Guercio, A., & Ward, B. (2025, January). Exploring ai text generation, retrieval-augmented generation, and detection technologies: a comprehensive overview. In 2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 00633-00639). IEEE.
Patel, A., Bakhtiyari, K., & Taghavi, M. (2011). Evaluation of cheating detection methods in academic writings. Library Hi Tech, 29(4), 623-640.
Patil, A., & Bomanwar, N. (2016). Survey on different plagiarism detection tools and Software's. International Journal of Computer Science and Information Technologies (IJCSIT), 7(5), 2191-2193.
Perkins, M., & Roe, J. (2024). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. Higher Education Policy, 37(3), 633-653.
Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). Game of Tones: Faculty detection of GPT-4 generated content in university assessments. arXiv preprint arXiv:2305.18081.
Plata, S., De Guzman, M. A., & Quesada, A. (2023). Emerging research and policy themes on academic integrity in the age of chat GPT and generative AI. Asian Journal of University Education, 19(4), 743-758.
Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-generated text be reliably detected? arXiv preprint arXiv:2303.11156.
Santra, P. P., & Majhi, D. (2023). Scholarly communication and machine-generated text: is it finally ai vs ai in plagiarism detection? Journal of Information and Knowledge, 175-183.
Tirumala, S. S., & Mayhew, C. A Chitchat on Using ChatGPT for Cheating.
Vasile, C. (2023). Artificial intelligence in education: Transformative potentials and ethical considerations. Journal of Educational Sciences and Psychology, 13(75), 2.
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.
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.
Zhou, C., Qiu, C., Liang, L., & Acuna, D. E. (2025). Paraphrase identification with deep learning: A review of datasets and methods. IEEE Access.

APPENDICES

1- Anova Table

Univariate Analysis of Variance

Between-Subjects Factors

		N
OriginalTool	ChatGPT	20
	Copilot	29
	Gemini	32
VersionType	Original	28
	Paraphrased	53

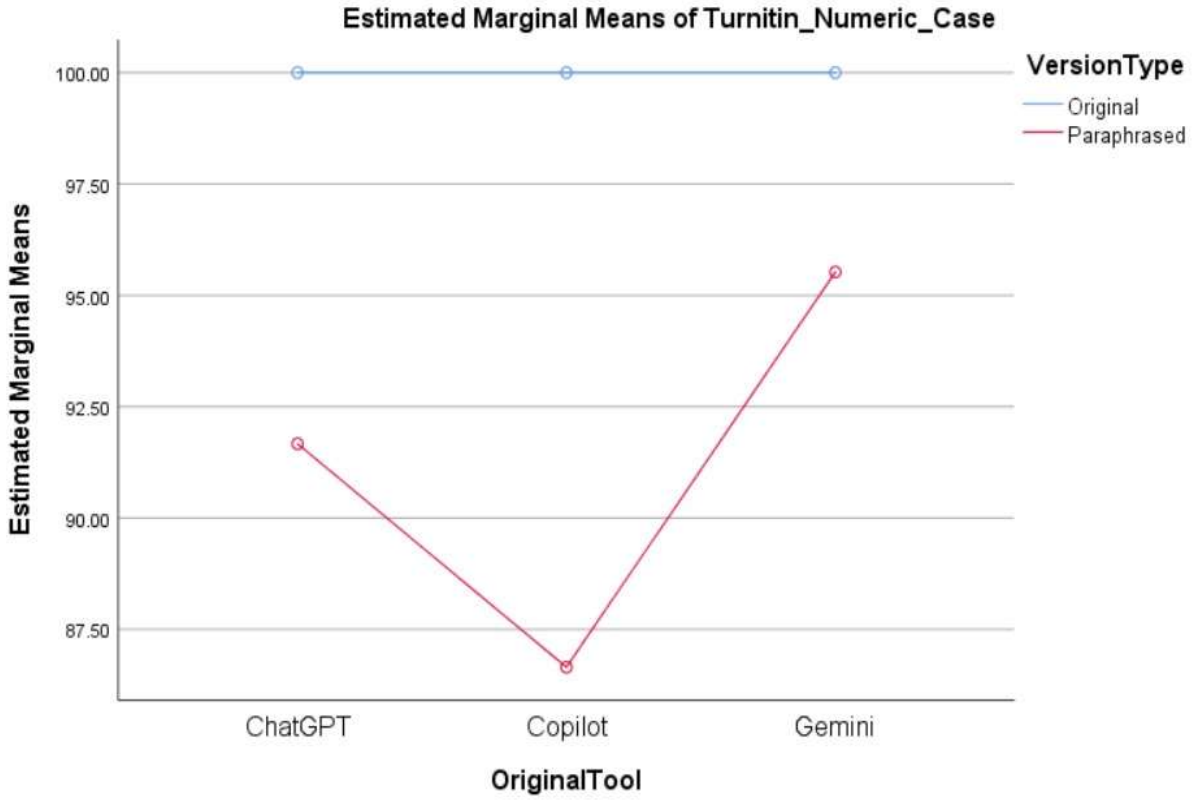
Tests of Between-Subjects Effects

Dependent Variable: Turnitin_Numeric_Case

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2194.829[a]	5	438.966	1.005	.421
Intercept	648249.341	1	648249.341	1484.067	.000
OriginalTool	263.011	2	131.505	.301	.741
VersionType	1347.161	1	1347.161	3.084	.083
OriginalTool * VersionType	263.011	2	131.505	.301	.741
Error	32760.455	75	436.806		
Total	755379.000	81			
Corrected Total	34955.284	80			

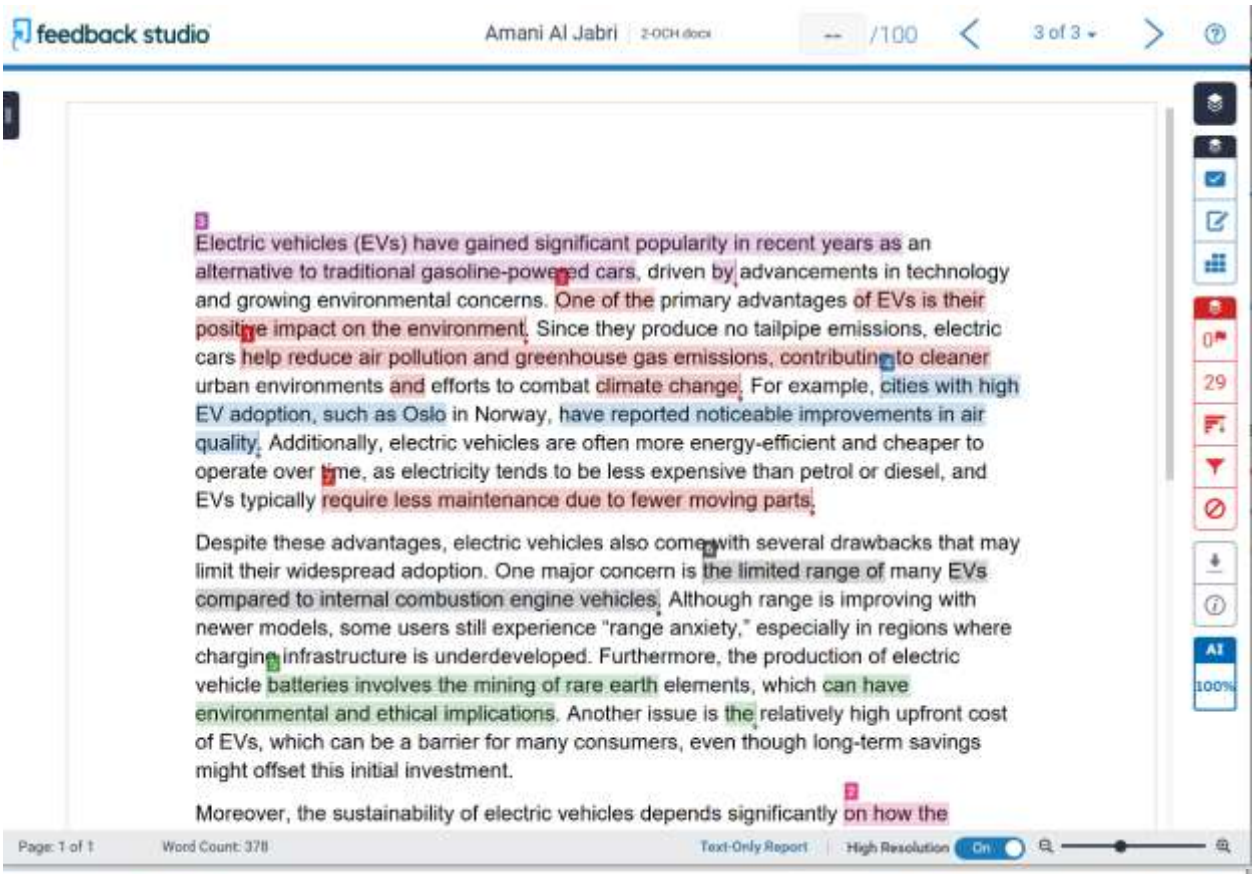
a. R Squared = .063 (Adjusted R Squared = .000)

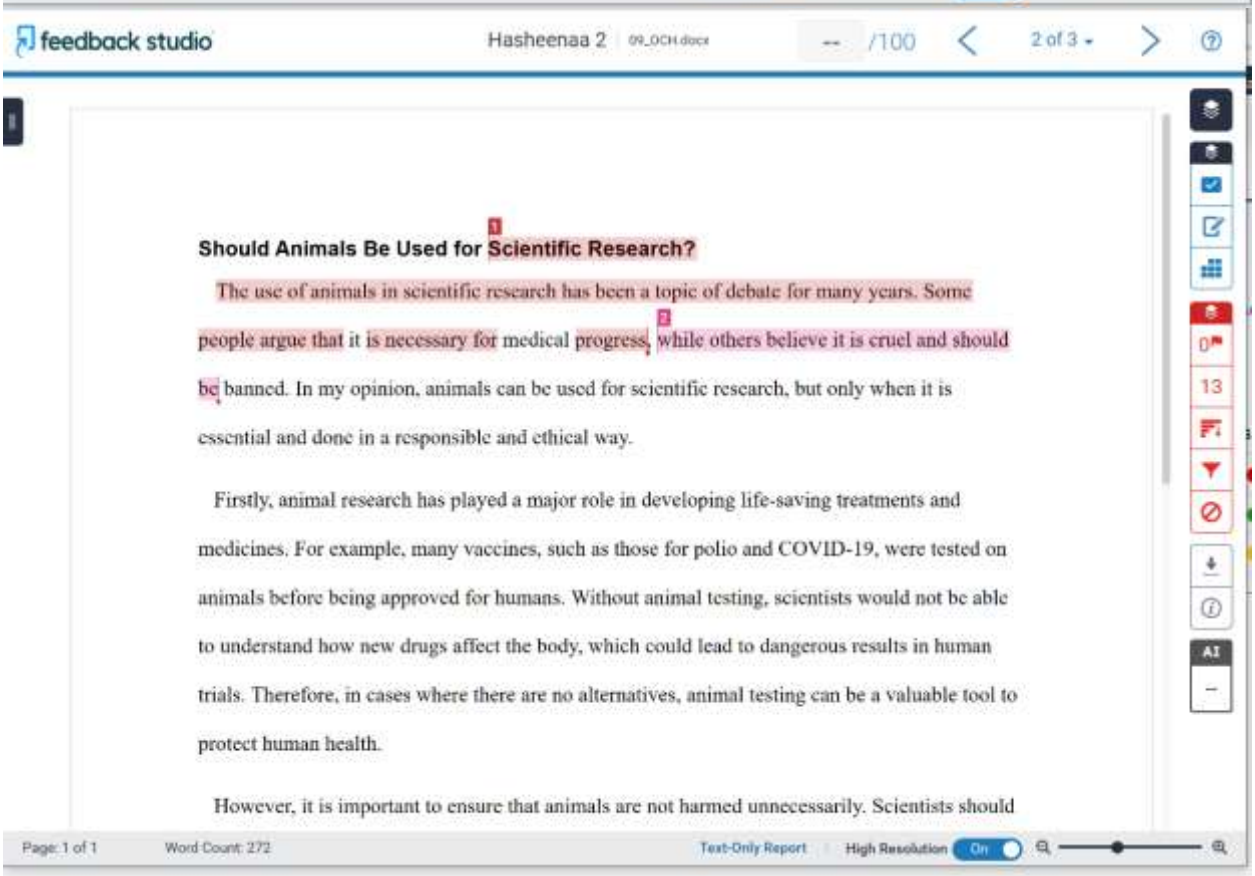
Profile Plots

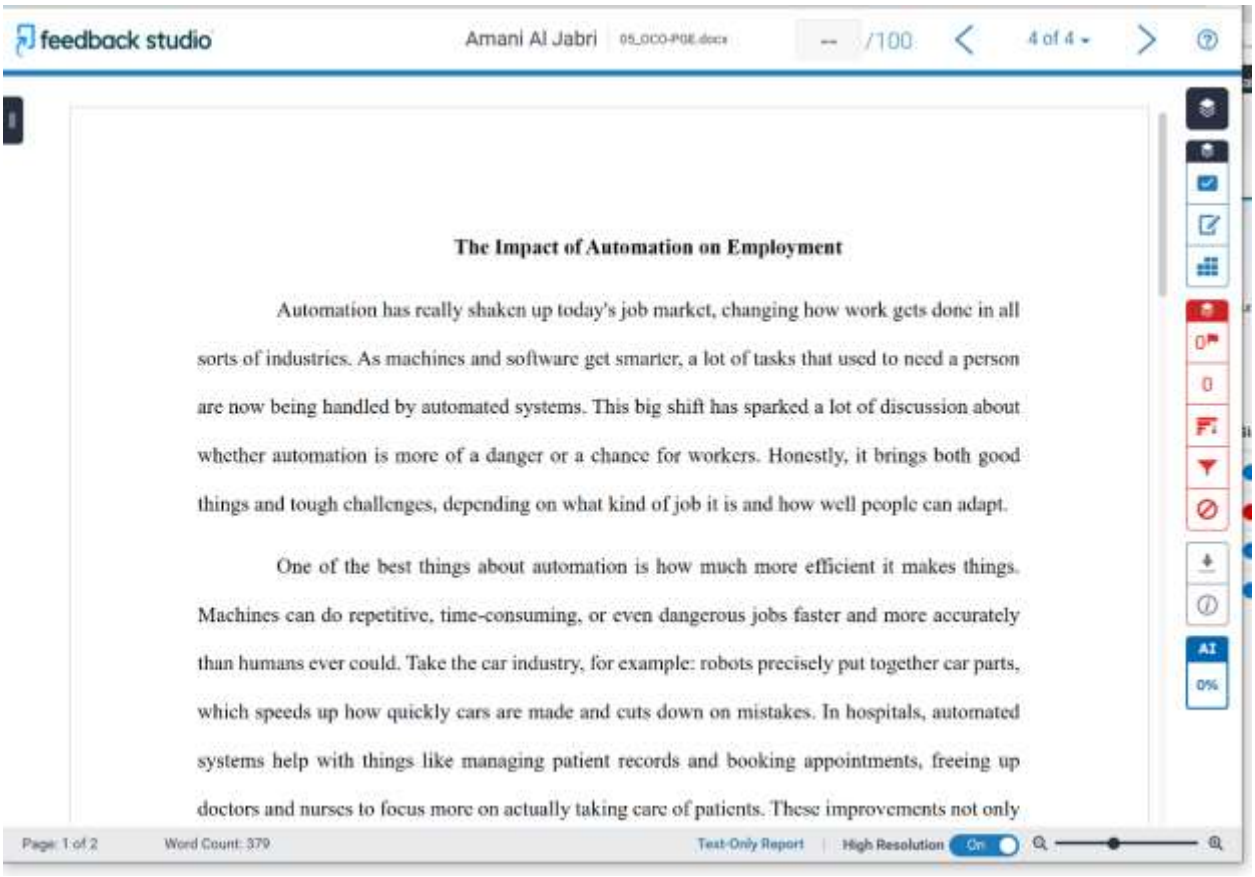


3-

Sample Turnitin Ai Reports


feedback studio
Amani Al Jabri
-- /100
3 of 3
Electric vehicles (EVs) have gained significant popularity in recent years as an alternative to traditional gasoline-powered cars, driven by advancements in technology and growing environmental concerns. One of the primary advantages of EVs is their positive impact on the environment. Since they produce no tailpipe emissions, electric cars help reduce air pollution and greenhouse gas emissions, contributing to cleaner urban environments and efforts to combat climate change. For example, cities with high EV adoption, such as Oslo in Norway, have reported noticeable improvements in air quality. Additionally, electric vehicles are often more energy-efficient and cheaper to operate over time, as electricity tends to be less expensive than petrol or diesel, and EVs typically require less maintenance due to fewer moving parts.
Despite these advantages, electric vehicles also come with several drawbacks that may limit their widespread adoption. One major concern is the limited range of many EVs compared to internal combustion engine vehicles. Although range is improving with newer models, some users still experience "range anxiety," especially in regions where charging infrastructure is underdeveloped. Furthermore, the production of electric vehicle batteries involves the mining of rare earth elements, which can have environmental and ethical implications. Another issue is the relatively high upfront cost of EVs, which can be a barrier for many consumers, even though long-term savings might offset this initial investment.
Moreover, the sustainability of electric vehicles depends significantly on how the
Page: 1 of 1
Word Count: 378
Text-Only Report
High Resolution
On
AI
100%


feedback studio
Hasheenaa 2
-- /100
2 of 3
Should Animals Be Used for Scientific Research?
The use of animals in scientific research has been a topic of debate for many years. Some people argue that it is necessary for medical progress, while others believe it is cruel and should be banned. In my opinion, animals can be used for scientific research, but only when it is essential and done in a responsible and ethical way.
Firstly, animal research has played a major role in developing life-saving treatments and medicines. For example, many vaccines, such as those for polio and COVID-19, were tested on animals before being approved for humans. Without animal testing, scientists would not be able to understand how new drugs affect the body, which could lead to dangerous results in human trials. Therefore, in cases where there are no alternatives, animal testing can be a valuable tool to protect human health.
However, it is important to ensure that animals are not harmed unnecessarily. Scientists should
Page: 1 of 1
Word Count: 272
Text-Only Report
High Resolution
On
AI


feedback studio
Amani Al Jabri
-- /100
4 of 4
The Impact of Automation on Employment
Automation has really shaken up today's job market, changing how work gets done in all sorts of industries. As machines and software get smarter, a lot of tasks that used to need a person are now being handled by automated systems. This big shift has sparked a lot of discussion about whether automation is more of a danger or a chance for workers. Honestly, it brings both good things and tough challenges, depending on what kind of job it is and how well people can adapt.
One of the best things about automation is how much more efficient it makes things. Machines can do repetitive, time-consuming, or even dangerous jobs faster and more accurately than humans ever could. Take the car industry, for example: robots precisely put together car parts, which speeds up how quickly cars are made and cuts down on mistakes. In hospitals, automated systems help with things like managing patient records and booking appointments, freeing up doctors and nurses to focus more on actually taking care of patients. These improvements not only
Page: 1 of 2
Word Count: 379
Text-Only Report
High Resolution
On
AI
0%