Tasks, Copilots, and the Future of Search: A Keynote at SIGIR 2023

Ryen W. White
Microsoft Research
Redmond, WA, USA
ryenw@microsoft.com

Abstract

Search is far from being a solved problem. While search engines may cope well with simple tasks, searchers and systems struggle as task complexity increases. Task is central to the search process, motivating the search and driving search behavior. Complex search tasks require more than support for rudimentary fact finding or re-finding. Various support options have been offered by search systems over time (e.g., query suggestions, contextual search) to help search engine users more effectively tackle complex tasks. The recent emergence of generative artificial intelligence (AI) and the arrival of assistive agents, or *copilots*, based on this technology, has the potential to offer further assistance to searchers, especially those engaged in complex tasks. The implications from these advances for the design of intelligent systems and for the future of search itself are significant. This overview of the keynote that I gave at the 2023 ACM SIGIR Conference introduces AI copilots and briefly presents some of the challenges and opportunities for researching, developing, and deploying search copilots.

Date: 26 July 2023.

1 Tasks in Search

Tasks drive the search process. The information retrieval (IR) and information science communities have long studied tasks in search [Shah et al., 2023b] and many information seeking models consider the role of task directly [Belkin, 1980; Dervin, 1998], including task complexity [Byström and Järvelin, 1995]. Recent estimates suggest that half of all Web searches are not answered.[1] Many of those searches are connected to complex search tasks. These tasks are ill-defined and/or multi-step, span multiple queries, sessions, and/or devices, and require deep engagement with search engines (many queries, backtracking, branching, etc.) to complete them [Hassan Awadallah et al., 2014]. Complex tasks also often have many facets and cognitive dimensions, and are closely connected to searcher characteristics such as domain expertise and task familiarity [Sarkar and Shah, 2021; White, 2018]; what is complex to one searcher, may not be complex to another.

[1] https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/

To date, there have been significant attempts to support complex search tasks via human intermediaries (e.g., librarians, subject matter experts) and search systems. Research on technologies to support complex tasks includes work on generating query and website suggestions [Hassan Awadallah et al., 2014; White et al., 2007], personalizing and contextualizing search [Bennett et al., 2012], and developing new search experiences, including those that span time and space [White et al., 2019; Agichtein et al., 2012]. We are now also seeing emerging search-related technologies in the area of generative AI [Najork, 2023] and in conversational search [Gao et al., 2023]. These more recent advances in particular have opened the door for copilots to finally realize our long-held vision of assistive agents that can guide searchers through the search process, augment their capabilities, and improve search task outcomes.

2 AI Copilots

Copilots[2] are applications of modern AI (e.g., foundation models such as GPT-4[3] and DALL·E 3[4]) to help people with complex cognitive tasks. Copilots have conversational user interfaces and copilot users engage with copilots via natural language, they are powered by foundation models, copilots are extensible with skills/tools/plugins, and they are scoped to specialized domains or applications (including search). Copilots are intentionally designed to keep humans firmly at the center of the task completion process to help them complete a broader range of tasks with higher quality outcomes in less time. We are also seeing an emergence of multi-agent systems that can use several specialized copilots working together, and with humans, to help people complete their tasks more efficiently and effectively [Wu et al., 2023].

Integrating copilots into search engines is complex, both on the algorithmic side and in terms of the user experience. Algorithmically, the search engine uses the top-ranked results for one or more auto-generated queries to ground the response of the copilot given a searcher request (prompt) [Lewis et al., 2020]. The answer generated can be supplemented or replaced by instant answers (weather, flights, etc.) and search results from the search index. These are shown to searchers as either two separate modalities (e.g., traditional search and conversational search, as in the case of Bing Chat[5]) or a single experience that unifies search and copilot results (in the case of Google Bard[6]). Figure 1 shows the conversational experience in the Bing Chat copilot. Answers from the copilots can also be integrated into the search engine result pages (SERPs) for selected informational queries. Provenance (knowing the source) is critical for building searcher trust and searchers will often need to verify generated answers before using them, ideally without leaving the search experience to preserve cognitive flow. Links to sources can be shown alongside answers (Figure 1). These links also help content creators, publishers, and advertisers, and incentivizes the continued content creation that is critical for developing future foundation models and copilots.

Copilots and conversational experiences are best viewed as a complement not a replacement to traditional search engines. Search engines have been around for decades and provide near

[2]Copilots is the Microsoft terminology for an AI-powered digital assistant designed to work alongside humans to augment their cognitive capabilities.

[3]https://openai.com/gpt-4

[4]https://openai.com/dall-e-3

[5]http://www.bing.com/chat; Recently rebranded as Microsoft Copilot.

[6]http://bard.google.com

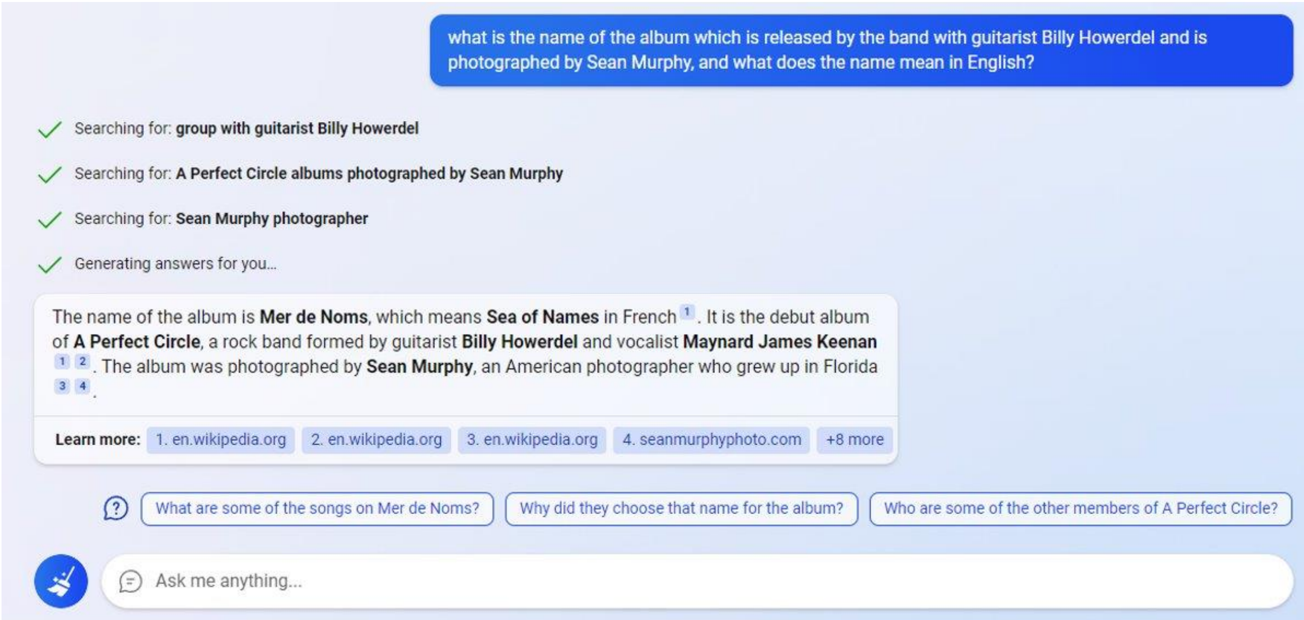

Figure 1. Screenshot of the conversational experience in Bing Chat, illustrating the richness of the requests that can be sent to the copilot, the internal queries that are generated by the system using a foundation model (in this case, GPT-4), and the completeness of the answer generated by that same model. Such a comprehensive answer would have taken a considerable amount of human effort to obtain with a traditional search engine, spanning at least several queries and several landing page visits.

instantaneous access to answers and resources for many intents. These existing and emerging modalities can and should work well together to help searchers tackle a wider range of tasks. The capabilities of copilots to better understand intentions and provide assistance beyond fact finding and basic learning/investigation will expand the task frontier, broadening the range of complex search tasks that can be completed, e.g., direct support for tasks requiring creative inspiration (Figure 2). This all moves us a step closer toward intelligent search systems that can help with all-task completion, comprising the universe of tasks for which people might need search support.

3 Challenges

Despite the promise of copilots, there are significant challenges with them that should be acknowledged and we must find ways to overcome. Those include issues with:

- *Hallucinations and Biases:* The copilot output shown in response to searcher requests, which may be skewed, garbled, or inaccurate. This includes the tendency of generative AI to hallucinate (generate incorrect or non-sensical responses), be affected by biases [Liang et al., 2021; Ji et al., 2023], and affect searchers' own biases [Leffer, 2023];
- *Learning:* The impacts that the copilots can have on searchers, especially on their learning processes. Learning is a core part of the search process and happens when people engage with result content [Marchionini, 2006; Vakkari, 2016]. By generating/synthesizing answers

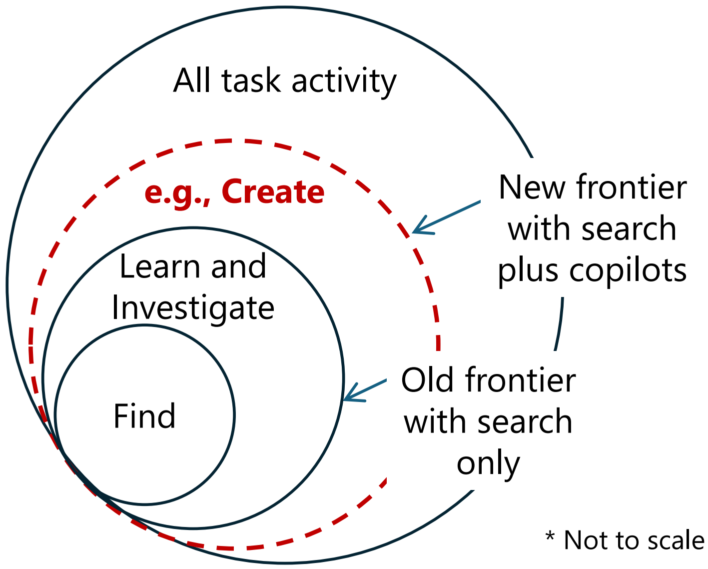

Figure 2. Visualizing the set of possible tasks that are possible with search only today (predominantly finding and learning/investigation) plus the expansion in the task frontier with search plus copilots (e.g., emerging AI support for creative inspiration, among other complex cognitive activities).

and reducing the need for searchers to engage with content directly, copilots are changing the learning experience during search, and;

- *Control:* Shifts in the degree of agency that humans have in the search process. Searchers want control and they want to understand how and why certain answers were provided [Bates, 1990]. When copilots are used, there is less control over answer generation, but more control over task/goal specification via more expressive natural language inputs and more strategic, less tactical search operations by searchers in general.

All of these need to be better understood. There are other challenges that are less related to the technology and are more about the usage of search engines, e.g., changing deeply ingrained habits that people may have for using search only (not copilots too), using keyword search only (not natural language), and using specific search engines that do not offer AI copilot functionalities.

4 Opportunities

For some time, members of the IR community have argued that the future of information access will involve personal search assistants with advanced capabilities, including natural language input, rich sensing, user/task/world models, and reactive and proactive experiences [White, 2016]. Some of these aspirations have thus far been out of reach, at least at the levels of performance required for large-scale deployment in flagship products. However, through the recent emergence of foundation models, technology is catching up with this vision. AI copilots will shape the future of information access and there are opportunities going forward in at least four areas:

- *Model Innovation:* Various advances in better modeling tasks so that copilots can represent searchers' intentions and goals, collecting feedback data to help align copilot outputs with searchers' preferences and values, augmenting model inputs using retrieval results and other forms of grounding and knowledge injection to generate more accurate copilot answers, and specializing models to specific tasks or applications and applying these models in a staged manner to minimize copilot scaling costs [Zhang et al., 2023].
- *Next-Generation Experiences:* Unified experiences spanning search and chat, including recommendations of copilot settings (e.g., foundation model temperature) and search modalities given task demands, supporting human learning, increasing automation while maintaining user agency and transparency, and expanding the scope of tasks that copilots can take all the way to completion, considering costs and benefits to searchers [Broder and McAfee, 2023].
- *Measurement:* Studying the performance of the copilots in various task settings (defining appropriate metrics, etc.), applying the underlying foundation models for tasks such as relevance judgment [Faggioli et al., 2023], intent classification [Shah et al., 2023a], and building more realistic simulations of searcher behavior to help more fully evaluate search systems [White et al., 2005], and understanding the searcher (e.g., mental models and changes in search behavior), the tasks where copilots perform best (and worst), and the capabilities of the copilots themselves.
- *Broader Implications:* Ensuring that copilots act responsibly with appropriate guardrails, exploring different business models, especially with an expansion in the task frontier into new types of task support, and ensuring that the negative impact on content creators is minimized (i.e., no "paradox of reuse" [Vincent, 2022], where fewer incentives for content creators mean less new content is created and hence less new data to train future foundation models), capitalizing on data and feedback from broad integration of copilots in existing applications, and various societal issues (e.g., sustainability, privacy, policy).

The opportunities are plentiful and there are likely many more than those outlined briefly in this section. The IR community should embrace this timely opportunity to take the lead on answering many of these questions. Their implications could be far reaching, extending beyond the realms of information access and into other domains such as productivity, education, and scientific discovery.

5 Concluding Remarks

It is clear that copilots will have a significant impact on how people access, manage, and use information. Copilots will augment and empower users to complete more of their tasks more accurately and more quickly. Task completion has been called the "last mile" of search interaction [White, 2018] and any progress from serving lists of starting points (SERP contents) to supporting the completion of more tasks is most welcome. Copilots will help with complex tasks in particular and we need to balance their use with the use of search engines, given search habits that have built up over decades and the massive costs involved in serving generative AI at scale. Careful search interface design is also required to help people quickly understand copilot capabilities and to unify search and copilots to simplify the search experience and preserve flow. Mechanisms to recommend the best modality for the current task would also help searchers utilize this new technology more effectively. These promising advances need to be balanced against implications for human learning and human control [Vakkari, 2016; Shneiderman, 2022], with AI safety as a critical underpinning, and adequate attention on societal issues from privacy to policy (and more).

6 Addendum

The slides for the keynote are available online.[7] This article has only scratched the surface of the broad range issues in this important area and there is not room to cover them in more detail in this extended abstract. Please see [White, 2023] for an in-depth examination on many of the topics covered here and more.

Acknowledgements

I wish to thank the three General Chairs of the ACM SIGIR 2023 Conference, Hsin-Hsi Chen, Wei-Jou (Edward) Duh, and Hen-Hsen Huang, for their kind invitation to give this keynote and for being such gracious hosts.

References

Eugene Agichtein, Ryen W White, Susan T Dumais, and Paul N Bennett. Search, interrupted: Understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 315–324, 2012.

Marcia J Bates. Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5):575–591, 1990.

Nicholas J Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5(1):133–143, 1980.

Paul N Bennett, Ryen W White, Wei Chu, Susan T Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. Modeling the impact of short- and long-term behavior on search personalization. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 185–194, 2012.

Andrei Z Broder and Preston McAfee. Delphic costs and benefits in web search: A utilitarian and historical analysis. *arXiv preprint arXiv:2308.07525*, 2023.

Katriina Byström and Kalervo Järvelin. Task complexity affects information seeking and use. *Information Processing & Management*, 31(2):191–213, 1995.

Brenda Dervin. Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2):36–46, 1998.

[7] http://ryenwhite.com/talks/pptx/WhiteSIGIR2023.pptx

Guglielmo Faggioli, Laura Dietz, Charles LA Clarke, Gianluca Demartini, Matthias Hagen, Claudia Hauff, Noriko Kando, Evangelos Kanoulas, Martin Potthast, Benno Stein, and Henning Wachsmuth. Perspectives on large language models for relevance judgment. In *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 39–50, 2023.

Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. *Neural Approaches to Conversational Information Retrieval*. Springer Nature, 2023.

Ahmed Hassan Awadallah, Ryen W White, Patrick Pantel, Susan T Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM CIKM International Conference on Information and Knowledge Management*, pages 829–838, 2014.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.

Lauren Leffer. Humans absorb bias from AI–and keep it after they stop using the algorithm. *Scientific American*, October 2023. URL https://www.scientificamerican.com/article/humans-absorb-bias-from-ai-and-keep-it-after-they-stop-using-the-algorithm/. Accessed: 2023-11-04.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, and Ruslan Salakhutdinov. Towards understanding and mitigating social biases in language models. In *Proceedings of the International Conference on Machine Learning*, pages 6565–6576, 2021.

Gary Marchionini. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

Marc Najork. Generative information retrieval. In *Proceedings of the 46th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1, 2023.

Shawon Sarkar and Chirag Shah. An integrated model of task, information needs, sources and uncertainty to design task-aware search systems. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 83–92, 2021.

Chirag Shah, Ryen W White, Reid Andersen, Georg Buscher, Scott Counts, Sarkar Snigdha Sarathi Das, Ali Montazer, Sathish Manivannan, Jennifer Neville, Xiaochuan Ni, Nagu Rangan, Tara Safavi, Siddharth Suri, Mengting Wan, Leijie Wang, and Longqi Yang. Using large language models to generate, validate, and apply user intent taxonomies. *arXiv preprint arXiv:2309.13063*, 2023a.

Chirag Shah, Ryen W White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar, and Nicholas Belkin. Taking search to task. In *Proceedings of the 2023 ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 1–13, 2023b.

Ben Shneiderman. *Human-Centered AI*. Oxford University Press, 2022.

Pertti Vakkari. Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1):7–18, 2016.

Nicholas Vincent. The paradox of reuse, language models edition, 2022. URL https://nmvg.mataroa.blog/blog/the-paradox-of-reuse-language-models-edition/. Accessed: 2023-11-04.

Ryen W White. *Interactions with Search Systems*. Cambridge University Press, 2016.

Ryen W White. Opportunities and challenges in search interaction. *Communications of the ACM*, 61(12):36–38, 2018.

Ryen W White. Navigating complex search tasks with AI copilots. *arXiv preprint arXiv:2311.01235*, 2023.

Ryen W White, Ian Ruthven, Joemon M Jose, and CJ Van Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems*, 23(3): 325–361, 2005.

Ryen W White, Mikhail Bilenko, and Silviu Cucerzan. Studying the use of popular destinations to enhance web search interaction. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 159–166, 2007.

Ryen W White, Adam Fourney, Allen Herring, Paul N Bennett, Nirupama Chandrasekaran, Robert Sim, Elnaz Nouri, and Mark J Encarnación. Multi-device digital assistance. *Communications of the ACM*, 62(10):28–31, 2019.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*, 2023.

Jieyu Zhang, Ranjay Krishna, Ahmed H Awadallah, and Chi Wang. EcoAssistant: Using LLM assistant more affordably and accurately. *arXiv preprint arXiv:2310.03046*, 2023.