



Short Communication

OpenAI ChatGPT generated content and similarity index: A study of selected terms from the library & information science

Swapan Kumar Patra^a and Deep Kumar Kirtania^b

^aDepartment of Library and Information Science, Sidho-Kanho-Birsha University, Purulia, West Bengal, India,
Email: skpatra@gmail.com

^bBankura Sammilani College, Bankura, West Bengal, India, Email: deepkrlis@gmail.com

Received: 05 May 2023; revised & accepted: 14 May 2023

With the increasing popularity of OpenAI ChatGPT, there are concerns about its use and abuse. In this paper, ChatGPT-generated contents are analysed with randomly selected terminologies related to library and information science discipline. The generated contents were checked through iThenticate, Turnitin and Urkund software to see the percentage of similarity. Only 7 percent similarity in iThenticate and Urkund and 13 percent in Turnitin was found.

Keywords: *ChatGPT, OpenAI, Artificial Intelligence, Similarity Index, Academic Integrity, Plagiarism, Library & Information Science*

Introduction

Artificial Intelligence (AI) is a field of computer science that focuses on developing machines and software that can perform tasks that typically require human intelligence. AI tools and techniques are widely used in various fields, such as visual perception, speech recognition, decision-making, language understanding and so on^{1,2}. ChatGPT is an AI language model developed by OpenAI. It is part of a class of AI models known as "transformer models" that use deep learning to generate human-like text based on a large trained corpus of textual data³. It is trained on a vast amount of data from various source including the internet, books, and other sources. ChatGPT is designed to understand and respond to a wide range of questions and topics. It is a useful tool for natural language processing, chatbot development, and other applications where natural language understanding and generation is required. It is a large language model that uses deep learning techniques to generate human-like responses to text-based inputs. It answers a wide range of questions and engage in conversations on a variety of topics. ChatGPT's popularity skyrocketed almost immediately after its release. In just 5 days the number of users of this technology reached 10,00,000 and by the end of January 2023 there was more than 100 million users⁴.

ChatGPT can be used to generate original text and ideas, which can help students and researchers avoid

the temptation to plagiarize⁵. Additionally, ChatGPT can assist educators in detecting instances of plagiarism in student work, allowing for early intervention and prevention. It can be a valuable tool in promoting academic integrity by helping to identify potential cases of plagiarism and providing guidance on how to avoid it. Students and researchers can use ChatGPT to check their writing for similarities to existing texts and to receive feedback on how to properly attribute sources and avoid unintentional plagiarism.

Several scholarly literatures have already been published on the subject. OpenAI ChatGPT generated literature review on the topic "Digital Twin in Healthcare" showed that paraphrased portions of content generated from ChatGPT had significant similarities³. Lund and Wang (2023)⁶ studied ChatGPT's history, technology, and its potential impact on academia and libraries via interview method. Ventayen's (2023)⁷ study showed the similarity index passed the institution's required similarity index. Apart from that, there has been a lot of work on applications of ChatGPT on various topics like academic performance⁸, education^{9,10,11}, education and lifelong learning¹², financial sector^{13,14,15}, forensic accounting¹⁶, information literacy¹⁷, law^{18,19}, pharmacology²⁰, public health²¹, psychology²², science²³ etc.

This paper examines similarity index in ChatGPT generated LIS content.

Table 1 — Similarity index of ChatGPT generated contents using various

Software	Similarity Index	No of Paragraph	Similarity Matching Paragraph	Not matching Paragraph
iThenticate	7%	83	18	65
Turnitin	13%	83	29	54
Urkund	7%	83	17	66

Method

Ten randomly chosen 20 topics in library and information science were used in ChaptGPT <https://chat.openai.com/chat> to generate content. The topics are: *Altmetrics, Archiving, Bibliometric Analysis, Cataloging, Circulation Services, Citation in Research, Data Science, Digital Preservation, Information Retrieval, Library and Information Science, Library Automation, Library Management, Open Access, Plagiarism, Public Library, Reference Services, Research Data Management, Resource Sharing, Scholarly Communication, and Scientometrics Analysis*. Subsequently, those contents were checked for similarity index with the help of iThenticate, Turnitin and Urkund plagiarism detecting tool.

Results

The similarity checking of the ChatGPT generated contents using different software are as follows iThenticate (7%), Urkund (7%) and Turnitin (13%) similarity. The similarity index of its content is within the permissible limit among the academia and other sectors. Moreover, paragraph wise checking of the contents, shows that out of 83 paragraphs, 65 paragraphs in iThenticate, 54 paragraphs in Turnitin and 66 paragraphs in Urkund have no similarity matching. Table 1 shows the similarity check of ChatGPT generated content using various plagiarism detection tools.

Conclusion

This study showed that most of the content created by ChatGPT is relatively less in the similarity index. However, there is still a lot of testing and research work to be done on ChatGPT. This study is based on only twenty sample queries. Perhaps a better result could be obtained with the inclusion of more words or questions to get more contents. Therefore, it is not possible to accurately assess the effectiveness of ChatGPT at the present time or from this small scope and coverage-based research work.

Although ChatGPT has recently started its functionality to public, researchers from all over the world of various subjects have started researching this topic. Effectiveness and application of ChatGPT in various subjects as well as how this tool can help students in academically should be researched. Besides, ethical issues related to ChatGPT, such as academic integrity, research ethics, plagiarism, and so on can also be the subject of research.

Acknowledgement

The preprint of the article is available at <https://www.qeios.com/read/Fo1CP6.3>. The authors are thankful to the reviewers for the insightful and constructive comments on the preprint version of the article.

References

- 1 Winston P H, *Artificial intelligence*, (Addison-Wesley Longman Publishing Co., Inc.) 1984, p. 1-15.
- 2 Russell S J, *Artificial intelligence a modern approach*, 4th edn (Pearson Education Inc), 2010, p. 5-20.
- 3 Aydın Ö and Karaarslan E, OpenAIChatGPT generated literature review: Digital twin in healthcare. (2022), Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4308687 (Accessed on 25 Feb 2023).
- 4 *CHATGPT sets record for fastest-growing app with 100 million users: Report*. Hindustan Times. (2023, February 2), Available at <https://www.hindustantimes.com/technology/chatgpt-sets-record-for-fastest-growing-app-with-100-million-users-report-101675323297627.html>(Accessed on February 22 Feb 2023).
- 5 Rudolph J, Tan S and Tan S, ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6 (1), (2023) 1-22.
- 6 Lund B D and Wang T, Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*, (2023). <https://doi.org/10.1108/LHTN-01-2023-0009> (Accessed on 25 Feb 2023).
- 7 Ventayen R J M, OpenAIChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4332664 (Accessed on 25 Feb 2023).
- 8 Alshater M, Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT, (2022). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312358 (Accessed on 25 Feb 2023).
- 9 Baidoo-Anu D and Owusu Ansah L, Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. (2023), Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4337484 (Accessed on 25 Feb 2023).
- 10 Zentner A, Applied Innovation: Artificial Intelligence in Higher Education, (2022). Available at <https://>

- papers.ssrn.com/sol3/papers.cfm?abstract_id=4314180 (Accessed on 25 Feb 2023).
- 11 Zhai X, ChatGPT user experience: Implications for education, (2022). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418 (Accessed on 25 Feb 2023).
 - 12 Mhlanga D, Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. *Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning*, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354422 (Accessed on 26 Feb 2023).
 - 13 Dowling M and Lucey B, ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters*, 103662 (2023). Available at <https://www.sciencedirect.com/science/article/pii/S1544612323000363> (Accessed on 27 Feb 2023).
 - 14 Yue T, Au D, Au C C and Iu K Y, Democratizing financial knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4346152 (Accessed on 25 Feb 2023).
 - 15 Zaremba A and Demir E, ChatGPT: Unlocking the Future of NLP in Finance, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4323643 (Accessed on 27 Feb 2023).
 - 16 Street D and Wilck J, 'Let's Have a Chat': Principles for the Effective Application of ChatGPT and Large Language Models in the Practice of Forensic Accounting, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4351817 (Accessed on 25 Feb 2023).
 - 17 Lund B and Agbaji D, Information Literacy, Data Literacy, Privacy Literacy, and ChatGPT: Technology Literacies Align with Perspectives on Emerging Technology Adoption within Communities. *Data Literacy, Privacy Literacy, and ChatGPT: Technology Literacies Align with Perspectives on Emerging Technology Adoption within Communities*, (2023).
 - 18 Armstrong A B, Who's Afraid of ChatGPT? An Examination of ChatGPT's Implications for Legal Writing. *An Examination of ChatGPT's Implications for Legal Writing* (2023).
 - 19 Bishop L, Can ChatGPT Think Like a Lawyer? A Socratic Dialogue. *A Socratic Dialogue*, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4338995 (Accessed on 25 Feb 2023).
 - 20 Nisar S and Aslam M S, Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology?, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4324310 (Accessed on 25 Feb 2023).
 - 21 De Angelis L, Baglivo F, Arzilli G, Privitera G P, Ferragina P, Tozzi A E and Rizzo C, ChatGPT and the Rise of Large Language Models: The New AI-Driven Infodemic Threat in Public Health. (2023), Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4352931 (Accessed on 25 Feb 2023).
 - 22 Uludag K, The Use of AI-Supported Chatbot in Psychology, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331367 (Accessed on 24 Feb 2023).
 - 23 Zhai X, ChatGPT for Next Generation Science Learning, (2023). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331313 (Accessed on 26 Feb 2023).