PDF Download
3726302.3730140.pdf
10 February 2026
Total Citations: 0
Total Downloads: 1733

Published: 13 July 2025

Citation in BibTeX format

SIGIR '25: The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval
July 13 - 18, 2025
Padua, Italy

Conference Sponsors:
SIGIR

DEMONSTRATION

A Flexible User Study Platform for Generative Information Retrieval

YIDONG LIANG, Beijing Institute of Technology, Beijing, China

ZHIJING WU, Beijing Institute of Technology, Beijing, China

YUCHEN HE, Beijing Institute of Technology, Beijing, China

FENGMING LIANG, Renmin University of China, Beijing, China

KEXIN LIU, Renmin University of China, Beijing, China

JIAXIN MAO, Renmin University of China, Beijing, China

Open Access Support provided by:

Renmin University of China

Beijing Institute of Technology

SIGIR '25: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (July 2025)
https://doi.org/10.1145/3726302.3730140
ISBN: 9798400715921

A Flexible User Study Platform for Generative Information Retrieval

Yidong Liang
School of Computer Science and Technology, Beijing Institute of Technology
Beijing, China
lyd@bit.edu.cn

Zhijing Wu[*]
School of Computer Science and Technology, Beijing Institute of Technology
Beijing, China
zhijingwu@bit.edu.cn

Yuchen He
School of Computer Science and Technology, Beijing Institute of Technology
Beijing, China
2084532709@qq.com

Fengming Liang
Gaoling School of Artificial Intelligence, Renmin University of China
Beijing, China
2544997551@qq.com

Kexin Liu
Gaoling School of Artificial Intelligence, Renmin University of China
Beijing, China
1378084216@qq.com

Jiaxin Mao[*]
Gaoling School of Artificial Intelligence, Renmin University of China
Beijing, China
maojiaxin@gmail.com

Abstract

User behavior and experience are important for improving information retrieval (IR) systems. While much research has focused on traditional IR systems, few studies have systematically examined user behavior and search experience with emerging generative IR systems. A key reason for this gap is the lack of publicly available toolkits to record user behavior and feedback in generative IR systems. We developed a comprehensive platform to collect user behavior and feedback on the generative IR system. This platform consists of: 1) a generative IR system that supports both API-based and customized retrieval-augmented generation (RAG) methods, 2) a user interface that logs various user behavior, including prompts, clicks, mouse movements, and scrolling, and 3) an annotation website that allows users to provide feedback. We believe the proposed platform has the potential to streamline data collection for user studies on generative IR systems, paving the way for future research on how users engage with and interact with these systems.

CCS Concepts

• **Information systems** → *Users and interactive retrieval.*

Keywords

Generative Information Retrieval Systems, User Behavior, User Study Platform

ACM Reference Format:
Yidong Liang, Zhijing Wu, Yuchen He, Fengming Liang, Kexin Liu, and Jiaxin Mao. 2025. A Flexible User Study Platform for Generative Information Retrieval. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3726302.3730140

1 Introduction

Information retrieval (IR) systems are essential tools for effectively accessing information. They help users quickly obtain specific resources they are searching for. One of the key goals of an IR system is to understand and satisfy users' information needs. However, user search intent and experience (e.g., satisfaction) are subjective and typically cannot be directly captured by the system. Instead, user behavior (e.g., mouse movement and click), which is more accessible, is regarded as implicit feedback reflecting user intent and experience. Therefore, user behavior and search experience have long been central topics in the IR community, and substantial research has been conducted on traditional IR systems [8, 13, 23].

Recently, the development of large language models (LLMs) has led to the emergence of generative IR systems (e.g., Perplexity.ai , Copilot , You Chat and Gemini). Generative IR systems generate responses with citations to answer user queries by either retrieving relevant documents to enhance a generative model with the retrieved information (retrieval-augmented generation), or by initially producing a response and then verifying or retrieving supporting evidence to attach appropriate citations (generation-augmented retrieval) [9]. As shown in Figure 3(b), users interact with generative IR systems through conversation. Users submit queries, and systems respond with generated text with citations. The interface of generative IR systems is completely different from the traditional "ten blue links", which has changed user behavior patterns.

User studies in the IR community usually involve collecting user behavior and explicit feedback and analyzing data [4, 16, 18]. Researchers often need to develop experimental platforms for data collection. However, developing an experimental platform from scratch requires labor and resources, which can be inefficient. This issue has been taken into account in traditional IR systems. Li et al. [14] has developed a logging tool, which enables researchers to collect user behaviors and explicit feedback on common search engines. However, there is no similar tool for generative IR systems, which may hinder the research on behavior modeling, search

[*] Corresponding author

SIGIR '25, Padua, Italy
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730140

experience analysis, and evaluation in this emerging IR scenario. We believe that a well-designed and publicly available user study platform is essential to advance research in this area.

In this work, we develop a user study platform for generative IR systems to collect user behavior and explicit feedback. Our platform consists of three main components:

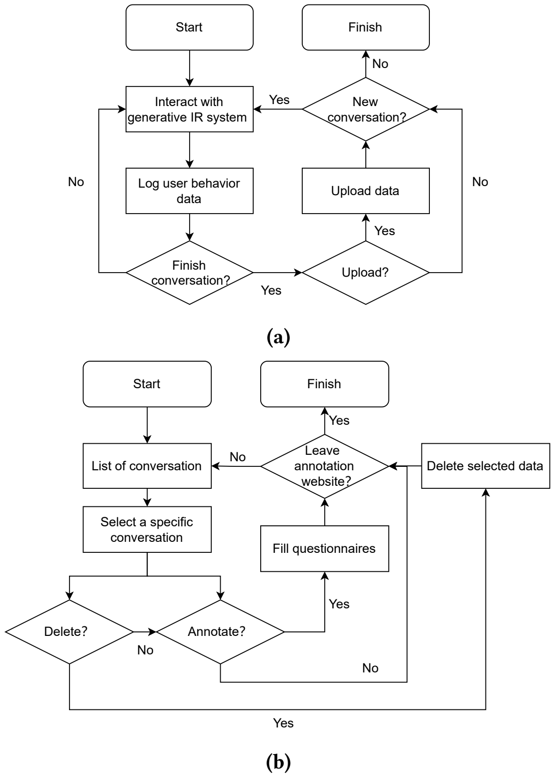
- A generative IR system that generates summarized text responses with citation links. The system is scalable and supports the use of APIs from existing generative IR systems. It also offers the ability to customize corpora and provides flexibility in selecting retrieval and generation algorithms in the RAG pipeline.
- A user interface where users interact with the generative IR system. User behaviors on the website, such as mouse movements, clicks, query reformulations, and timestamps, are recorded. Users can choose whether to upload data, which helps protect their privacy. Researchers can configure the layout of the interface components, such as whether to display the citation link list. They can also define which user behaviors to record.
- An annotation website for collecting explicit feedback, where users annotate data and provide feedback, such as satisfaction and link usefulness. The website supports researchers in customizing the annotation content to suit their specific research needs.

To the best of our knowledge, this is the first publicly available user study platform designed for the generative IR system scenario, including an extensible generative IR system, user interface and annotation website. Researchers can conduct various studies using the platform. For example, researchers can modify the interface to study design impacts and adjust the retrieval-then-generation pipeline to analyze the effects of different retrieval corpora, algorithms, and generation methods on users. The source code is available at https://github.com/BITLYDG/GUS-Platform.

2 Related Work

Research on user behavior and search experience contributes to various tasks within the field of traditional IR. User behavior and search experience help to design better interfaces [3, 11, 30], improve ranking algorithms [1, 15, 21], and inspire the design of evaluation metrics [5, 7, 20]. For example, adding an answer module to the search engine results page (SERP) can help users complete their search tasks quickly with less effort [27]. Implicit feedback, such as clicking and browsing, effectively improves ranking performance [1, 19]. In addition, user behavior can be used for evaluation [28]. Research on user behavior and search experience is typically conducted through user studies [2, 6, 10], where participants complete search tasks using platforms developed by researchers in a controlled lab environment. On traditional IR systems, a unified general recorder for user behavior and experience has already been developed [14].

On generative IR systems, user-centric research remains limited and typically focuses on specific scenarios or phenomena [17, 24–26, 29]. For example, generative IR systems pose risks such as increasing selective exposure, creating echo chambers, and contributing to opinion polarization [22]. Gienapp et al. [9] provided theoretical insights into the user model and evaluation on generative IR systems.



(a)

(b)

Figure 1: The workflow of users interacting with the (a) generative IR system / (b) annotation website.

The lack of user studies on generative IR systems limits understanding of user behavior and search experience. We developed a user study platform to promote user-centric research.

3 System Design

3.1 Functional design

The platform needs to have three main functions: 1) implementing a generative IR system; 2) recording user behavior and conversation content; and 3) collecting user explicit feedback. In this section, we explain the design of three functions.

Generative IR system: A generative IR system generates summarized text responses with citation links. First, the platform should support established APIs for generative IR systems. Additionally, in user studies, researchers may explore the effects of specific system variables, such as retrieval or generation models. Therefore, the generative IR module should be extensible, allowing researchers to modify configurations as needed. Furthermore, the interface of the system should be similar to commercial generative search engines.

User behavior and conversation content collection: The platform should automatically collect various user behaviors. The collection function needs to maintain high accuracy and stability. Our platform allows users to experiment remotely to get a more realistic user behavior and search experience. The behavior collected should be reliably transmitted to the back-end. To minimize the impact on user behavior and search experience, behavior recording

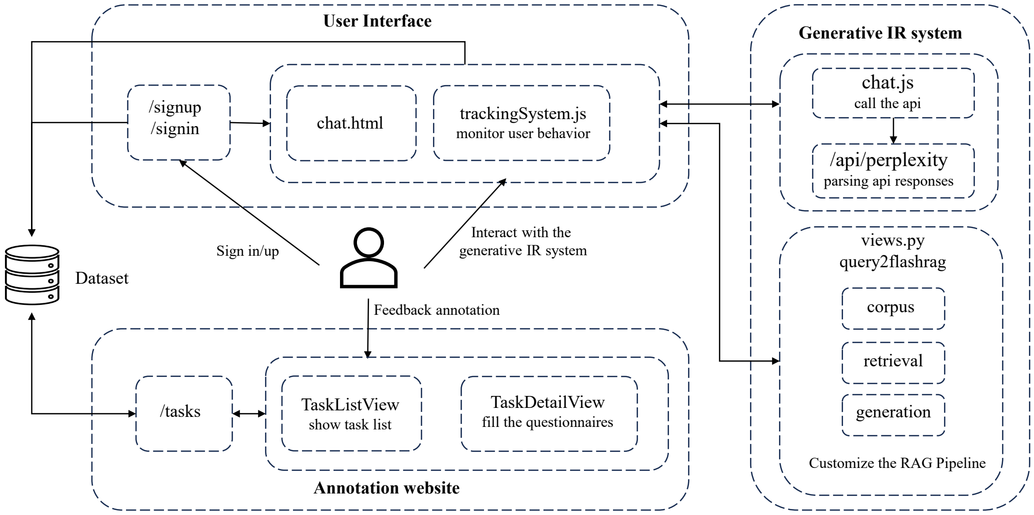


Figure 2: Framework of the platform.

should be as unobtrusive as possible. The data is collected based on the conversation level. The data is saved when the user starts a new conversation or closes the page. We have collected the following types of user behavior and conversation content:

- Behavior: The platform automatically records user behaviors such as mouse movements, scrolling, dwell time, and clicks on the generative IR system. The timestamp for these behaviors is also recorded. For click behaviors, the URLs, text, and position information of clicked links are recorded.
- Conversation content: Beyond user behavior data, conversation content is also important. The platform records queries and responses generated by the generative IR system. It comprehensively captures textual content and link-related information within the responses. In addition, timestamps corresponding to queries and responses are recorded. Some meta-information (e.g., user device and browser) is also recorded.

Explicit feedback collection: In addition to behavior and search data, our platform contains an annotation website to collect explicit feedback from users. In user studies, explicit feedback plays an important role, such as user satisfaction, perceived usefulness of links, and users' search intent. The annotation website will show all user-submitted conversations and indicate whether they are annotated. If the user does not want to keep a conversation, he can delete it. The annotation interface for each conversation will present the recorded queries and responses with query-level and conversation-level feedback questionnaires. Researchers can configure the questionnaires to collect the necessary feedback.

3.2 User interaction workflow

The interaction between users and the platform is shown in Figure 1. After logging into the platform, users interact with the generative IR system. The user-system interaction is similar to that in commercial generative search engines, where users submit queries and receive text responses with citation links. The platform automatically records both user behavior and the conversation content. When the user restarts a new conversation or closes the page, the data from the previous conversation will be uploaded to the database. For the conversation that the user does not want to share with us, the user can also choose not to upload it, which can protect the user's privacy and increase their willingness to use the platform. Users accessing the annotation platform can view all uploaded conversations along with the annotation status of each one. Users can delete any conversation they do not want to share, or click on a conversation to enter its annotation interface. In the annotation interface, users fill out query-level and session-level questionnaires to provide explicit feedback.

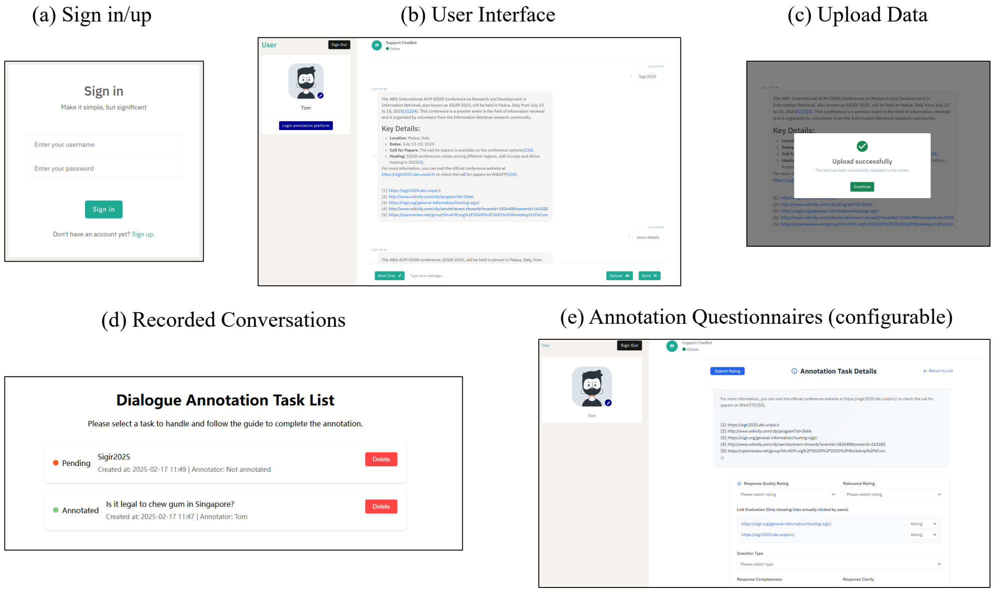
4 Implementation

Our platform is a Django project developed using Python. The architecture of the platform is shown in Figure 2. The main components are a generative IR system, a user interface, and an annotation website. In this section, we will describe the implementation of these three parts in detail.

Generative IR system: On the one hand, we call the API and parse the response to implement streaming output. In the public project, we implemented a reliable generative IR system using Perplexity.ai API as an example. On the other hand, we provide a retrieval-and-generate pipeline. This pipeline, based on flashRAG[1] [12] , allows researchers to configure the retrieval model, generation model, and corpus more flexibly.

User interface: The front-end of the user interface mainly includes the HTML files of the interface and the JavaScript files that record user behavior and conversation content. These JavaScript files define various behavior recording functions and initialization functions, which can be easily called. The low coupling and high cohesion between the HTML files of the interface and the JavaScript files used for data collection make it easy for researchers to modify and customize any one of them. The back-end mainly includes a

[1] https://github.com/RUC-NLPIR/FlashRAG


(a) Sign in/up

Sign in

Make it simple, but significant

Enter your username

Enter your password

Don't have an account yet? Sign up.

(b) User Interface

Sigir2025

The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, also known as SIGIR 2025, will be held in Padua, Italy from July 13 to 18, 2025[1][2][4]. This conference is a premier event in the field of information retrieval and is organized by volunteers from the Information Retrieval research community.

Key Details:

- **Location**: Padua, Italy
- **Dates**: July 13-18, 2025
- **Call for Papers**: The call for papers is available on the conference website[2][4].
- **Hosting**: SIGIR conferences rotate among different regions, with Europe and Africa hosting in 2025[3].

For more information, you can visit the official conference website at https://sigir2025.dei.unipd.it/ or check the call for papers on WikiCFP[2][4].

[1]: https://sigir2025.dei.unipd.it
[2]: http://www.wikicfp.com/cfp/program?id=2666
[3]: https://sigir.org/general-information/hosting-sigir/
[4]: http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=182648&ownerid=163282
[5]: https://openreview.net/group?id=ACM.org%2FSIGIR%2F2025%2FWorkshop%2FeCom

more details

The 48th ACM SIGIR conference, SIGIR 2025, will be held in person in Padua, Italy, from

(c) Upload Data

The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, also known as SIGIR 2025, will be held in Padua, Italy from July 13 to 18, 2025[1][2][4]. This conference is a premier event in the field of information retrieval and is organized by volunteers from the Information Retrieval research community.

Key Details:

[2]: http://www.wikicfp.com/cfp/program?id=2666
[3]: https://sigir.org/general-information/hosting-sigir/
[4]: http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=182648&ownerid=163282
[5]: https://openreview.net/group?id=ACM.org%2FSIGIR%2F2025%2FWorkshop%2FeCom

Upload successfully

The data has been successfully uploaded to the server.

(d) Recorded Conversations

Dialogue Annotation Task List

Please select a task to handle and follow the guide to complete the annotation.

● Pending Sigir2025
Created at: 2025-02-17 11:49 | Annotator: Not annotated

● Annotated Is it legal to chew gum in Singapore?
Created at: 2025-02-17 11:47 | Annotator: Tom

(e) Annotation Questionnaires (configurable)

For more information, you can visit the official conference website at https://sigir2025.dei.unipd.it/ or check the call for papers on WikiCFP[2][4].

[1]: https://sigir2025.dei.unipd.it
[2]: http://www.wikicfp.com/cfp/program?id=2666
[3]: https://sigir.org/general-information/hosting-sigir/
[4]: http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=182648&ownerid=163282
[5]: https://openreview.net/group?id=ACM.org%2FSIGIR%2F2025%2FWorkshop%2FeCom

Response Quality Rating: Please select rating
Relevance Rating: Please select rating

Link Evaluation (Only showing links actually clicked by users)

https://sigir.org/general-information/hosting-sigir/ Rating
https://sigir2025.dei.unipd.it/ Rating

Question Type: Please select type

Response Completeness
Response Clarity

Figure 3: Some main interfaces of the platform.

database that stores conversation data and user information, as well as the above-mentioned generative IR system.

Annotation website: The annotation website shares a database with the user interface. The data collected on the user interface is uploaded to the database. The annotation website obtains the data from the database and presents the conversation content on the annotation interface. The annotation interface reuses the user interaction interface to avoid inaccurate explicit feedback caused by the large gap between the two interfaces. We set up query-level and conversation-level questionnaires on the annotation interface, and users provide explicit feedback by filling out these questionnaires. Overall, the annotation website extends the user interface by incorporating questionnaire display and recording, along with some modifications to the interaction.

5 Demonstration

In this section, we will demonstrate the functions implemented by our user study platform. Figure 3 shows the main interface of the user study platform.

When users first access the platform, they are required to register a new account (Figure 3(a)). The registration form can also be easily modified according to the demand for user information. Once logged in, users are redirected to the user interaction interface (Figure 3(b)), where they interact with the generative IR system according to the instructions from researchers or their daily usage habits. At the same time, the website will record user behavior and conversation. Researchers can customize the content they wish to collect by modifying the JavaScript files. After a conversation ends, users can click the "upload" button to upload the data (Figure 3(c)). Users will be reminded whether to upload data when starting a new conversation or closing the page. On the annotation website, users can view the conversation records and have the option to annotate or delete conversations they do not wish to share (Figure 3(d)). Considering that a single conversation may include multiple queries and responses, we have designed both query-level and conversation-level feedback questionnaires to collect more comprehensive explicit feedback (Figure 3(e)). These feedback questionnaires can also be easily customized by modifying Django files. Users can view and annotate their submitted data at any time. Since our platform is in the form of websites, it allows for remote user studies. Researchers can access the user behavior and explicit feedback from the back-end database.

6 Conclusion

User-centric research has been the focus of research in the IR community. User behavior and search experience have been extensively studied in traditional IR systems. These studies have contributed to the development of IR, such as SERP design and system evaluation. There is still insufficient research on user behavior and search experiences in generative IR systems, a novel paradigm that has emerged in recent years. We developed a user study platform that includes a generative IR system, a user interface, and an annotation website. This platform allows researchers to collect user behavior and search experiences on generative IR systems. The platform can be flexibly configured for different user studies. Users can conduct experiments remotely through our platform, which is not restricted to laboratory settings. To the best of our knowledge, this is the first user study platform designed for generative IR systems. We believe the platform will facilitate user-centric research in this area.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62302040), the China Postdoctoral Science Foundation (No. 2022TQ0033), and the Beijing Institute of Technology Research Fund Program for Young Scholars.

References

[1] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 19–26.

[2] Ioannis Arapakis, Xiao Bai, and B. Barla Cambazoglu. 2014. Impact of response latency on user behavior in web search. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 103–112. doi:10.1145/2600428.2609627

[3] Ioannis Arapakis, Luis A Leiva, and B Barla Cambazoglu. 2015. Know your onions: understanding the user experience with the knowledge module in web search. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 1695–1698.

[4] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology* 67, 11 (2016), 2635–2651.

[5] Olivier Chapelle, Donald Metlzer, Ya Zhang, and Pierre Grinspan. 2009. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM conference on Information and knowledge management*. 621–630.

[6] Jia Chen, Jiaxin Mao, Yiqun Liu, Fan Zhang, Min Zhang, and Shaoping Ma. 2021. Towards a better understanding of query reformulation behavior in web search. In *Proceedings of the web conference 2021*. 743–755.

[7] Aleksandr Chuklin, Pavel Serdyukov, and Maarten De Rijke. 2013. Click model-based information retrieval metrics. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 493–502.

[8] Michael J Cole, Chathra Hendahewa, Nicholas J Belkin, and Chirag Shah. 2015. User activity patterns during information search. *ACM Transactions on Information Systems* 33, 1 (2015), 1–39.

[9] Lukas Gienapp, Harrisen Scells, Niklas Deckers, Janek Bevendorff, Shuai Wang, Johannes Kiesel, Shahbaz Syed, Maik Fröbe, Guido Zuccon, Benno Stein, Matthias Hagen, and Martin Potthast. 2024. Evaluating Generative Ad Hoc Information Retrieval. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC, USA) *(SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 1916–1929. doi:10.1145/3626772.3657849

[10] Jiepu Jiang, Daqing He, and James Allan. 2014. Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 607–616.

[11] Jimmy Jimmy, Guido Zuccon, Bevan Koopman, and Gianluca Demartini. 2019. Health cards for consumer health search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 35–44.

[12] Jiajie Jin, Yutao Zhu, Xinyu Yang, Chenghao Zhang, and Zhicheng Dou. 2024. FlashRAG: A Modular Toolkit for Efficient Retrieval-Augmented Generation Research. *CoRR* abs/2405.13576 (2024). arXiv:2405.13576 https://arxiv.org/abs/2405.13576

[13] Diane Kelly and Leif Azzopardi. 2015. How many results per page? A Study of SERP Size, Search Behavior and User Experience. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 183–192. doi:10.1145/2766462.2767732

[14] Hanyu Li, Hongyu Lu, Songhao Huang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2021. Privacy-Aware Remote Information Retrieval User Experiments Logging Tool. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 2615–2619. doi:10.1145/3404835.3462793

[15] Xiangsheng Li, Jiaxin Mao, Chao Wang, Yiqun Liu, Min Zhang, and Shaoping Ma. 2019. Teach machine how to read: reading behavior inspired relevance estimation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 795–804.

[16] Mengyang Liu, Yiqun Liu, Jiaxin Mao, Cheng Luo, Min Zhang, and Shaoping Ma. 2018. "Satisfaction with Failure" or "Unsatisfied Success": Investigating the Relationship between Search Success and User Satisfaction. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1533–1542. doi:10.1145/3178876.3186065

[17] Sijie Liu, Yuyang Hu, Zihang Tian, Zhe Jin, Shijin Ruan, and Jiaxin Mao. 2024. Investigating Users' Search Behavior and Outcome with ChatGPT in Learning-oriented Search Tasks. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*. 103–113.

[18] David Maxwell, Leif Azzopardi, and Yashar Moshfeghi. 2017. A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 135–144. doi:10.1145/3077136.3080824

[19] Rishabh Mehrotra, Imed Zitouni, Ahmed Hassan Awadallah, Ahmed El Kholy, and Madian Khabsa. 2017. User interaction sequences for search satisfaction prediction. In *Proceedings of the 40th International ACM SIGIR conference on research and development in information retrieval*. 165–174.

[20] Alistair Moffat and Justin Zobel. 2008. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Syst.* 27, 1, Article 2 (Dec. 2008), 27 pages. doi:10.1145/1416950.1416952

[21] Taesup Moon, Georges Dupret, Shihao Ji, Ciya Liao, and Zhaohui Zheng. 2010. User behavior driven ranking without editorial judgments. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1473–1476.

[22] Nikhil Sharma, Q. Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1033, 17 pages. doi:10.1145/3613904.3642459

[23] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. 2018. User intent, behaviour, and perceived satisfaction in product search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 547–555.

[24] Siddharth Suri, Scott Counts, Leijie Wang, Chacha Chen, Mengting Wan, Tara Safavi, Jennifer Neville, Chirag Shah, Ryen W. White, Reid Andersen, Georg Buscher, Sathish Manivannan, Nagu Rangan, and Longqi Yang. 2024. The Use of Generative Search Engines for Knowledge Work and Complex Tasks. *CoRR* abs/2404.04268 (2024). doi:10.48550/ARXIV.2404.04268 arXiv:2404.04268

[25] Johanne R. Trippas, Sara Fahad Dawood Al Lawati, Joel Mackenzie, and Luke Gallagher. 2024. What do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC, USA) *(SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 2703–2707. doi:10.1145/3626772.3657914

[26] Albatool Wazzan, Stephen MacNeil, and Richard Souvenir. 2024. Comparing Traditional and LLM-based Search for Image Geolocation. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval* (Sheffield, United Kingdom) *(CHIIR '24)*. Association for Computing Machinery, New York, NY, USA, 291–302. doi:10.1145/3627508.3638305

[27] Zhijing Wu, Mark Sanderson, B. Barla Cambazoglu, W. Bruce Croft, and Falk Scholer. 2020. Providing Direct Answers in Search Results: A Study of User Behavior. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1635–1644. doi:10.1145/3340531.3412017

[28] Emine Yilmaz, Milad Shokouhi, Nick Craswell, and Stephen Robertson. 2010. Expected browsing utility for web search evaluation. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (Toronto, ON, Canada) *(CIKM '10)*. Association for Computing Machinery, New York, NY, USA, 1561–1564. doi:10.1145/1871437.1871672

[29] Ines Zelch, Matthias Hagen, and Martin Potthast. 2024. A User Study on the Acceptance of Native Advertising in Generative IR. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval* (Sheffield, United Kingdom) *(CHIIR '24)*. Association for Computing Machinery, New York, NY, USA, 142–152. doi:10.1145/3627508.3638316

[30] Jie Zou, Mohammad Aliannejadi, Evangelos Kanoulas, Maria Soledad Pera, and Yiqun Liu. 2023. Users meet clarifying questions: Toward a better understanding of user interactions for search clarification. *ACM Transactions on Information Systems* 41, 1 (2023), 1–25.