

Exploring ChatGPT in Academic Information Retrieval: A Critical Review

A Review of:

Liu Z, Wang X, Li L. ChatGPT-Assisted Information Retrieval: A Comparative Study of User Behavior in Academic Information Retrieval. In: Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries. 2024: 1-3. Article 3702595. doi: 10.1145/3677389.3702595

Reviewed by: Shafigheh Mousavi^{1*} 

¹ Department of Medical Library and Information Science, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Article Type:

Critical Review Article

Article History:

Received: 19 Aug 2025

Revised: 23 Aug 2025

Accepted: 14 Sep 2025

ePublished: 17 Sep 2025

Keywords:

ChatGPT,
Information Retrieval,
User Behavior,
Scientific Search,
Critique and
Evaluation

Abstract

Objective. To compare scientific search behavior with and without ChatGPT assistance.

Design/Method. Comparative study with two randomly assigned groups of graduate students; five tasks based on cognitive levels.

Tools/Settings. ChatGPT for Google extension version 3.5 and web search; data collected via screen recording.

Metrics. Number of searches, query length, SERP clicks, SERP browsing time, number/time of URL visits, total retrieval time.

Key Findings. Reduced clicks, reduced number/time of URL visits, and reduced total retrieval time with ChatGPT; increased query length and SERP browsing time; significant differences mainly in T1, T2, and T5.

Conclusion. ChatGPT excels in recall, understanding, and creation tasks; limited effect on analysis and evaluation.

Mousavi S. Exploring ChatGPT in Academic Information Retrieval: A Critical Review. *Depiction of Health*. 2025; 16(3): xxx-xxx. doi: 10.34172/doh.2025.25. (Persian)

Extended Abstract

Introduction

In the current era, with the rapid expansion of artificial intelligence technologies, tools like ChatGPT are quickly changing patterns of human interaction with

information. Understanding how these tools affect scientific information retrieval and academic user behavior is increasingly important. The article "ChatGPT-Assisted Information Retrieval: A

* Corresponding author; Shafigheh Mousavi, E-mail: shafigheh.mousavi@gmail.com



Comparative Study of User Behavior in Academic Information Retrieval" (Liu et al., 2024), published in the proceedings of the 24th ACM/IEEE joint conference on digital libraries (JCDL'24), investigates fundamental differences in user behavior when searching for scientific information with ChatGPT assistance compared to traditional methods. The article is innovative in its topic and study design; however, if a larger sample had been included, its findings would have been more reliable for decision-making.¹

This study, utilized a structured approach and designed five tasks corresponding to the cognitive levels of "recall, understanding, analysis, evaluation, and creation" and compared the performance of two user groups (with and without ChatGPT). The findings indicate that ChatGPT can enhance search efficiency in some of these tasks.

Commentary

The study randomly assigned 20 first-year graduate students from various disciplines to two groups. The experimental group completed five search tasks designed according to the cognitive levels "recall, understanding, analysis, evaluation, and creation" using the ChatGPT for Google extension (version 3.5), while the control group performed the same tasks using a standard search engine. Behavioral data were collected via screen recording and manually annotated. Metrics included the number of searches, average query length, SERP (search engine results page) clicks, SERP browsing time, number of URLs selected, time spent on URLs, and total information retrieval time (seconds). The findings indicate that in all tasks, using ChatGPT reduced SERP clicks, the number and duration of URL visits, and total retrieval time. In contrast, the average query length and SERP browsing time increased with ChatGPT. Statistically significant differences were mainly observed in tasks T1 (recall), T2 (understanding), and T5 (creation). In the conclusion section, the authors highlighted ChatGPT's advantage in fact/concept-oriented tasks (recall and understanding) and content creation tasks.

This study provides valuable insights into the application of ChatGPT in optimizing information retrieval processes, highlighting its role in enhancing search efficiency and its integration with traditional search engines, which represents an important step toward understanding how AI tools can be incorporated into information-seeking practices. Nonetheless, details regarding participant selection were not provided, which may introduce potential bias in the results.²

The research tasks were designed based on the cognitive levels framework, enhancing conceptual coherence and allowing assessment of participant

performance across different cognitive domains. However, the lack of baseline assessment of participants' information retrieval skills prior to the study may have influenced the results and reduced the accuracy of comparisons.

The collection of actual user behavioral data, including click counts, time spent browsing web pages, and the number of URLs visited, strengthened the validity and reliability of the findings. Nonetheless, environmental and individual factors such as internet connectivity, participants' familiarity with the subject, or fatigue could have affected the outcomes and were not accounted for in the study.

Direct comparison of traditional search and ChatGPT-assisted search enabled objective and quantitative evaluation of the advantages and limitations of each approach. However, the study only included first-year master's students, and although it was mentioned that they were selected from various disciplines, the specific fields of study were not reported; this issue may limit the generalizability of the findings. The use of ChatGPT version 3.5 was clearly specified, which is justified given the study's timeline, providing an accurate picture of the tool's capabilities at that time. However, the fast development of newer ChatGPT versions means that the results only show the tool's status at the time of the study, and comparing with newer versions may be needed in future research.³

Overall, the structured task design, use of real behavioral data, and the simultaneous analysis of traditional and ChatGPT-assisted search are key strengths of the study. However, methodological limitations and sample selection issues identify areas for future improvements to increase validity and generalizability. Nevertheless, the study provides valuable insights into the potential of ChatGPT-assisted information retrieval and offers important guidance for future research.

Conclusion

The study by Liu et al. (2024) represents a valuable step toward examining the role and impacts of ChatGPT in scientific information retrieval. However, for future research, it is essential that researchers pay attention to certain limitations that may arise in the design, execution, and reporting of studies in order to provide more replicable and efficient results. Accordingly, it is recommended that future studies expand the sample size and include more diverse participant groups to examine the impact of ChatGPT across different levels of expertise and experience. Comparative evaluations of various versions of ChatGPT, especially newer ones, can also help identify the strengths and limitations of each version and offer practical guidance for optimal tool

selection. Assessing participants' initial search skills prior to the study and conducting qualitative evaluations of outputs by domain experts can enhance the accuracy and validity of the findings.² At the same time, considering intervening and contextual variables such as individual experience, type of query, familiarity with the subject, and level of information literacy, along with extending the scope of research applications to diverse domains such as medicine, law, education, and business, will provide a more comprehensive picture of ChatGPT's

effectiveness. Designing integrated and interconnected tasks, formulating clear hypotheses and assumptions, and using transparent evaluation criteria accompanied by reporting inter-rater reliability can further strengthen the theoretical and methodological framework of future studies and yield more reliable and practical results. However, future researchers need to consider some types of limitations that may occur in designing, executing, and reporting research in order to present more replicable and efficient results.^{4,5}

Proof

نگاهی انتقادی به بازیابی اطلاعات علمی با کمک چت جی پی تی

مروری نقادانه بر:

ژو لئو، شو یان وانگ، و لی لی. بازیابی اطلاعات با کمک ChatGPT: یک مطالعه مقایسه‌ای رفتار کاربران در بازیابی اطلاعات علمی. در: مجموعه مقالات بیست و چهارمین کنفرانس مشترک ACM/IEEE بر کتابخانه‌های دیجیتال. ۱۶ دسامبر ۲۰۲۴. صفحات ۱-۳. شماره مقاله: ۳۷۰۲۵۹۵
doi: [10.1145/3677389.3702595](https://doi.org/10.1145/3677389.3702595)

مرور کننده: شفیقه موسوی^{۱*}

^۱ گروه کتابداری و اطلاع‌رسانی پزشکی، دانشکده مدیریت و اطلاع‌رسانی پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

چکیده

اطلاعات مقاله

نوع مقاله:

مقاله مروری نقادانه

سابقه مقاله:

دریافت: ۱۴۰۴/۰۵/۲۸

اصلاح نهایی: ۱۴۰۴/۰۶/۰۱

پذیرش: ۱۴۰۴/۰۶/۲۳

انتشار برخط: ۱۴۰۴/۰۶/۲۶

کلیدواژه‌ها:

چت جی پی تی،
بازیابی اطلاعات،
رفتار کاربران،
جستجوی علمی،
نقد و ارزیابی

هدف. مقایسه رفتار جستجوی علمی با/ و بدون یاری ChatGPT

طرح/ روش. مطالعه مقایسه‌ای با دو گروه تصادفی از دانشجویان کارشناسی ارشد؛ ۵ وظیفه مبتنی بر سطوح شناختی **ابزار/ تنظیمات.** افزونه ChatGPT for Google نسخه ۳/۵ و جستجوی وب؛ گردآوری داده‌ها از طریق ضبط صفحه **سنجش‌ها.** تعداد جستجو، طول کوئری، کلیک‌های SERP، زمان مرور SERP، تعداد/ زمان بازدید URL‌ها، زمان کل بازیابی

یافته‌های کلیدی. کاهش کلیک‌ها، کاهش تعداد/ زمان بازدید URL و کاهش زمان کل با ChatGPT؛ افزایش طول کوئری و زمان مرور SERP؛ تفاوت‌های معنادار عمدتاً در T5 و T2، T1 و **جمع‌بندی.** برتری ChatGPT در وظایف یادآوری، فهمیدن و خلق؛ اثر محدود در تحلیل و ارزیابی

مقدمه

منتشر شده است و به واکاوی تفاوت‌های بنیادین در رفتار کاربران هنگام جستجوی اطلاعات علمی با یاری ChatGPT در مقایسه با روش‌های سنتی پرداخته است. با وجود این که مقاله در حیطه موضوعی خود دارای نوآوری در موضوع و طراحی مطالعه است؛ اما اگر حجم نمونه بزرگی را شامل می‌شد نتایج آن برای تصمیم‌گیری قابل اعتمادتر می‌بود.

این پژوهش با رویکردی ساختاریافته، پنج وظیفه متناسب با سطوح شناختی «یادآوری، درک، تحلیل، ارزیابی

در عصر حاضر و با گسترش روزافزون فناوری‌های هوش مصنوعی، ابزارهایی نظیر ChatGPT به سرعت در حال تغییر الگوهای تعامل انسان با اطلاعات هستند. در این میان، درک چگونگی تأثیر این ابزارها بر بازیابی اطلاعات علمی و رفتار کاربران دانشگاهی از اهمیت فزاینده‌ای برخوردار است. مقاله: بازیابی اطلاعات با کمک چت جی پی تی: یک مساله مقایسه‌ای رفتار اطلاعات در بازیابی اطلاعات علمی (لئو و همکاران، ۲۰۲۴) در کتابچه بیست و چهارمین کنفرانس مشترک ACM/IEEE بر کتابخانه‌های دیجیتال

* پدیدآور رابط: شفیقه موسوی، آدرس ایمیل: shafigheh.mousavi@gmail.com

و خلق» را طراحی و عملکرد دو گروه از کاربران (با/ و بدون بهره‌گیری از ChatGPT) را مورد مقایسه قرار داده است. یافته‌های این مطالعه نشان می‌دهد که به‌کارگیری ChatGPT قادر است کارایی فرآیند جستجو را در برخی از این وظایف ارتقاء بخشد.^۱

با توجه به نوظهور بودن ادغام ChatGPT در فرآیندهای جستجوی علمی و کاربرد مستقیم آن در حوزه علم کتابداری، بررسی و نقد آن برای محققین و حرفه‌مندان این رشته خالی از لطف نیست.

در مطالعه انجام گرفته در مقاله مورد نقد، بیست نفر از دانشجویان سال اول مقطع کارشناسی ارشد از رشته‌های مختلف به‌صورت تصادفی به دو گروه تقسیم شدند. گروه آزمایش با استفاده از افزونه ChatGPT for Google (نسخه 3.5)، پنج وظیفه جستجو را که مبتنی بر سطوح شناختی «یادآوری، درک، تحلیل، ارزیابی و خلق» طراحی شده بودند، به انجام رساندند و گروه کنترل نیز به جستجوی اطلاعات در حوزه پنج وظایف تعیین شده با بهره‌گیری از موتور جستجو اقدام کردند. داده‌های رفتاری کاربران از طریق ضبط صفحه نمایش جمع‌آوری و سپس به صورت دستی نشان‌گذاری شد. معیارهای اندازه‌گیری شامل تعداد جستجوها، متوسط طول جستجوها، تعداد کلیک‌های بر روی نتایج موتور جستجو (SERP: Search Engine Result Page)، مدت زمان جستجو در SERP، تعداد URLها انتخاب شده، مدت زمان بازدید از URLها و زمان کل فرایند بازیابی اطلاعات (ثانیه) بودند.

یافته‌ها حاکی از آن است که در تمامی وظایف، استفاده از ChatGPT منجر به کاهش تعداد کلیک‌های SERP، تعداد و مدت زمان بازدید از URLها و زمان کل بازیابی شده است. در مقابل، متوسط طول سوال پرس و جو (Query) و مدت زمان مرور SERP در حالت استفاده از ChatGPT افزایش نشان می‌دهد. تفاوت‌های معنادار آماری عمدتاً در وظایف T1 (یادآوری)، T2 (درک) و T5 (خلق) بوده‌اند. نویسندگان در بخش نتیجه‌گیری، برتری ChatGPT را برای وظایف واقعیت‌محور/ مفهومی (مانند یادآوری و درک) و همچنین وظایف خلق محتوا، برجسته‌تر گزارش کرده‌اند.

یادداشت

این مطالعه بینش‌های ارزشمندی درباره کاربرد ChatGPT در بهینه‌سازی فرآیند بازیابی اطلاعات ارائه

می‌دهد و نقش آن را در ارتقای کارایی جستجو و جایگاه آن در مدل‌های ترکیبی با موتورهای جستجو روشن می‌کند، که گامی مهم در درک چگونگی ادغام ابزارهای هوش مصنوعی در فرآیندهای جستجوی اطلاعات است، با این حال جزئیات نحوه انتخاب شرکت‌کنندگان ارائه نشده است که می‌تواند موجب سوگیری احتمالی نتایج شود.^۲

وظایف پژوهش مبتنی بر چارچوب سطوح شناختی طراحی شده‌اند و این امر انسجام مفهومی مطالعه را افزایش داده و امکان ارزیابی عملکرد شرکت‌کنندگان در حوزه‌های مختلف شناختی را فراهم کرده است، با این حال، عدم ارزیابی مهارت‌های پایه شرکت‌کنندگان پیش از مطالعه ممکن است بر نتایج تأثیر بگذارد و دقت مقایسه‌ها را کاهش دهد.

جمع‌آوری داده‌های رفتاری واقعی کاربران، شامل تعداد کلیک‌ها، مدت زمان مرور صفحات وب و تعداد URLهای بازدید شده، اعتبار و روایی یافته‌ها را افزایش داده است، با این حال، عوامل محیطی و فردی مانند کیفیت اتصال اینترنت، آشنایی شرکت‌کنندگان با موضوع و خستگی آنها می‌توانست بر نتایج تأثیرگذار باشد و مطالعه این موارد را در نظر نگرفته است.

مقایسه مستقیم کارایی جستجوی سنتی و جستجوی یاری‌گرفته شده از ChatGPT امکان ارزیابی عینی و کمی مزایا و محدودیت‌های هر روش را فراهم کرده است، با این حال، مطالعه صرفاً شامل دانشجویان سال اول کارشناسی ارشد بوده با این حال، مطالعه صرفاً شامل دانشجویان سال اول کارشناسی ارشد بوده و با وجود اشاره به انتخاب از رشته‌های مختلف، جزئیات مربوط به رشته‌ها ذکر نشده است؛ موضوعی که می‌تواند تعمیم‌پذیری یافته‌ها را با محدودیت مواجه سازد.

استفاده از نسخه ۳/۵ ChatGPT با توجه به زمان انجام پژوهش قابل توجیه است و یافته‌ها تصویری واقعی از ابزار در زمان مطالعه ارائه می‌دهند، با این حال، پیشرفت سریع نسخه‌های جدیدتر این ابزار باعث می‌شود نتایج فقط بازتاب وضعیت گذشته باشد و بررسی نسخه‌های جدیدتر در مطالعات آینده ضروری است.^۳

در مجموع، طراحی ساختار یافته وظایف، استفاده از داده‌های رفتاری واقعی و بررسی همزمان روش سنتی و یاری‌گرفته شده از ChatGPT از نقاط قوت مطالعه محسوب می‌شوند، با این حال، محدودیت‌های روش شناختی و

و تجربه بررسی شود. همچنین ارزیابی تطبیقی نسخه‌های مختلف ChatGPT، به‌ویژه نسخه‌های جدیدتر، می‌تواند نقاط قوت و محدودیت‌های هر نسخه را مشخص کرده و راهنمایی عملی برای انتخاب بهینه ابزار ارائه دهد. سنجش مهارت‌های اولیه شرکت‌کنندگان پیش از مطالعه و ارزیابی کیفی خروجی‌ها توسط کارشناسان حوزه‌های مرتبط می‌تواند دقت و اعتبار نتایج را افزایش دهد.^۲ در عین حال، توجه به متغیرهای مداخله‌گر و زمینه‌ای مانند تجربه فردی، نوع پرسش، آشنایی با موضوع و سطح سواد اطلاعاتی، و گسترش دامنه کاربردی پژوهش به حوزه‌های متنوع مانند پزشکی، حقوق، آموزش و تجارت، تصویری جامع‌تر از کارایی ChatGPT ارائه خواهد کرد. طراحی وظایف یکپارچه و مرتبط، تدوین فرضیه‌ها و مفروضات کلیدی و استفاده از معیارهای ارزیابی شفاف همراه با گزارش توافق بین ارزیابان، می‌تواند چارچوب نظری و روش‌شناختی پژوهش‌های آینده را تقویت کند و یافته‌های قابل اتکا و کاربردی‌تری ارائه نماید.^{۴،۵}

انتخاب نمونه، نکاتی هستند که باید در مطالعات بعدی مورد توجه قرار گیرند تا اعتبار و تعمیم‌پذیری یافته‌ها افزایش یابد. با این وجود، پژوهش دیدگاه‌های مهمی درباره پتانسیل ارزیابی اطلاعات با کمک ChatGPT ارائه می‌دهد و راهنمایی ارزشمندی برای تحقیقات آینده فراهم می‌کند.

جمع‌بندی

مطالعه لیو و همکاران گامی ارزشمند در راستای بررسی نقش و تأثیرات ChatGPT در فرآیند ارزیابی اطلاعات علمی به شمار می‌آید. با این حال، برای پژوهش‌های آینده ضروری است که پژوهشگران به برخی از محدودیت‌هایی که ممکن است در طراحی، اجرا و گزارش پژوهش رخ دهد توجه کنند تا نتایجی تکرارپذیرتر و کارآمدتر ارائه شود. از این رو توصیه می‌شود در پژوهش‌های آینده دامنه نمونه افزایش یابد و گروه‌های شرکت‌کننده متنوع‌تر انتخاب شوند تا تأثیر ChatGPT در سطوح مختلف تخصص

References

1. Liu Z, Wang X, Li L. ChatGPT-assisted information retrieval: A comparative study of user behavior in academic information retrieval. In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'24). 2024; 1-3. doi: 10.1145/3677389.3702595
2. Bujang MA, Sa'at N, Sidik TMITAB, Joo LC. Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malays J Med Sci*. 2018; 25(4): 122-130. doi: 10.21315/mjms2018.25.4.12
3. Rojas M, Rojas M, Burgess V, Toro-Pérez J, Salehi S. Exploring the Performance of ChatGPT Versions 3.5, 4, and 4 With Vision in the Chilean Medical Licensing Examination: Observational Study. *JMIR Med Educ*. 2024; 10: 1-6. doi: 10.2196/55048
4. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)*. 2012; 22(3): 276-282. doi: 10.11613/BM.2012.031
5. Cole R. Inter-rater reliability methods in qualitative case study research. *Sociol Methods Res*. 2023; 53(4): 1944-1975. doi: 10.1177/00491241231156971