COLLEGE LIBRARIES
Vol. 39 No. 04 December 2024 pp. 3-11

Experimental Insights of Similarity Detection Tools in Recognizing Contents generated through Chatsonic, an Artificial Intelligence Tool

Kuheli Dutta
Assistant Librarian, Presidency University, Kolkata, West Bengal
Prof. (Dr.) Arun Kumar Chakraborty*
Professor and Head, Department of Library and Information Science, NSOU, Kolkata, West Bengal
(*Corresponding Author)

Abstract

The present paper is an original experimental study that aims to evaluate the efficiency of three similarity detection tools - Drillbit, Grammarly, and iThenticate- and their basic features and capability in detecting contents created through an artificial intelligence tool- Chatsonic. Contents were created in three different languages- Bengali, Hindi, and English through the Chatsonic Artificial Intelligence tool. Then, these contents were checked using the three similarity detection tools - Drillbit, Grammarly, and iThenticate.The study shows the extent of similarity and AI detection varies from one tool to another. The situation is a matter of concern for documents related to regional Indian languages. However, the point to be noted is that similarity detection tools like Grammarly and DrillBit can detect AI-generated content that is not humanized. DrillBit is good for checking similarities in Indian regional languages to some extent. Grammarly is good for students who are novices in writing. For publishing journal articles with standard publishers, iThenticate can be considered due to its wide quality data coverage for similarity checks and global acceptance. The present paper may assist in giving a deep thought about whether the AI detection feature of Similarity Detection Tools should be given importance while subscribing to the resources or not.

Keywords: Artificial Intelligence, Chatsonic, DrillBit, Grammarly, iThenticate, Plagiarism, Plagiarism checker, Similarity Detection Tools.

1. Introduction:

Academic Integrity is a matter of concern nowadays. To maintain academic integrity and to reduce the possibility of plagiarism in academic content, many active measures are being taken constantly by the University Grants Commission (UGC), India. Checking academic content, especially Ph. D. and Dissertations and Ph. D. related research articles, has been made mandatory through a similarity detection tool also known as a Plagiarism detection tool. Checking journal articles through similarity detection tools is also mandatory and mostly done through Turnitin or iThenticate similarity detection tools by renowned publishers. Thus, the role of these similarity detection tools is noticeable.

As part of the ShodhShuddhi Program, INFLIBNET Centre is offering DrillBit-Extreme

"Plagiarism Detection Software" to Higher Education Institutions (HEIs) w.e.f. October 1, 2023, Shodhshuddhi. (n.d.). DrillBitis is a proprietary Cloud-based global plagiarism checker developed by DrillBit Soft Tech. DrillBit senses the similarities and irregularities in text formatting and manipulations. It supports 18 Indian regional languages and different formats of documents to be uploaded to check plagiarism using OCR Technology. It also claims AI Tools Generated Text Detection (DrillBit, n.d.).

Grammarly is a writing assistant tool that takes AI help (optional) to improve writing quality with proper grammar or reference-related suggestions. In addition to that, it has features to check similarity and AI-generated content (Grammarly, n.d.).

iThenticate is a premier software developed by Turnitin Solutions to stimulate academic integrity using cutting-edge technology to prevent plagiarism worldwide. It covers resources in many languages, including English, but Indian regional language coverage is not at all satisfactory (iThenticate, n.d.).

As DrillBit, Grammarly, and iThenticateare well-known similarity detection tools, the paper intends to evaluate their efficiency by creating a few pieces of content with the help of Chatsonic - an artificial intelligent content creator. Chatsonic is a part of the Writesonic Artificial Intelligent tool (Writesonic, n.d.).

The authors of this paper felt the need to evaluate the efficiency of the similarity detection software in AI-generated content detection as the use of AI is increasing day by day. Authors also felt that if any additional costs are to be paid how cost-effective will it be? ChatGPT is well known but at the same time, many other AI tools generate content. Thus authors thought of trying to generate content through the Chatsonic AI tool and check the efficiency of the similarity detection software.

2. Literature Review

Dehouche (2021) generated three types of content (academic essay, talk, and opinion piece) via AI Dungeon using GPT-3. He expressed that there might be increasing scientific misconduct and thus publishing standards require immediate revision. Sabeeh and Khaled (2021) provided examples of different types of plagiarism/similarity detection tools. Text and source-code plagiarism were highlighted. Gao, et al. (2023) emphasized limiting the borders of virtuous and tolerable use of large language models to assist in scientific writing. King (2023) tried to use ChatGPT to create some content and also asked ChatGPT regarding the chances of plagiarism in higher education. It was found that the ChatGPT could understand and reply to complex requests. Khalil and Er (2023) expressed their deep concern regarding the unethical use of chatbots like ChatGPT by providing students with a convenient source to easily produce academic content without using their own academic and creative talent. Kleebayoon and Wiwanitkit (2023) advised that students and academicians must be made aware of how to use AI to avoid plagiarism ethically. Jarrah et al. (2023) searched databases like Google Scholar, ProQuest, Scopus, and Science Direct with specific keywords related to ChatGPT in academic research, education, publishing, and ethical challenges. Patra, S. K., &Kirtania, D. K. (2023) through their study observed that most of the content created by ChatGPT is relatively less in the similarity index. Many articles are there on ChatGPT and its efficiency in generating content, which showed some concerns about revising the publication standards after implementing AI tools. Though there were studies on ChatGPT, there are limited studies on other AI tools for academic writing. Further, the efficiency of similarity detection tools to detect AI-generated content through other AI tools like Chatsonic, etc., was not done yet. In this context, this study was taken

to analyze the effectiveness of two similarity detection tools - Drillbit, Grammarly, and iThenticate against the content generated through the Chatsonic-AI tool.

3. Objective of the Study

The objective of the study was to evaluate the efficiency of the DrillBit, Grammarly, and iThenticate similarity detection tools in checking and detecting the contents created through Chatsonic - An Intelligent Artificial Intelligence (AI) Chatbot. Here, one point is to be mentioned that the detection of AI-generated text was given more importance than similarity detection as the contents were created solely through AI tools.

4. Methodology

An account was created for free in the Chatsonic intelligent AI chatbot, a part of the Writesonic AI tool. Then, ten topics were created in different languages with proper prompts. Thus, the contents were copied in a notepad to retain them only as text and copied to different word files. The total number of contents created is ten (2 in Bengali, 2 in Hindi, and 6 in English).

Settings for both the similarity detection software were kept the same.Bibliography, references, quotes, phrases, and small sources were not excluded in the settings part for similarity detection. All repositories were selected for checking.

5. Scope and Limitations of the Study

Only DrillBit, Grammarly, and iThenticate software were considered plagiarism checkers or similarity detection tools in this study as the authors had access to only these three well-known tools. iThenticat is a globally accepted similarity detection software. DrillBit is recommended by the UGC and INFLIBNET for plagiarism checking and Grammarly is also well known in academic writings. The study was based on only 10 documents in English, Bengali, and Hindi languages created through ChatSonic-an, an Artificial Intelligence ChatBot. English, Bengali, and Hindi language were selected as the authors faced higher demands in similarity detection in these languages at their workplace in West Bengal.

6. Data Analysis and Interpretations

A total of ten documents were created with the help of a Chatsonic artificial intelligent content creator with different topic prompts and in three languages.

Table 1: Details of Topics and Instructions provided to Chatsonic along withlanguage details of the contents created

Sl No.	Topic with Instructions	Language of the content created
1	Write a detailed article on the topic of plagiarism. Please ensure that the article provides a comprehensive understanding of what plagiarism is, its consequences, and how to avoid it. The article should be written in a formal tone and should be around 500 words in length. Please include examples and references to support your points.	English
2	Write a comprehensive paragraph discussing the Indian Copyright Act. Provide an overview of the act, its purpose, and its key provisions. Additionally, highlights the enforcement mechanisms and penalties for copyright infringement under the Indian Copyright Act. What is IPR? Explain the different types of IPR. Explain the importance of IPR.	English

3	Write an article about Rabindranath Tagore in Bengali in 500 words.	Bengali
4	Write a paragraph on the primary education system in India in Bengali in 500 words. Add information about the challenges in the primary education system. Explore solutions for these challenges.	Bengali
5	Write about Hindi literature and its famous authors in Hindi. Summarize the major works of these authors.	Hindi
6	Write an essay on environmental and ICT effects positive and negative in Hindi. Summarize the positive impact of ICT on the environment. Summarize the positive impact of ICT on the environment in Hindi.	Hindi
7	Recent Advancements in the Application of Artificial Intelligence	English
8	Write an article on renewable energy and its application in India. The article should have abstract keywords that reference statistical data. Explain the role of renewable energy in ICT.	English
9	Write an article on the application of statistics, and metrics in journal articles in 500 words with citations and references.	English
10	Write an article on Publication ethics.	English

Thereafter, the contents created through AI were uploaded one by one to DrillBit, Grammarly, and iThenticate to check how efficiently they detect the contents created through the Chatsonic Artificial intelligence tool.Similar settings were maintained in both the plagiarism detection software.

It is to be mentioned that 100% of all the contents were created through the Artificial intelligence tool-Chatsonic only. The similarity result is as follows:

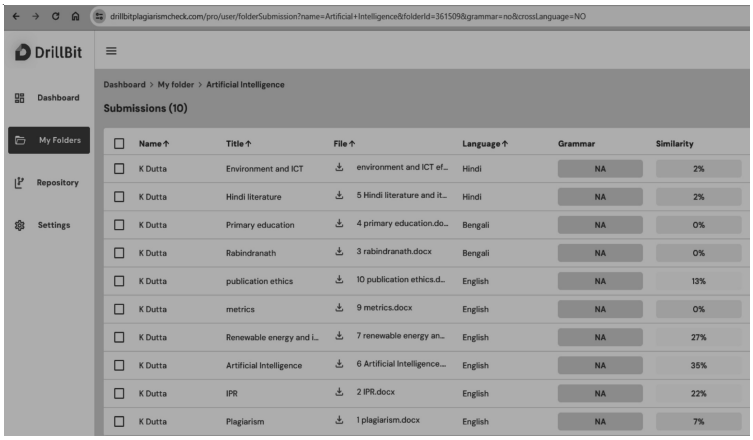

Figure 1: Dashboard of DrillBit Plagiarism Checker with documents uploaded for the study

Figure 1 shows - that for both the documents created in the Bengali language, the similarity is 0%. For both the documents created in the Hindilanguage, the similarity is 2%. For documents created in English language, one shows 0% similarity and other shows similarity percentage from 7 to 35.

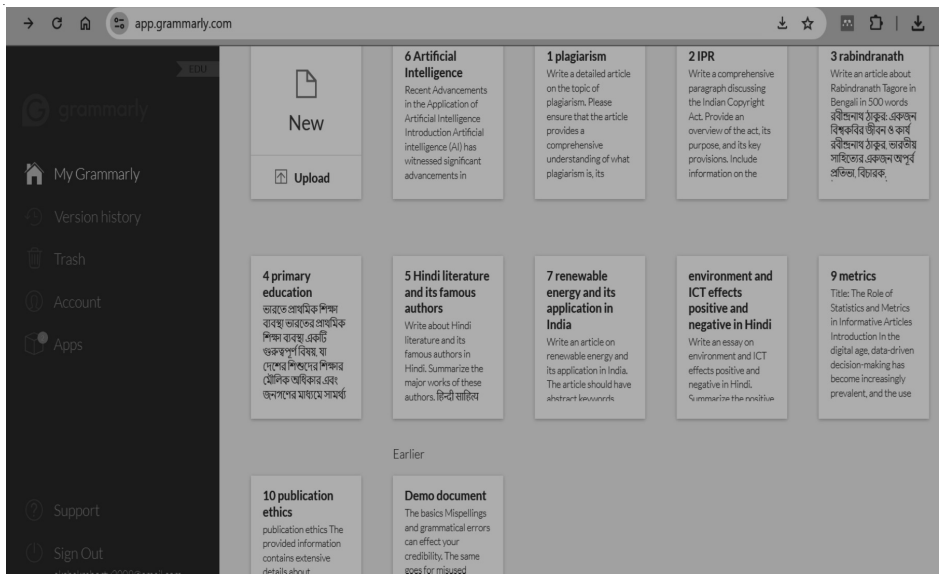

Figure 2: Dashboard of Grammarly with documents uploaded for the study

Figure 2 shows all the documents uploaded to Grammarly. Clicking on them reveals the similarity and AI detection percentage along with other modifications required to improve the quality of the content.

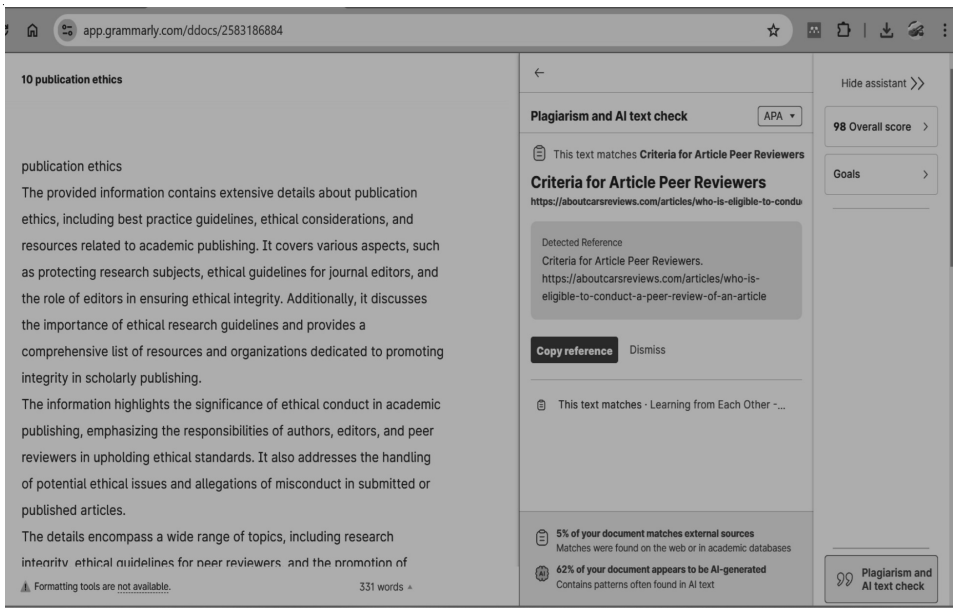

Figure 3: Dashboard of Grammarly showing Similarity and AI percentage of an individual file

Figure 3 shows Grammarly detects 5% of matches from external sources and 62% of matches as AI-generated text. Grammarly also suggested references not included in the study by identifying their sources.

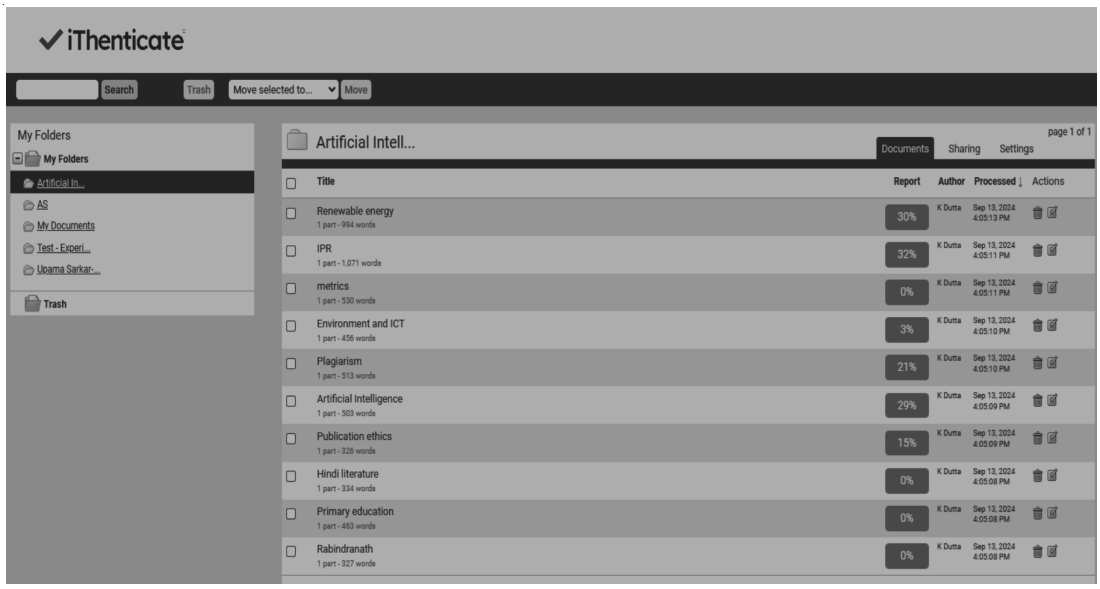

Figure 4: Dashboard of iThenticate with documents uploaded for the study

Figure 4 shows that the similarity is 0% for both documents created in Bengali. For both the documents created in the Hindi language, the similarity is 0% for one document and 3% for another. For documents created in English, one shows 0% similarity, and the other shows a similarity percentage from 15 to 32.

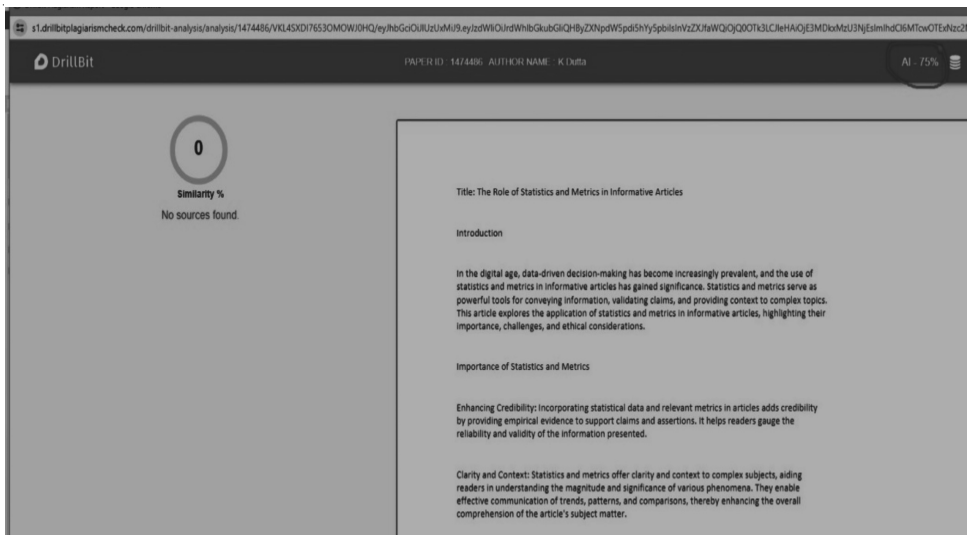

Figure 5: Dashboard of DrillBit Plagiarism Checker showing report with Similarity% and AI%

Figure 5 illustrates the interface of DrillBit plagiarism checker, showing Similarity percentage (0%) and AI-generated content (75%) for the content on topic metrics.

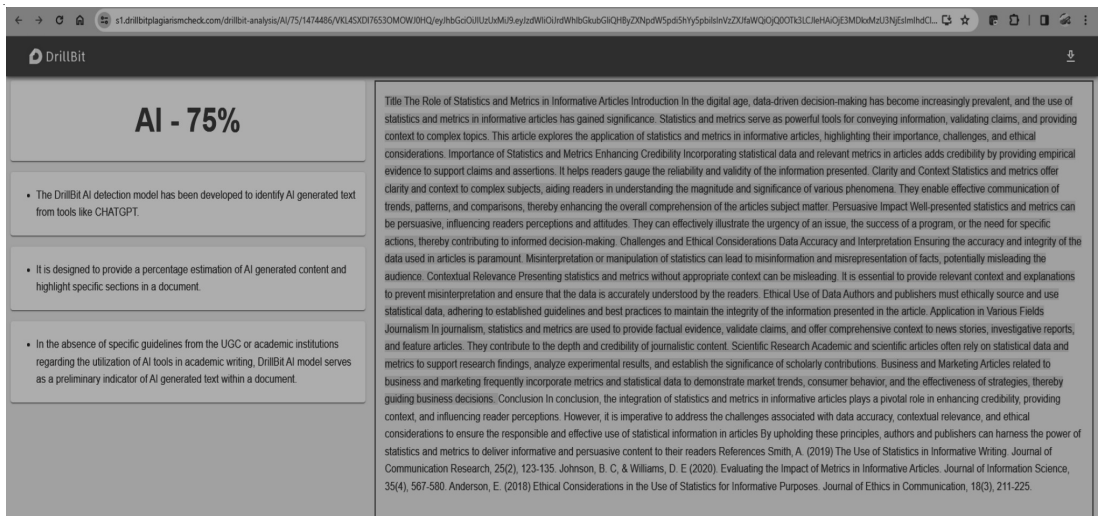

Figure 6: Dashboard of DrillBit Plagiarism Checker showing Detected AI written contents along with percentage

Figure 6 illustrates the interface of the DrillBit plagiarism checker, showing AI-generated content (75%) highlighted along with some notes for the content on metrics.

Table 2: Similarity% and AI Detection% shown by DrillBit, Grammarly, and iThenticate

Sl No.	Topic	Language of the content created Similarity	Detection			AI Detection (without humanization)		
			Drill Bit	Grammarly	iThenticate	DrillBit	Grammarly	iThenticate
1	Plagiarism	English	7%	6%	21%	0%	72%	AI detection not been included yet
2	IPR.	English	22%	17%	32%	64%	53%	
3	Rabindranath	Bengali	0%	0%	0%	0%	12%	
4	Primary education system	Bengali	0%	0%	0%	0%	0%	
5	Hindi literature	Hindi	2%	0%	0%	0%	0%	
6	Artificial Intelligence	English	35%	25%	29%	0%	27%	
7	Renewable energy	English	27%	12%	30%	70%	48%	
8	Environment and impact of ICT in Hindi	Hindi	2%	0%	3%	0%	0%	
9	Metrics	English	0%	4%	0%	75%	85%	
10	Publication Ethics	English	13%	5%	15%	56%	62%	

Table 2 clarifies that the result of similarity detection and Artificial Intelligence (AI)tool written content detection varies from one tool to another.

While considering AI text detection, the result shows that DrillBit and Grammarly can detect the AI text. As observed, AI text detection is often better in Grammarly (0% - 3 documents, other documents with 12%,27%,48%,53%,62%, 72%, 85%) than in DrillBit (0% - 6 documents, other documents with 56%, 64%, 70%, 75%).But in fact, 100% of texts were AI-generated.

While considering documents written in Bengali and Hindi, similarity detection in Grammarly, DrillBit, and iThenticate shows very poor results or almost 0% similarity. Only Grammarly detected 12% of AI-generated text for Bengali documents in one out of two documents.

It is worth mentioning that one English document with 0% similarity shows 75% AI written by DrillBit. iThenticate efficiently finds similarities or sources of AI-generated content to some extent, resulting in greater similarity percentages.

7. Discussions

The extent of similarity and AI detection varies from one tool to another. The situation is a matter of concern for documents related to regional Indian languages. However, the point to be noted is that similarity detection tools like Grammarly and DrillBit can detect AI-generated content that is not humanized. DrillBit is good for checking similarities in Indian regional languages to some extent. Grammarly is good for students who are novices in writing. For publishing journal articles with standard publishers, iThenticate can be considered due to its wide quality data coverage for similarity checks. All the similarity detection tools have some common features like including or excluding quotes, bibliographies, etc., and detection of AI written context. But in addition to this, their features vary as per the target end users. The selection of the tools needs proper justification with clear and specific predefined objectives. The main point to consider is that while selecting a similarity detection tool, whether the ability to detect AI-generated content will be given importance or not as the study shows even if the tools have AI content detection features, the percentage of detection varies.

Conclusion

Every similarity or plagiarism detection tool has some limitations. On the other hand, it is also proved that there are some mechanisms to detect similarity and AI written content to reduce plagiarism and enhance publication ethics. These tools are in the constant modification stage as per demand. Policies should be there to justify the extent of usage of AI-written content in academic writing. Many humanization tools are also available, which make identification of AI-generated creations undetectable. Techniques to identify that also need to be implemented. Some tools can detect AI-generated content for free for some limited words/documents. The study can be done with more similarity detection tools covering more languages. Studies may also be done with AI-generated content after humanizing them.

References

Dehouche, N. (2021). Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). Ethics in Science and Environmental Politics,21,17-23.

DrillBit. (n.d.). FAQ. Retrieved from https://www.drillbitplagiarism.com/faq

Gao, C.A. et al. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digit. Med. 6, 75 (2023). Retrieved from https://doi.org/10.1038/s41746-023-00819-6

Grammarly. (n.d.). Features. Retrieved from https://www.grammarly.com/features

INFLIBNET. (n.d.). Retrieved from https://etd2023.inflibnet.ac.in/docs/PPT/CP1.pdf

iThenticate. (n.d.). iThenticateFAQs. Retrieved from https://www.ithenticate.com/resources#ithenticate-faq

Jarrah, A. M, et al (2023). "Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say?" Online Journal of Communication and Media Technologies.

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of Plagiarism Detection. Interacción. DOI:10.48550/arXiv.2302.04335

King, M.R., &chatGPT (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. Cellular and Molecular Bioengineering, 16, 1-2.

Kleebayoon, A., &Wiwanitkit, V. (2023). Artificial Intelligence, Chatbots, Plagiarism and Basic Honesty: Comment. Cellular and Molecular Bioengineering, 16, 173-174.

Patra, S. K., &Kirtania, D. K. (2023). OpenAIChatGPT generated content and similarity index: A study of selected terms from the library & information science. Annals of Library and Information Studies, 70(2), 99-101.

Sabeeh, M., & Khaled, F. (2021). Plagiarism Detection Methods and Tools: An Overview. Iraqi Journal of Science.

Shodhshuddhi. (n.d.). Retrieved from https://shodhshuddhi.inflibnet.ac.in/

Writesonic. (n.d.). Retrieved from https://writesonic.com/chat