

Research Article

OpenAI ChatGPT Generated Content and Similarity Index: A study of selected terms from the Library & Information Science (LIS) discipline

Deep Kumar Kirtania¹, Swapan Kumar Patra²

1. Bankura Sammilani College, Bankura, India; 2. Sidho-Kanho-Birsha University, India, Lagda

With the increasing popularity of OpenAI ChatGPT there are concerns about its use and abuse from all over the globe. Few have appreciated this effort and others are skeptical about it. Whether a proponent or opponent of this tool, no one can deny the fact that this new AI will revolutionize every aspect of human being. In this context, the ChatGPT contents generated through OpenAI ChatGPT are analyzed with the selected terminologies of Library and Information Science (LIS) discipline. About ten popular terms from the LIS field were put on Chat GPT (<https://chat.openai.com/>) to generate contents. Those contents were checked through Turnitin software (a web-based plagiarism detection service) to see the percentage of similarity. The study observed that, after checking the content with the plagiarism tool, only 13 percent similarity was found from these ten contents. The observation is significant in the context of academic integrity. The findings are perhaps useful for library information science professionals in any academic institution to generate contents.

Corresponding authors: Deep Kumar Kirtania, deepkrlis@gmail.com; Swapan Kumar Patra, skpatra@gmail.com

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, such as visual perception, speech recognition, decision-making, and language understanding (Winston, 1984; Russell, 2010). ChatGPT is an artificial intelligence 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 corpus of text data that it has been trained on (Aydın & Karaarslan, 2022). ChatGPT is designed to understand and respond to a wide range of questions and topics, making it 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 is trained on a vast amount of data from various source including the internet, books, and other sources. 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 lakhs and by the month of January 2023 the number of users reached more than 100 million users (Hindustan Times 2nd February, 2023).

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. ChatGPT can also be used to generate original text and ideas, which can help students and researchers avoid the temptation to plagiarize (Rudolph, Tan & Tan, 2023). Additionally, ChatGPT can assist educators in detecting instances of plagiarism in student work, allowing for early intervention and prevention.

However, the content generated by ChatGPT must be checked to see if it is authentic or unique. This paper selects a few important terminologies from the Library and Information Science field.

The study aims to find the similarity of the content of a few terms generated through ChatGPT in this field.

Literature review

ChatGPT is an AI language model created by OpenAI. As an AI language model, it is trained on a large dataset of text, to understand and generate natural language responses to questions and statements asked by its users.

Although ChatGPT is a completely new concept, there is a great interest in this topic among scholars all over the world. A number of good quality informative research papers have already been published on the subject. Aydın and Karaarslan (2022) studied the OpenAI ChatGPT generated literature review on the

topic Digital Twin in Healthcare. In this research, literature review was created using ChatGPT and similarity was checked with the help of Authenticate plagiarism tool. This work showed that paraphrased portions of content generated from ChatGPT had significant similarities. Lund and Wang (2023) the paper includes ChatGPT, history and technology of GPT, and its potential impact on academia and libraries via interview method. Ventayen (2023) studied the similarity Index of Artificial Intelligence-Based (ChatGPT) based Content. The study showed the generated results' similarity index passed the institution's required similarity index. Apart from that, there has been a lot of work on applications of Chat DPT on various topics like Academic performance (Alshater, 2022), Education (Baidoo-Anu & Owusu Ansah, 2023; Zentner, 2022; Zhai, 2022), Education and Lifelong Learning (Mhlanga, 2023), Financial Sector (Dowling & Lucey, 2023; Yue, Au, Au & Iu, 2023; Zaremba & Demir, 2023), Forensic Accounting (Street & Wilck, 2023), Information Literacy (Lund & Agbaji, 2023), Law (Armstrong, 2023; Bishop, 2023), Pharmacology (Nisar & Aslam, 2023), Public Health (De Angelis et al, 2023), Psychology (Uludag, 2023), Science (Zhai, 2023) etc. From these works on ChatGPT has been elaborated and our understanding of the application of ChatGPT has also been increased. This work is going to add value in evaluating the contents generated by the ChatGPT in the field of Library and Information Science.

Methodology

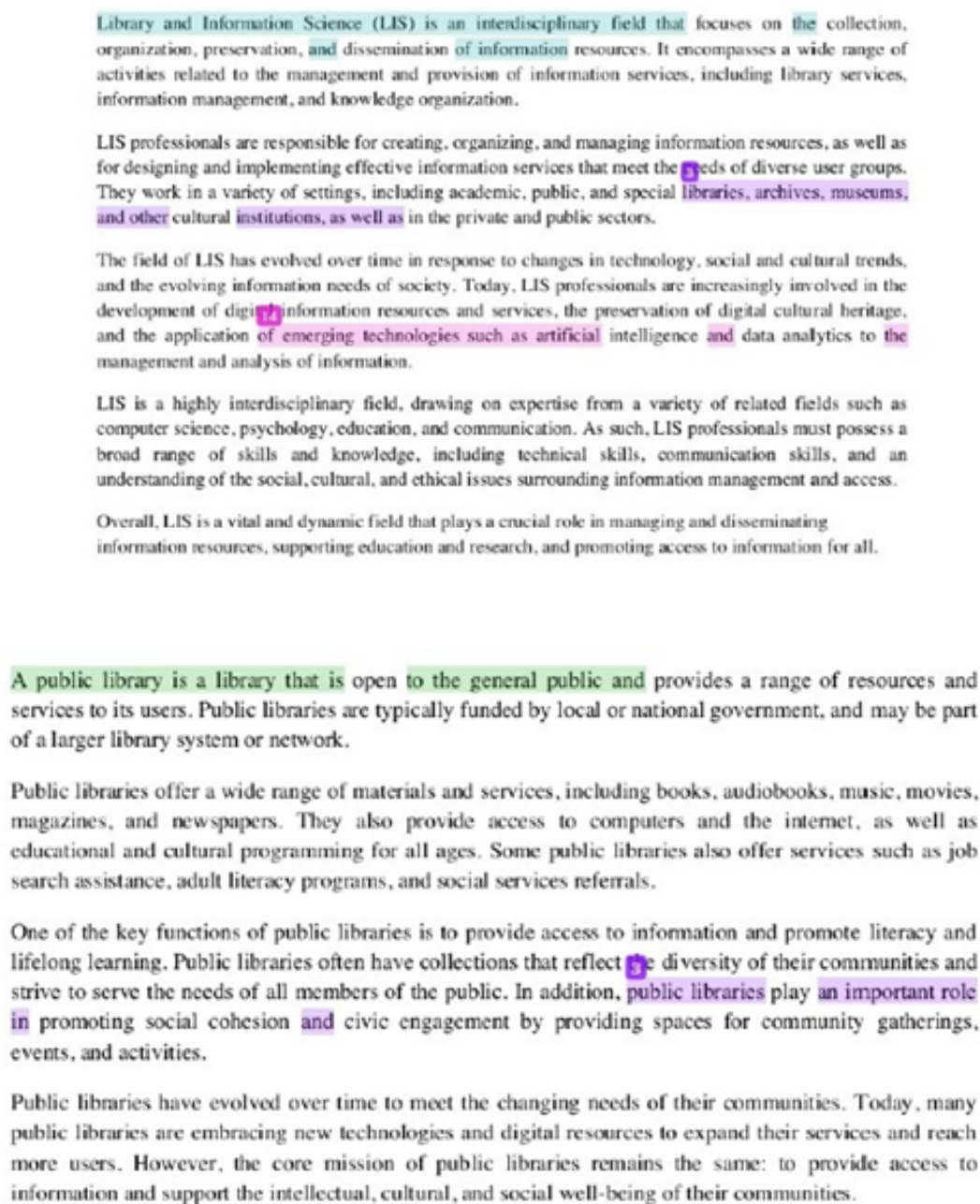
The methodology used in this paper is quite simple and practical. In the first phase of the paper generated content from the Chat GPT on very ten popular topics in library and information science. The topics are; *Library and Information Science, Public Library, Bibliometric Analysis, Library Automation, Library Management, Information Retrieval, Scholarly Communication, Citation in research, Plagiarism, Research data management*. Subsequently, those contents were checked for similarity index with the help of Turnitin plagiarism detecting tool, to check the similarity among the text. Turnitin is a web-based plagiarism detection service that helps educators and institutions detect plagiarism. It compares submitted papers against a database of other papers, as well as online sources and other publicly available materials, to identify any instances of copying or plagiarism. Finally, an attempt has been made to arrive at appropriate conclusions by analysing these results from the present study.

Results

The similarity checking of the ChatGPT generated contents using the Turnitin software shows about 13% similarity from the above ten contents. 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 51 paragraphs, 32 paragraphs i.e., about sixty percent of paragraphs have no similarity matching. The following figures shows the Similarity Check of ChatGPT Generated Content using Turnitin.

Figure 1. Similarity Check of ChatGPT Generated Content using Turnitin.



Bibliometric analysis is a method used to study and analyze scholarly literature, particularly scientific publications. It involves using statistical and quantitative methods to measure various aspects of publications, including their frequency, impact, and productivity.

Bibliometric analysis can be used to identify key authors, institutions, and research areas within a field, as well as to track trends over time. It can also be used to evaluate the impact of individual publications or researchers, and to inform decisions about funding and resource allocation.

Bibliometric analysis relies on various measures and indicators, such as citation counts, h-indices, and journal impact factors, to assess the quality and impact of research output. However, it is important to note that these measures are not without limitations, and should be used with caution and in conjunction with other forms of evaluation.

Overall, bibliometric analysis is a powerful tool that can provide valuable insights into the structure and dynamics of the scientific literature, and inform decision-making in academia and beyond.

Library automation is the process of using computer software and technology to manage and automate various library operations and services. This includes tasks such as cataloging, circulation, inventory management, and interlibrary loan.

Library automation has several benefits, such as improving the efficiency and accuracy of library services, reducing staff workload, and enhancing access to library resources for users. With automation, libraries can provide more effective and timely services, and can free up staff time to focus on more complex tasks and user engagement.

Library automation systems typically include a range of modules or components, such as cataloging, circulation, acquisitions, serials management, and discovery interfaces. These systems can be customized to meet the specific needs of different types of libraries, such as academic, public, and special libraries.

In addition to traditional library automation systems, libraries are also exploring the use of emerging technologies such as artificial intelligence, machine learning, and the Internet of Things to further automate and enhance library operations and services.

Overall, library automation has revolutionized the way libraries operate and has contributed to the development of more efficient and effective library services that better meet the needs of their users.

Library management is the process of planning, organizing, directing, and controlling library resources and services to meet the needs of library users. Effective library management involves developing and implementing policies and procedures to ensure the smooth operation of library services and programs.

Library management encompasses a wide range of activities, including collection development and management, cataloging and classification, circulation, reference and information services, outreach and programming, and facility management. Library managers are responsible for overseeing these activities and ensuring that they are carried out effectively and efficiently.

Library managers must also be skilled in personnel management, including hiring, training, and supervising staff, as well as developing and maintaining positive relationships with other stakeholders, such as library users, community organizations, and funding agencies.

Effective library management requires a combination of technical and managerial skills, as well as a thorough understanding of the mission and goals of the library and the needs of its users. Library managers must be able to balance the competing demands of different stakeholders, prioritize tasks and resources, and adapt to changes in technology, funding, and other factors that can affect library operations.

In recent years, library management has become increasingly complex, with the rapid development of digital technologies and the growing diversity of library users and services. Library managers must therefore be able to navigate these changes and adapt to new trends and challenges in order to provide high-quality, relevant library services and programs to their communities.

Information retrieval is the process of finding relevant information from a large and complex set of data or information resources. It involves the use of specialized tools and techniques to search, retrieve, and evaluate information in response to specific user needs or queries.

Information retrieval systems may include search engines, databases, or other tools designed to help users find information in different formats, such as text, images, or multimedia content. These systems rely on sophisticated algorithms and techniques to analyze and index information resources, and to match user queries with relevant information.

The field of information retrieval has evolved over time, in response to changes in technology and the growing volume and diversity of information available online. Today, information retrieval systems are used in a wide range of applications, from web search engines and digital libraries to e-commerce and social media.

Information retrieval is a complex process that requires a combination of technical and analytical skills, as well as an understanding of the needs and preferences of different user groups. Effective information retrieval systems must be designed with user needs in mind, and must be able to handle diverse queries and search parameters.

In addition to technical considerations, information retrieval also raises important ethical and social issues, such as privacy, bias, and the quality and reliability of information. As such, information retrieval professionals must be aware of these issues and work to design systems that are fair, transparent, and accessible to all.

Overall, information retrieval is a crucial process that enables users to access and use the vast and growing array of information resources available in today's digital age.

Scholarly communication refers to the ways in which scholars and researchers communicate their work and ideas to one another and to the wider community. It encompasses a wide range of activities, including publishing research articles, presenting at conferences, sharing data and code, and engaging in public outreach and education.

The goal of scholarly communication is to advance knowledge and understanding in a particular field of study, and to ensure that research findings are widely disseminated and accessible to other scholars, practitioners, and the general public.

Scholarly communication is evolving rapidly in response to changes in technology, funding, and academic culture. The rise of digital publishing and open access has made it easier for researchers to share their work with a global audience, while new tools and platforms for data sharing and collaboration have made it possible to collaborate and communicate in new and innovative ways.

Scholarly communication also raises important ethical and social issues, such as the need to ensure that research is conducted with integrity and rigor, the need to protect the rights of authors and other stakeholders, and the need to ensure that research is accessible and relevant to diverse communities.

Effective scholarly communication requires a combination of technical and interpersonal skills, as well as an understanding of the needs and preferences of different audiences. It also requires a willingness to engage in ongoing dialogue and debate with other scholars and practitioners, in order to advance knowledge and understanding in a particular field of study.

Overall, scholarly communication is a critical component of the research process, enabling scholars and researchers to share their findings, build connections, and make meaningful contributions to their fields of study and to society more broadly.

Citation is the practice of acknowledging and referencing sources of information used in research. Proper citation is an essential component of the research process, as it allows readers to trace the sources of ideas, data, and arguments presented in a research paper, and it gives credit to the authors and publishers of the sources used.

Citation styles vary across disciplines, but all generally include the author's name, the title of the work, and the date of publication. Some citation styles also include information about the publisher, the volume and issue number, and the page numbers of the source material.

In addition to giving credit to the sources used, citation also helps to establish the credibility and authority of a research paper. By citing reputable sources and following established citation conventions, researchers demonstrate their knowledge and understanding of the field, and they show that their work is grounded in established scholarship.

Citation also helps to prevent plagiarism, which is the use of someone else's work without proper attribution. By citing sources and giving credit where it is due, researchers demonstrate their ethical commitment to academic integrity and the responsible conduct of research.

In summary, proper citation is an essential component of the research process. It helps readers to trace the sources of information used in a research paper, gives credit to the authors and publishers of those sources, establishes the credibility and authority of the paper, and helps to prevent plagiarism.

Plagiarism is the act of using someone else's words, ideas, or work without giving proper credit or attribution. It is a serious violation of academic integrity and can have serious consequences, including loss of credibility, academic penalties, and legal action.

Plagiarism can take many forms, including copying and pasting text from a source without proper citation, paraphrasing someone else's work without proper attribution, or submitting someone else's work as one's own.

Plagiarism is a problem in many fields, including academia, journalism, and creative writing. It can undermine the credibility of the person who commits it, as well as the credibility of the work in which it appears.

To avoid plagiarism, it is important to always give proper credit to the sources of ideas, words, or images that are used in one's work. This can be done by using proper citation styles, such as MLA or APA, and by providing attribution for direct quotes, paraphrased material, and images.

It is also important to be aware of the different types of plagiarism and to understand the consequences of committing plagiarism. Academic institutions often have strict policies and procedures in place to prevent and detect plagiarism, and can impose penalties ranging from a failing grade to expulsion.

Overall, avoiding plagiarism is essential to maintaining academic and professional integrity, and is an important part of the ethical practice of research and scholarship.

Research data management refers to the process of collecting, organizing, storing, and sharing data generated during research. Effective research data management is critical to ensuring the quality and integrity of research, as well as to complying with ethical, legal, and regulatory requirements.

Research data management involves a number of different activities, including developing data management plans, selecting appropriate storage and backup options, ensuring data security and privacy, and sharing data with other researchers or the public.

Good data management practices also help to ensure that data are discoverable, accessible, and reusable. This allows other researchers to verify results, build on existing research, and create new knowledge.

Effective research data management requires collaboration and communication among researchers, data managers, and other stakeholders. It also requires careful attention to the needs and preferences of different research communities, and a willingness to adopt new technologies and standards as they emerge.

Overall, effective research data management is essential to ensuring the quality and integrity of research, and to advancing knowledge and understanding in a particular field of study. It requires careful planning, attention to detail, and a commitment to ethical and responsible research practices.

Conclusion

There is no doubt that artificial intelligence will be at the forefront of technology in the coming days. ChatGPT is one of the latest applications of artificial intelligence but has established itself as one of the most popular tools around the world today. As an AI language model, ChatGPT generated responses are based on patterns and structures from vast amounts of textual data. The response from ChatGPT says that the responses are not intended to be used for academic purposes, such as writing papers or completing assignments. They are intended to provide general information and guidance on a wide range of topics.

In the current research paper, we verified the content created by ChatGPT with plagiarism software to see how much similar is the text. For the present paper we have chosen the subject of library and information science. In this case we have observed that only 13 percent similarity matching text was found from ten ChatGPT created content. From this result it can be easily said that most of the content created by ChatGPT is relatively less in the similarity index or plagiarism free text. However, there is still a lot of testing and research work to be done on Chat GPT. 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.

Limitation of the study and scope for further research

This study is based on only ten sample queries. Perhaps a better result could be obtained with the inclusion of more words or questions to get more contents.

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.

References

- Armstrong, A. B. (2023). Who's Afraid of ChatGPT? An Examination of ChatGPT's Implications for Legal Writing. *An Examination of ChatGPT's Implications for Legal Writing (January 23, 2023)*.
- Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. *Available at SSRN 4308687*.
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Available at SSRN 4337484*.
- Bishop, L. (2023). Can ChatGPT Think Like a Lawyer? A Socratic Dialogue. *A Socratic Dialogue (January 26, 2023)*.
- CHATGPT sets record for fastest-growing app with 100 million users: Report. Hindustan Times. (2023, February 2). Retrieved February 22, 2023, from <https://www.hindustantimes.com/technology/chatgpt-sets-record-for-fastest-growing-app-with-100-million-users-report-101675323297627.html>
- De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the Rise of Large Language Models: The New AI-Driven Infodemic Threat in Public Health. *Available at SSRN 4352931*.
- Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters*, 103662.
- Lund, B., & Agbaji, D. (2023). 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* (January 14, 2023).

- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library Hi Tech News*.
- M Alshater, M. (2022). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT. *Available at SSRN*.
- Mhlanga, D. (2023). 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* (February 11, 2023).
- Nisar, S., & Aslam, M. S. (2023). Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology?. *Available at SSRN 4324310*.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1).
- Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc..
- Street, D., & Wilck, J. (2023). 'Let's Have a Chat': Principles for the Effective Application of ChatGPT and Large Language Models in the Practice of Forensic Accounting. *Available at SSRN 4351817*.
- Uludag, K. (2023). The Use of AI-Supported Chatbot in Psychology. *Available at SSRN 4331367*.
- Ventayen, R. J. M. (2023). OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents. *Available at SSRN 4332664*.
- Winston, P. H. (1984). *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc..
- Yue, T., Au, D., Au, C. C., & Iu, K. Y. (2023). Democratizing financial knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology. *Available at SSRN 4346152*.
- Zaremba, A., & Demir, E. (2023). ChatGPT: Unlocking the Future of NLP in Finance. *Available at SSRN 4323643*.
- Zentner, A. (2022). Applied Innovation: Artificial Intelligence in Higher Education. *Available at SSRN 4314180*.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN 4312418*.
- Zhai, X. (2023). ChatGPT for Next Generation Science Learning. *Available at SSRN 4331313*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.