CHAPTER 11.

GENERATIVE AI CAN DO MY RESEARCH FOR ME ✨ *RESEARCHERS SHOULD ALWAYS EVALUATE AND CONTEXTUALIZE SEARCH RESULTS*

Leslie Allison, Tiffany DeRewal, and Amy Reed
Rowan University

According to a 2023 survey of generative artificial intelligence (GenAI) users ages 14–22, the most common use for GenAI is as a search tool (Hopelab et al., 2024). Notably, this survey was conducted *before* many tech companies began trumpeting the search capabilities of AI. As of this writing, Google, Meta, Microsoft, and OpenAI have all integrated or plan to integrate GenAI search tools into their existing platforms, all making some version of a promise that GenAI is, in the words of OpenAI, a "faster and easier" internet search tool than a traditional search engine (OpenAI, 2024).

As any college student who has written a term paper knows, searching for relevant information can be one of the most time-consuming aspects of research. Who wouldn't want a tool that could make finding information "faster and easier"? Instead of merely producing a ranked list of web links, GenAI search tools generate tidy outline or paragraph overviews in response to queries. Many even provide weblinks to accompany the generated text. If students experiment with one of any number of free GenAI search tools available today, they might surmise that they don't need to provide precise keywords to get relevant and accessible responses, nor are they directed to paywalled content they cannot access. Indeed, it's tempting to think of GenAI as a super-charged, one-stop research solution.

Yet, there are other important criteria besides speed and ease-of-use that help information-seekers determine whether a particular research tool is worthwhile. In this chapter, we argue that the most important criteria for evaluating a research tool is *transparency*. If researchers need to "check the work," does the tool make it possible to do so? All search tools—but especially

DOI: https://doi.org/1010.37514/PER-B.2026.2777.2.11

GenAI search tools—are fallible, so it's crucial that humans retain agency as researchers. GenAI search tools increase the risk of obtaining inaccurate and irrelevant information, while at the same time stripping the context that might help a researcher evaluate a particular source. Furthermore, GenAI search tools operate via different, opaque processes, meaning that *how* and *why* they get things wrong is different, depending on the tool. For this reason, we advocate for research practices that allow users to retain control—making decisions about what information is relevant and credible.

IDENTIFYING CREDIBLE, RELEVANT, AND COMPLETE INFORMATION IS THE CENTRAL GOAL OF RESEARCH

When we search for information, we want to know that the information we find is accurate or credible and that it actually answers the question at hand. Unfortunately, current GenAI tools commonly fail on these accounts (Allison & DeRewal, 2024; Shah & Bender, 2024).

Take accuracy, for example: perhaps you remember when Google's AI confidently recommended adding glue to pizza (Robison, 2024)? This AI "hallucination" made the rounds online because it was so absurd and immediately recognizable as false information. While such examples are humorous and show the limits of GenAI search, as writing instructors, we worry that such laughable examples falsely give the impression that AI errors are easy to spot. Such an impression fundamentally mischaracterizes information-seeking processes. Queries about whether glue belongs on pizza don't replicate genuine informational searches: most people already know that glue is inedible. Real research is driven by inquiry—a need to ask a question that may not have a clear answer (ACRL Board, 2016). For example, what happens when AI-augmented search results recommend mixing dangerous chemicals (Turner, 2024) or voting at a nonexistent polling location (Nelson & Angwin, 2024)?

While all search tools can provide inaccurate, irrelevant, and/or incomplete information, GenAI search tools put researchers at a double disadvantage because of their lack of transparency. First, as Shyam Sharma (2026) explains, these tools lack transparency because they are designed to sound credible and authoritative, even when the information they provide is inaccurate, biased, or incomplete. This "impression of omniscience," as Sharma explains, makes it easy to be duped, especially when seeking new information. Second, the search processes used by GenAI are opaque, distancing researchers from original sources (Shah & Bender, 2024) and harming researchers' ability to evaluate and contextualize information. Understanding the basics of how GenAI search tools work can help researchers make strategic choices about whether and when these tools fit a research need, an essential skill recommended in the MLA *Student Guide to AI Literacy* (2024).

HOW DOES AI-AUGMENTED SEARCH (NOT) WORK?

GenAI chatbots and search tools are built on foundation models that make statistical predictions based on word associations calibrated over extremely large data sets. These tools cannot understand context, nor can they determine accuracy (Shah & Bender, 2024). And while the most recent GenAI tools combine their underlying large language model (LLM) technology—the foundation models that generate text—with external website content retrieval, which allows the tools to access current web sources, this development has not solved problems with accuracy, relevance, or completeness.

Contemporary GenAI search tools commonly add weblinks to LLM-generated content through a process known as retrieval-augmented generation, or RAG (Tay, 2024). In the RAG process, current websites are crawled and indexed, becoming additional sources of data for the LLM: "semantically relevant" text retrieved from those web sources is fed to the LLM to augment a user's query, and weblinks can also be appended to the generated text output (Gienapp et al., 2024; Tay 2024). While GenAI companies claim that RAG ensures "authoritative" responses from "top-tier" sources, both anecdotal and experimental evidence shows that the RAG process of extracting text from indexed sources to augment a prompt does not ensure that the generated output will be accurate or that it will correspond to the supplied weblinks (Ho, 2024; "How," 2024; OpenAI, n.d.). These tools produce:

- Inaccurate information (Gwon et al., 2024; Jaźwińska & Chandrasekar, 2025; Liu et al., 2023)
- Incorrect/misattributed citations (Jaźwińska & Chandrasekar, 2024, 2025; Memon & West, 2024)
- Citations to deficient sources (Jaźwińska & Chandrasekar, 2024)
- Responses that distort, overgeneralize, or misrepresent the information from original sources (Jaźwińska & Chandrasekar, 2024; Memon & West, 2024; Peters & Chin-Yee, 2025)

Importantly, such problems persist even when GenAI tools are provided with more data from original sources, prompted to limit their search to particular types of sources (such as "only look for scholarly sources"), instructed to "behave" in certain ways (such as "act like a PhD-level researcher"), or used with "deep reasoning" capabilities (Chen et al., 2025; Liu et al., 2024; Magesh et al., 2025; Peters & Chin-Yee, 2025).

Significantly, the RAG process *decontextualizes* information from web sources: when text is extracted from a source to augment a prompt, there is no guarantee that the selected text reflects its authors' intentions or the situation in which they were writing (Memon & West, 2024). As any good researcher knows, context is *essential* to understanding, evaluating, and applying a source of information effectively. Thus, even with the ongoing fine-tuning and enhancements of contemporary GenAI search tools, it is not surprising that the decontextualizing process of RAG, layered upon the decontextualizing process by which LLMs draw on their pretraining to generate text, continues to yield unreliable results. These processes produce content that persuasively *mimics* the end-product of an authentic, successful research experience, even as they subvert that experience through their very design, inhibiting transparency and limiting the researcher's ability to evaluate the content being generated.

What does this mean for a student researcher? While a GenAI search system may provide end users with links to accompany generated text, those links do *not* mean that the text represents an accurate summary of those sources, that any kind of "fact-checking" has occurred, or that the text is related to the links at all. Furthermore, users have no way of knowing how, or how many, internet sources were indexed to produce those links, nor can they be certain those links will direct them to sites that are regarded as trustworthy or relevant sources of information.

RECLAIMING AGENCY IN THE SEARCH PROCESS

Even as we recognize the deficiencies of GenAI search tools, we also acknowledge that they are becoming increasingly difficult to avoid. Major tech companies aim to make GenAI pervasive throughout all stages of search, and we are approaching an era where any user searching for digital sources may be required to engage with a chatbot-like interface, even when using traditional library databases such as JSTOR or EBSCO (EBSCO, 2023; Guthrie & LaPensee, 2024). Yet researchers still have agency in determining whether or how they use the information that these tools provide. Here are a few guidelines for conducting research in the era of GenAI:

Take the Word "Research" Literally—Re-Search—and Consult Multiple Search Tools

The best researchers don't rely on one tool, nor stop at one search query. Rather than relying on one tool, such as Elicit, Google, or even JSTOR or Project Muse, effective researchers work across multiple platforms. No matter the search tool, they look for sources over and over again, using different prompts and keywords, asking different questions, and exploring different angles. A good researcher will spend a lot of time just looking for sources—and reading them critically—before starting to draw any conclusions. This process, of course, can be hard to do when up against a deadline. But good information-seeking practices are, at their core, recursive processes, not linear.

Know the Search Capabilities of the Product Being Used, and Prioritize Platforms that Value Transparency

To evaluate the comprehensiveness, relevance, and credibility of research output, researchers need to know what sources the GenAI product has access to and how it accesses them. GenAI tools are less transparent on both counts than traditional search tools. For example, it's a common assumption that all of the GenAI search tools available today can browse the internet and provide live web links. In truth, the browsing capabilities of major LLM-based tools, such as ChatGPT, Microsoft Copilot, and Claude, are highly variable, constantly shifting, and often dependent on version model, subscription, or login status. Effective use of any of these search tools requires knowing their relationship to web browsing, a relationship not always made clear to the user. Likewise, not every search platform has transparent practices for indexing sources. Perplexity states it relies on "trusted news sources, academic papers, and established blogs," without elaborating on how they determine what counts as trusted or established (Perplexity, n.d.). In contrast, JSTOR publishes lists of every scholarly journal it hosts, allowing researchers to know where knowledge gaps might lie (JSTOR, n.d.; JSTOR, 2021).

Be Wary of Summarization Features

GenAI search tools claim to summarize and synthesize sources on your behalf, providing you a fast, easy answer. Yet of all GenAI research features, source summarization is perhaps the most shaky. When scholarly AI platforms like Elicit summarize scholarly papers, for instance, they are often limited to summarizing the abstract, rather than the article's full text (Elicit, 2024). That means that users are not provided a comprehensive summary of a full paper, but merely a summary of a summary. Furthermore, GenAI scholarly search tools' ability to interpret context and domain-specific knowledge is prone to error. Elicit notes, for instance, that the "models aren't explicitly trained to be faithful to a body of text by default," and as such, "can miss the nuance of a paper" (Elicit, 2024). It's important to always keep in mind that the technological process underlying these tools, called automatic text summarization (or ATS), is still nascent, and cannot perform at the same level as human summarization (Luo et al., 2024; Shakil et al., 2024).

Ask a Human

Sometimes, the best methods of finding sources for research don't involve search engines at all. Asking a librarian, teacher, or expert in your field for recommendations can be one of the "faster and easier" ways to find high-quality sources. Because librarians and subject matter experts intimately understand field-specific research inquiries, a key component of the ACRL framework, they can open up lines of questions that more novice researchers would not have known to ask. These questions, in turn, can lead more directly to relevant sources, many of which may be housed offline, in proprietary databases, or specialized indexes that GenAI search products simply cannot access. This process can be invaluable to any researcher, regardless of skill level.

Over the coming decade, GenAI will play an increasingly larger role in our search processes, but at the end of the day, research is a deeply human, social act of asking questions about our world. We encourage building search habits that will retain researcher agency, keeping humans central to evaluating the credibility and quality of the information.

REFERENCES

Association of College and Research Libraries Board. (2016, January 11). *Framework for information literacy for higher education*. https://www.ala.org/sites/default/files/acrl/content/issues/infolit/Framework_ILHE.pdf

Allison, L., & DeRewal, T. (2024, October). Where knowledge begins? Generative search, information literacy, and the problem of friction. *Critical AI, 2*(2), https://doi.org/10.1215/2834703X-11556038

Chen, Y., Benton, J., Radhakrishnan, A., Uesato, J., Denison, C., Schulman, J., Somani, A., Hase, P., Wagner, M., Roger, F., Mikulik, V., Bowman, S. R., Leike, J., Kaplan, J., & Perez, E. (2025). *Reasoning models don't always say what they think*. arXiv. http://arxiv.org/abs/2505.05410

EBSCO. (2023, September 21). *EBSCO Information Services pursues generative artificial intelligence (AI) opportunities*. [Press release]. https://tinyurl.com/bdbz8d8a

Elicit. (2024). *Information and advice from the Elicit team*. Elicit Help Center. https://support.elicit.com/en/articles/549569

Gienapp, L., Scells, H., Deckers, N., Bevendorff, J., Wang, S., Kiesel, J., Syed, S., Fröbe, M., Zuccon, G., Stein, B., Hagen, M., & Potthast, M. (2024). Evaluating generative ad hoc information retrieval. *Proceedings of the 47th International*

ACM SIGIR Conference on Research and Development in Information Retrieval, 1916–1929.

Guthrie, K., & LaPensee, B. (2024, June 25). Generative AI on JSTOR: What we're learning from early usage, real-world applications, and user feedback. *JSTOR Blog*. https://about.jstor.org/blog/generative-ai-on-jstor-what-were-learning/

Gwon, Y. N., Kim, J. H., Chung, H. S., Jung, E. J., Chun, J., Lee, S., & Shim, S. R. (2024). The use of generative AI for scientific literature searches for systematic reviews: ChatGPT and Microsoft Bing AI performance evaluation. *JMIR Medical Informatics 12*(1), e51187.

Ho, V. (2024, June 20). *Why AI sometimes gets it wrong—And big strides to address it*. Microsoft Source. https://news.microsoft.com/source/features/company-news/why-ai-sometimes-gets-it-wrong-and-big-strides-to-address-it/

Hopelab, Center for Digital Thriving at Harvard Graduate School of Education, & Common Sense Media. (2024). *Teen and young adult perspectives on generative AI: Patterns of use, excitements, and concerns* (pp. 1–35).

https://www.commonsensemedia.org/sites/default/files/research/report/teen-and-young-adult-perspectives-on-generative-ai.pdf

Jaźwińska, K., & Chandrasekar, A. (2024, November 27). *How ChatGPT search (mis) represents publisher content*. Columbia Journalism Review. https://www.cjr.org/tow_center/how-chatgpt-misrepresents-publisher-content.php

Jaźwińska, K., & Chandrasekar, A. (2025, March 6). *AI search has a citation problem*. Columbia Journalism Review. https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php

JSTOR. (2021). *Editorial review process. Support for publishers*. https://support.publishers.jstor.org/hc/en-us/articles/360044318434-Editorial-Review-Process

JSTOR. (n.d.). *Multidiscipline Archive Collections. About JSTOR*. Retrieved December 13, 2024, from https://about.jstor.org/librarians/journals/multi-discipline/

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics, 12*, 157–173.

Liu, N., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 7001–7025). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.467

Luo, M., Xue, B., & Niu, B. (2024) A comprehensive survey for automatic text summarization: Techniques, approaches, and perspectives. *Neurocomputing*, *603*, 128280.

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-Free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies, 22*(2), 216–242.

Memon, S. A., & West, J. D. (2024, February 18). *Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable*. Center for an Informed Public. https://www.cip.uw.edu/2024/02/18/search-engines-chatgpt-generative-artificial-intelligence-less-reliable/

Modern Language Association of America. (2024). *Student guide to AI literacy*. MLA Style Center. https://style.mla.org/student-guide-to-ai-literacy/

Nelson, A., & Angwin, J. (2024, March 8). Beware of AI apps this election season, they could mislead you. *The Los Angeles Time*s. https://www.latimes.com/opinion/story/2024-03-08/primaries-voting-elections-ai-misinformation-plaforms-chatgpt

OpenAI. (n.d.). *Retrieval augmented generation (RAG) and semantic search for GPTs*. OpenAI Help Center. Retrieved December 13, 2024, from https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts

Perplexity. (n.d.). *Advice and Answers from the Perplexity Team*. Retrieved December 13, 2024, from https://www.perplexity.ai/hub/faq/how-does-perplexity-work

Perplexity. (n.d.). *What is Perplexity?* Retrieved December 13, 2024, from https://www.perplexity.ai/hub/faq/what-is-perplexity

Peters, U., & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. *Royal Society Open Science, 12*(4), 241776.

Robison, K. (2024, May 23). *Google promised a better search experience—Now it's telling us to put glue on our pizza. The Verge*.
https://www.theverge.com/2024/5/23/24162896/google-ai-overview-hallucinations-glue-in-pizza

Shah, C., & Bender, E. M. (2024). Envisioning information access systems: What makes for good tools and a healthy web? *ACM Trans. Web, 18*(3), 33:1–33:24.

Shakil, H., Farooq, A., & Kalita, J. (2024). Abstractive text summarization: State of the art, challenges, and improvements. *Neurocomputing, 603*, 128255.

Sharma, S. (2026). AI knows everything: AI can perpetuate ignorance, prejudices, and epistemic-rhetorical harms globally. In C. Basgier, A. Mills, M. Olejnik, M. Rodak, & S. Sharma (Eds.), *Bad ideas about AI and writing: Generative practices for teaching, learning, and communication*. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2026.2777.2.02

Tay, A. (2024, May 21). Retrieval Augmented Generation and academic search engines—Some suggestions for system builders. *Aaron Tay's Musings about Librarianship*. https://musingsaboutlibrarianship.blogspot.com/2024/05/retrieval-augmented-generation-and.html

Turner, B. (2024, May 24). Google's AI tells users to add glue to their pizza, eat rocks and make chlorine gas. *Live Science*.
https://www.livescience.com/technology/artificial-intelligence/googles-ai-tells-users-to-add-glue-to-their-pizza-eat-rocks-and-make-chlorine-gas