Comparative Analysis of AI-Generated Research Content: Evaluating ChatGPT and Google Gemini

Irfan ul haq Akhoon

`irfanakhoon@kashmiruniversity.net`

University of Kashmir

Mashood Yousuf Khan

University of Kashmir

Tajamul Ahmad Bhat

University of Kashmir

Additional Declarations: No competing interests reported.

Abstract

Background:

The advent of Natural Language Generation (NLG) models like ChatGPT and Google Bard has transformed academic writing by automating the creation of research articles. This study aims to evaluate the effectiveness of these AI tools in academic content generation, with a focus on their authenticity, relevance, and potential for plagiarism. Additionally, it explores the ethical concerns associated with AI-generated research articles.

Methods:

The research employs a comparative analysis of articles generated by ChatGPT and Google Gemini, using Turnitin, a plagiarism detection tool, to assess the originality of the content. Key parameters, such as citation accuracy, reference authenticity, and the similarity index, were examined to evaluate the validity and ethical use of these AI tools.

Results:

The findings reveal that while ChatGPT and Google Gemini generate coherent articles, both tools frequently produce fabricated citations and references. ChatGPT adhered to APA citation styles but used non-existent sources, while Google Gemini presented some authentic sources but failed to follow proper citation formats. The similarity index for AI-generated content was lower than anticipated, but the repetition of limited sources compromised the comprehensiveness of the work.

Conclusion:

AI tools like ChatGPT and Google Gemini hold potential for streamlining research article generation, but human supervision remains critical. The study emphasizes the need for ethical guidelines and robust content verification methods to mitigate issues such as plagiarism and fabricated data. Researchers and institutions are encouraged to adopt AI tools responsibly, ensuring their use enhances academic integrity rather than undermines it.

Introduction

In recent years, the emergence of large language models (LLMs), such as ChatGPT and Google Gemini, has introduced a new paradigm in academic research. These Natural Language Generation (NLG) models offer researchers the ability to streamline tasks like research planning, content generation, and data analysis, thereby alleviating some of the cognitive load associated with academic writing. The

potential time savings can allow researchers to focus on novel experimental designs and theoretical developments, leading to breakthroughs in various disciplines (Liebrenz et al., 2023).

ChatGPT, developed by OpenAI, represents a significant advancement in artificial intelligence (AI) technology. As part of the Generative Pre-Trained Transformer (GPT) family, ChatGPT leverages deep learning techniques such as supervised learning and reinforcement learning to generate coherent, contextually relevant text. Since its launch in November 2022, it has gained widespread popularity, with applications ranging from writing academic papers to creating computer programs and performing complex data analyses (Gonsalves, 2023; van Dis et al., 2023). Recent studies have explored the role of ChatGPT in the academic realm; Kasneci et al. (2023) emphasizing its potential to assist researchers by generating literature reviews, summarizing articles, and identifying research gaps.

Google Gemini, initially launched as Google Bard, is a conversational generative AI tool developed by Google, designed to build upon the strengths of large language models. Originally based on the LaMDA (Language Model for Dialogue Applications) family of LLMs, it has since evolved through the integration of Google's more advanced Gemini architecture, which focuses on improving both conversational capabilities and contextual understanding in AI-generated content. Released as a response to the increasing adoption of ChatGPT, Google Gemini has rapidly gained traction as a valuable tool for generating informative and contextually appropriate responses across various academic fields. Recent study by Dowling & Lucey (2023), has examined its potential for academic writing, highlighting both its strengths and areas of improvement compared to ChatGPT.

Given the rapid adoption of these tools, understanding their application in the generation of academic content is critical. AI tools like ChatGPT and Google Gemini can reduce time-consuming processes like literature review and article drafting, potentially accelerating the publication process and alleviating writer's block (Kim, 2023). However, the integration of AI in research also raises ethical concerns, particularly around issues such as plagiarism, the fabrication of sources, and the reliability of AI-generated references (Aydın & Karaarslan, 2022). These concerns are magnified in academic settings, where the accuracy and originality of content are paramount.

This study examines the evolution of NLG technology and its application in generating research articles, focusing on the capabilities and limitations of ChatGPT and Google Gemini. In particular, we investigate the authenticity of AI-generated content by assessing its originality and adherence to citation standards using advanced plagiarism detection tools like Turnitin. By highlighting both the potential benefits and ethical challenges, this research aims to provide a comprehensive evaluation of how these AI tools can contribute to academic writing, especially in the field of Library and Information Science.

Several recent studies have begun to address these concerns. For example, Stokel-Walker (2023) reported that ChatGPT has been listed as a co-author on research articles, which has sparked a debate about the role of AI in academic authorship. Similarly, Gao et al. (2022) compared AI-generated abstracts to original scientific abstracts, revealing that while AI models can generate plausible content, they frequently introduce inaccuracies and inconsistencies. This study builds on these findings by offering a

detailed analysis of both ChatGPT and Google Gemini, assessing their effectiveness in different phases of research writing, including the introduction, methodology, literature review, and conclusion.

By focusing on the specific challenges and opportunities presented by these AI models, this research aims to contribute to a more informed understanding of their role in modern academic research, particularly within the field of Library and Information Science.

Literature Review

The application of artificial intelligence (AI) in academic writing has been a growing area of interest, particularly with the introduction of advanced Natural Language Generation (NLG) models like ChatGPT and Google Gemini. Several studies have evaluated the efficacy of these tools in producing academic content, raising important discussions about their reliability, ethical implications, and impact on scholarly research.

Zhai (2022) conducted a notable experiment using **ChatGPT** to compose an academic paper on "Artificial Intelligence for Education." His findings revealed that while the generated writing was coherent and informative, it was only partially accurate and sometimes lacking in depth and critical analysis. This raised questions about the ability of AI models to produce high-quality academic work without human oversight.

Similarly, **Chen (2023)** explored ChatGPT's capacity for scientific writing, particularly its use in language translation. His study demonstrated ChatGPT's potential benefits for translating academic content from Chinese to English, highlighting its usefulness in bridging language barriers in research. However, concerns remained about the accuracy and nuance in translation, which could impact the interpretation of complex scientific concepts.

Aydın and Karaarslan (2022) examined ChatGPT's ability to generate literature reviews in the context of digital twins for healthcare. While the model successfully generated a literature review, the authors discovered that the text contained significant instances of plagiarism and inadequate paraphrasing. These findings emphasize the necessity of using AI tools with caution, particularly in contexts where originality and proper citation are critical.

Another key issue with AI-generated content is the question of authorship. **Stokel-Walker (2023)** reported that ChatGPT has been credited as a co-author in at least four research papers. For instance, **O'Connor and ChatGPT (2023)** published an editorial in *Nurse Education in Practice*, where ChatGPT was listed as an author. However, the attribution of authorship to AI-generated work has sparked considerable debate in the academic community. Prominent publishers, including *Science*, *Nature*, and the *JAMA Network*, have explicitly stated that AI tools cannot be acknowledged as authors due to their lack of accountability and the inability to contribute meaningfully to the intellectual content of a paper (Brainard, 2023).

In response to these controversies, publishing companies have started updating their authorship guidelines. **Van Dis et al. (2023)** and **Liebrenz et al. (2023)** emphasized the need for strict guidelines

when using AI tools like ChatGPT in academic writing. They argue that while these tools can assist in certain aspects of research, the final responsibility must always lie with human researchers. Publishers like *Springer-Nature*, *Elsevier*, and *Taylor & Francis* have updated their policies, stating that AI-generated content must be properly disclosed and cannot be listed as an author (Nature, 2023; Springer-Nature, 2023; Taylor & Francis, 2023).

As for Google's contributions, **Dowling and Lucey (2023)** conducted a comparative analysis of **Google Gemini** (initially launched as Google Bard) and ChatGPT, evaluating their capabilities in academic content generation. They found that while both tools were able to produce coherent research articles, Google Gemini often struggled with maintaining context and consistency in longer texts. This comparative research revealed that although AI models have made significant strides in academic writing, human monitoring is essential to ensure quality, relevance, and ethical compliance.

Gao et al. (2022) also compared AI-generated abstracts from ChatGPT with original scientific abstracts. Their findings revealed that while ChatGPT could generate plausible content, it frequently introduced factual inaccuracies and lacked critical insight, further reinforcing the need for human intervention in AI-assisted writing.

Scope

This study focuses on evaluating research content generated by two of the most popular AI text-generation tools:

·   ChatGPT

·   Google Gemini (initially launched as Google Bard).

By analyzing these aspects, the study aims to provide insights into the capabilities and limitations of ChatGPT and Google Gemini, contributing to ongoing discussions about the role of AI in academic research.

Objectives:

·   To generate research articles using ChatGPT and Google Gemini.

·   To evaluate the similarity ratio of AI-generated content using advanced plagiarism detection tools.

·   To manually review the generated content in terms of structure, including the number of pages, citations, and references.

·   To assess the authenticity and accuracy of citations and references generated by ChatGPT and Google Gemini.

Methodology

To evaluate the capabilities and limitations of ChatGPT and Google Gemini in generating academic content, the researchers selected two demo research topics: "Adoption of Artificial Intelligence in Libraries" and "Impact of Social Media Platforms on Library Services: An Assessment." These topics were chosen to represent diverse yet relevant themes in the field of Library and Information Science, allowing for a comprehensive assessment of how these AI tools perform across various sections of academic writing.

The study employed a set of predefined prompts to direct ChatGPT and Google Gemini in generating different sections of the research articles. The focus was on generating key components, including the **introduction**, **problem statement**, **research gaps**, **methodology**, **literature review** (inclusive of citations and references), **conclusion**, and **references**. Through this structured exploration, the study aimed to evaluate the effectiveness, coherence, and accuracy of both tools in generating these distinct elements of academic writing.

AI Versions Used:

· **ChatGPT version 3.5** was utilized for this study, as it represents a widely-used iteration of the model with proven capabilities in academic writing.

· **Google Gemini** (formerly Google Bard) was evaluated to compare its output to ChatGPT's, focusing on its capacity to generate coherent and relevant academic content.

Prompts Used: To guide the AI tools in generating each section of the research articles, the following prompts were used:

· **Introduction Prompt**: "Write an introduction for the research topic 'Research Topic' and provide the sub-sections: Background, Problem Statement, and Research Gap."

· **Literature Review Prompt**: "Write a literature review for the research topic 'Research Topic' with in-text citations and references in APA style."

· **Conclusion Prompt**: "Write a conclusion for the research topic 'Research Topic'."

After generating the articles, the output from each AI tool was analyzed for several key factors:

1. **Coherence and Completeness**: Each section was evaluated for logical flow, depth of content, and the clarity of arguments presented.

2. **Citations and References**: The citations and references generated by the AI tools were manually checked for authenticity, accuracy, and adherence to APA citation style.

3. **Plagiarism Detection**: Using the Turnitin plagiarism detection tool, the similarity ratio of the generated content was evaluated to identify any instances of potential plagiarism or over-reliance on existing sources.

4. **Content Structure**: The generated articles were also reviewed for proper structuring, including page length, organization, and how well the AI addressed the required sub-sections.

This methodology provides a structured and detailed evaluation of ChatGPT and Google Gemini, enabling the researchers to assess their capabilities in contributing to various stages of academic research writing, and to identify the critical challenges and ethical considerations in employing AI tools in scholarly contexts.

Data Analysis

After collecting the AI-generated content from both ChatGPT and Google Gemini (formerly Google Bard), a total of four research papers were generated: two from ChatGPT and two from Google Gemini. The analysis focused on several key aspects: the number of citations generated, the authenticity of citations, the similarity ratio, and the overall quality of the generated content.

Citation Analysis

Upon reviewing the citations generated by both tools, a minimal difference was found in the total number of citations: Google Gemini generated 18 citations across its two articles, while ChatGPT generated 17. However, a deeper inspection revealed significant issues with citation authenticity. All citations generated by ChatGPT, though properly formatted in APA style, were fabricated. In contrast, Google Gemini produced authentic citations in one of its articles, but the other article contained fabricated citations, and none of the citations adhered to proper APA formatting. These findings are summarized in Tables 1 and 2.

Table 1: Number of Citations Generated

Article No	Google Gemini	ChatGPT
1	9	8
2	9	9
Total	18	17

Table 2: Relevance of References Generated in Terms of APA Style (APA Format Compliance)

Article No	Google Gemini	ChatGPT
1	0 (0.0%)	8 (100%)
2	0 (0.0%)	9 (100%)
Total	0 (0.0%)	17 (100%)

Repetition of Citations

A further analysis was conducted to examine the frequency of repeated citations within individual articles. In one article generated by Google Gemini, only three unique sources were cited, each repeated three times, indicating a reliance on a small number of sources. Similarly, the other article by Google Gemini followed this pattern, generating content based on limited sources. In contrast, ChatGPT produced articles with 8 and 9 unique citations, each referencing different sources. This is outlined in Tables 3 and 4.

Table 3: Number of Double/Triple Repeated Citations in a Single Article

Article No	Google Gemini	ChatGPT
1	0 / 3 / 0	0 / 0 / 0
2	1 / 1 / 1	0 / 0 / 0
Total	1 / 4 / 1	0 / 0 / 0

Table 4: Number of References Generated (Source Articles Consulted)

Article No	Google Gemini	ChatGPT
1	3	8
2	3	9
Total	6	17

Similarity Index Analysis

Contrary to previous studies which indicated that 30-40% of AI-generated content tends to be plagiarized (Aydın & Karaarslan, 2022), the similarity ratios of the AI-generated content in this study were remarkably low. Three of the articles showed a similarity ratio of just 3%, while one article generated by Google Gemini showed a similarity ratio of 8%, as detailed in Table 5 and Figure 1.

Table 5: Percentage of Similarity Index of AI-Generated Content

Article No	Google Gemini	ChatGPT
1	3.0%	3.0%
2	8.0%	3.0%
Total	11%	6.0%

Detection of AI-Generated Content

Turnitin's AI detection tool was used to assess the degree to which the generated content could be recognized as AI-generated. Surprisingly, despite the content being fully generated by AI, Turnitin

detected that 77% to 94% of the content was AI-generated (see Table 6 and Figure 2). These results either raise questions about the efficacy of AI detection tools like Turnitin or demonstrate the sophistication of AI tools in generating content that can evade such detection systems.

Table 6: Percentage of AI-Generated Content Detected by Turnitin

Article No	Google Gemini	ChatGPT
1	94%	82%
2	77%	80%

Other Findings

· **Problem Statements**: Both AI tools generated problem statements that were entirely hypothetical and lacked supporting references. This suggests that neither ChatGPT nor Google Gemini is currently capable of identifying original research gaps from the literature, as they rely on generalizations rather than access to specific scholarly databases.

· **Research Gaps**: Since both ChatGPT and Google Gemini cannot access the majority of academic articles, they fail to provide proper references when identifying gaps in the literature. This presents a significant limitation when employing these tools for comprehensive research purposes.

Conclusion

This study aimed to evaluate the capabilities and limitations of AI-based Natural Language Generation (NLG) models—specifically ChatGPT and Google Gemini—in generating academic research articles. Through a comprehensive analysis of generated content, citations, similarity ratios, and ethical concerns, several critical insights were identified.

The results show that while both ChatGPT and Google Gemini can produce coherent and structured academic content, significant challenges arise in terms of citation authenticity and research quality. Both AI tools were found to fabricate citations and references, particularly ChatGPT, which consistently generated citations in correct APA style but with nonexistent sources. Google Gemini, on the other hand, produced authentic sources in some instances but struggled with citation formatting. Furthermore, both tools demonstrated an over-reliance on a limited number of sources, which undermines the comprehensiveness and diversity required in academic writing.

The study also found that the similarity ratio of AI-generated content was lower than expected, contradicting previous research that indicated a higher degree of plagiarism in AI-generated text. Nevertheless, this raises questions about the robustness of existing plagiarism detection tools such as Turnitin, particularly in detecting the full extent of AI-generated content.

In addition, neither ChatGPT nor Google Gemini could effectively identify specific research gaps, as they rely on generalized knowledge and lack access to current scholarly databases. This limitation is a significant drawback when employing these tools for generating high-quality academic research articles.

The findings emphasize the need for human supervision in utilizing AI tools for academic writing. While these models can assist in various aspects of content generation and research writing, their current limitations—such as the generation of fabricated citations and reliance on limited sources—indicate that they are not yet ready to be used autonomously. Ethical guidelines, stronger AI detection tools, and proper acknowledgment of AI usage are essential to ensure the integrity of academic research.

Overall, AI tools like ChatGPT and Google Gemini have the potential to support and enhance the research process, but their use must be carefully managed, with human researchers retaining full responsibility for the quality and authenticity of the work produced. As AI technology continues to evolve, so too must the ethical frameworks and detection tools that regulate its use in academia.

Suggestions

Based on the findings and conclusions of this study, the following suggestions are proposed to ensure the responsible use of AI tools like ChatGPT and Google Gemini in academic research:

1. **Human Supervision and Verification**: AI-generated content should always be thoroughly reviewed by human researchers. While AI can assist in generating drafts, identifying research gaps, or providing basic structure, human intervention is necessary to ensure accuracy, coherence, and the validity of sources and citations.
2. **Improved AI Detection Tools**: Current plagiarism detection tools, such as Turnitin, need to be enhanced to better identify AI-generated content. Developers should focus on integrating sophisticated AI detection capabilities into existing plagiarism software to keep up with advancements in AI language models.
3. **Clear Ethical Guidelines**: Academic institutions and publishers should establish and enforce clear guidelines for the use of AI in research. Researchers should be required to disclose the use of AI tools in the methodology or acknowledgment sections of their work, and policies should address the ethical implications of AI-generated content, including authorship, citation accuracy, and plagiarism risks.
4. **Strengthening AI Citation Practices**: As both ChatGPT and Google Gemini were found to fabricate citations, it is crucial to either improve these tools' access to legitimate databases or limit their use in generating references. Researchers should manually verify all citations and references generated by AI to avoid reliance on inaccurate or fabricated sources.
5. **AI as an Assistive Tool, Not a Substitute**: AI tools should be regarded as complementary resources rather than substitutes for genuine research and writing. Researchers should rely on AI to aid in brainstorming, structuring papers, or overcoming writer's block, but the critical and creative aspects of research should remain under the control of human authors.

6. **Training and Awareness for Researchers**: Researchers, especially early-career academics, should be trained on the potential risks and ethical considerations of using AI in research. Workshops, seminars, and guidelines should be provided to help researchers understand both the benefits and limitations of AI tools in academic writing.

7. **Future Research on AI Development**: Further studies should focus on improving the capabilities of AI models in generating accurate and ethical academic content. Research should explore ways to integrate AI with reliable scholarly databases and to minimize the generation of fabricated citations. Moreover, future research should address how AI tools can assist researchers more effectively in tasks such as literature reviews and data analysis without compromising the integrity of the research process.

Declarations

Author Contribution

All authors worked equally and are agreed with the content and that all gave explicit consent to submit and that they obtained consent from the responsible authorities at the institute/organization where the work has been carried out, before the work is submitted.

References

1. Aydın Ö, Karaarslan E (2022) OpenAI ChatGPT generated literature review: Digital twin in healthcare. Emerg Comput Technol 2(1):22–31

2. Aydın Ö, Karaarslan E (2022) OpenAI ChatGPT generated literature review: Digital twin in healthcare. Emerg Comput Technol 2(1):22–31

3. Brainard J (2023) As scientists explore AI-written text, journals hammer out policies. Science 379(6628):1478–1480. https://doi.org/10.1126/science.adh2937

4. ChatGPT Five priorities for research. Nature, 614(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

5. Chen TJ (2023) ChatGPT and other artificial intelligence applications speed up scientific writing. J Chin Med Association 86(4):383–388. https://doi.org/10.1097/JCMA.0000000000000900

6. Dowling M, Lucey B (2023) ChatGPT for (finance) research: The Bananarama conjecture. Finance Res Lett 53:103662. https://doi.org/10.1016/j.frl.2023.103662

7. Dowling M, Lucey B (2023) ChatGPT for (finance) research: The Bananarama conjecture. Finance Res Lett 53:103662. https://doi.org/10.1016/j.frl.2023.103662

8. Gao CA, Howard FM, Markov NS et al (2023) Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. Npj Digit Med 6:75. https://doi.org/10.1038/s41746-023-00819-6

9. Gao CA, Howard FM, Markov NS et al (2023) Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digit Med 6:75. https://doi.org/10.1038/s41746-023-00819-6

10. Gonsalves RA (2023), January 17 Using ChatGPT as a creative writing partner – part 2: Music. *Towards Data Science*. https://towardsdatascience.com/using-chatgpt-as-a-creative-writing-partner-part-2-music-d2fd7501c268

11. Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günnemann S, Hüllermeier E, Krusche E, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Stadler M, Kasnec G (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103(April 2023). https://doi.org/10.1016/j.lindif.2023.102274

12. Kim SG (2023) Using ChatGPT for language editing in scientific articles. Maxillofacial Plast Reconstr Surg 45(1):13. https://doi.org/10.1186/s40902-023-00381-x

13. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A (2023) Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. Lancet Digit Health 5(3):e105−e106. https://doi.org/10.1016/S2589-7500(23)00019-5

14. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A (2023) Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. Lancet Digit Health 5(3):e105−e106. https://doi.org/10.1016/S2589-7500(23)00019-5

15. Liu G, Yang L (2019) Study on research progress in academic fields. J Sci Technol Stud 48(6):112−124

16. O'Connor S, ChatGPT (2023) Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? Nurse Educ Pract 66:103537. https://doi.org/10.1016/j.nepr.2022.103537

17. Stokel-Walker C (2023) ChatGPT listed as author on research papers: Many scientists disapprove. Nature 613(7945):620−621. https://doi.org/10.1038/d41586-023-00107-z

18. Stokel-Walker C (2023) ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613(7945), 620−621. https://doi.org/10.1038/d41586-023-00107-z Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023)

19. Van Dis EA, Bollen J, Zuidema W, Van Rooij R, Bockting CL (2023) ChatGPT: Five priorities for research. Nature 614(7947):224−226. https://doi.org/10.1038/d41586-023-00288-7

20. Zhai X, ChatGPT User Experience : Implications for Education (December 27, 2022). Available at SSRN: https://ssrn.com/abstract=4312418 or http://dx.doi.org/10.2139/ssrn.4312418

Figures

Figure 1

Similarity Index of AI Generated Articles and their Sources

Figure 2

Percentage of AI generated Content Detected by Turnitin