

## Perils of Searching Unfamiliar Topics

Marydee Ojala\*

Corresponding author: marydee@xmission.com

### Abstract

If you've ever been employed as a generalist, I salute you. Working at the reference desk of a public or academic library is tough. You just never know what topic will arise. And somehow, you're supposed to know everything about everything, ranging from anthropology to zoology, with current news events thrown into the mix. We're all familiar with the librarian motto that we don't actually need to know everything, we just need to know where to find it. That is becoming more difficult in an age of AI-generated information.

### Keywords

Internet search, Digital library, Online retrieval

Information professionals working in higher education as liaisons to specific departments and those in specialized libraries have a somewhat easier time of it. They develop a good understanding of the sources in their area, whether it's business, chemistry, or history. When I worked for a financial institution, I became intimately familiar with banking terminology and finance statistics. If you work for a pharmaceutical company, a government agency, or an association, you immerse yourself in relevant and reliable resources.

Problems arise when information that sounds perfectly plausible, even to those in the know, isn't actually correct. I am not simply referring to hallucinations. By now, info pros are well aware of the dangers posed by generative AI (gen AI) creating fake citations[1,2]. Thanks to RAG (retrieval-augmented generation), hallucinations have decreased. In fact, when I tested several reports generated by deep research on Perplexity and Gemini, which contained dozens of citations in their references, all of them were correct. Yes, it was laborious to check each one of them, but worth the effort. And yes, I was surprised.

Errors can crop up that are created outside the

realm of AI but amplified by the large language models (LLMs) used by gen AI. A recent case in point is "vegetative electron microscopy." Sounds like a real thing, right? It isn't. Not, frankly, that I would know if I found that phrase in search results, since it's far from my area of expertise.

As Retraction Watch reported ([retractionwatch.com/2025/02/10/vegetative-electron-microscopy-fingerprint-paper-mill](https://retractionwatch.com/2025/02/10/vegetative-electron-microscopy-fingerprint-paper-mill/)), the phrase was likely introduced by a scanning error in an article published in 1959[3]. The word "vegetative" was in one column and immediately adjacent to it in the second column was "electron microscopy"—the scan jumped the columns to combine the words into one phrase[4]. Particularly in the early days of OCR scanning, such a mistake was not uncommon.

What's new is the ingestion of older scholarly articles into the training sets for LLMs. Suddenly, search chatbots can spit out "vegetative electron microscopy" as valid terminology. This amplification of fake phrases is sometimes referred to as "digital fossils" or "tortured phrases." It reminds me of what was known in journalism circles as the "Nexis effect" decades ago. A reporter might get a number

wrong in an article that was added, full text, to Nexis. Other reporters searched Nexis, found the article, and repeated the incorrect statistic. Suddenly, it became accepted truth[5].

This particular digital fossil has not permeated the scientific literature. A Dialog search found only three articles, one of which was about the Retraction Watch article, one a retraction, and the third about a disgraced researcher. Google Scholar found 12. One was the original 1959 article, one a retraction, and two were the same article, as originally published and as corrected. The remainder used the phrase as if it were a real thing.

This begs the question of what to do if the non-specialist searcher finds a digital fossil but doesn't know that's what it is. Without specialized knowledge (and lacking the time to check Retraction Watch constantly), it's hard to know what is a true technical term and what is a scanning error or a randomly introduced nonsense phrase.

You can use common sense. If it's a Very Important Technology, why are there only a handful of mentions in Google Scholar and next to none in your library's databases? In the case of "vegetative electron microscopy," why were there no articles published between 1959 and 2019? You can also ask an expert. Your requestor may be as early to the field as you are, so do a reality check with a department head, association analyst, or established researcher.

On a lighter note, WIRED reported on Google AI Overviews nonsense ([wired.com/story/google-ai-overviews-meaning/](https://www.wired.com/story/google-ai-overviews-meaning/)). Make up an idiom—WIRED used "You can't lick a badger twice"—enter it into Google, and get delightfully zany results from AI Overviews, including its hallucinations about

what your gibberish means and its origin story. Just don't use it in scholarly articles.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to show sincere thanks to those technicians who have contributed to this research.

## References

- [1] Bourner, T., Greener, S. (2016). Searching in unfamiliar fields. *Research Methods for Postgraduates*, 113-120.
- [2] Debowski, S. (2001). Wrong way: Go back! An exploration of novice search behaviours while conducting an information search. *The Electronic Library*, 19(6), 371-382.
- [3] O'Mahony, S., Cohen, S. (2022). Navigating the promises and perils of researching emerging phenomena in strategy and organizations. *Strategic Organization*, 20(4), 872-885.
- [4] Salmerón, L., Kammerer, Y., García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior*, 29(6), 2161-2171.
- [5] Salmerón, L., Kammerer, Y., García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior*, 29(6), 2161-2171.

*Marydee Ojala* is a career information professional and editor of *Online Searcher*.