



AI versus web search

Description

Discussions of AI in web search are often overtaken by concerns over the risks associated with automated user profiling (see post [Surveillance capitalism and its discontents](#)). Setting aside that raft of concerns, it's worth reviewing claims about how AI facilitates efficient and helpful web search.

AI enabled search

Gemini (<https://gemini.google.com/>) is the name of Google's new prototype AI-enabled multi-modal search engine. Who better to converse with about the role of AI in search engines than the Gemini conversational platform (chat bot). What follows is what we discussed (in my own words).

Indexing is key to a search engine's operations. Technically, at its most basic it uses linked tree data structures and hash tables for the indexes stored and distributed on extensive server networks. Many search engines now employ neural networks as part of that process. The accepted wisdom identifies three roles for AI in web search.

Alternative interpretations: Soon after a web page goes online and independently of any search event, the search engine indexes the words on the page. AI-enabled search engines can now produce page summaries, sometimes using synonyms. It can include those extra terms in the index. That's similar to processes by which large language models (ChatGPT, Claude, Gemini) produce text summaries.

Note that these summaries are not designed necessarily for human readability, but enable the search engine to deliver web pages that address key words and concepts other than those expressed explicitly on the web pages. AI-enabled search engines will process user **queries** in a similar way. If I type in a search query "underground domestic heating carbon footprint" then the search engine might run that through an LLM (large language model) prediction engine (like Gemini or chatGPT) and through its inference mechanism send terms such as "geothermal" as part of the query.

Entity recognition: Neural networks as used in LLMs are good at recognizing and classifying entities within webpages and search queries. So the query "underground domestic heating carbon

footprint?• likely falls within classes of queries about *heating, sustainability* and *buildings* rather than *underground railways, beachcombers* or *chemistry*. This LLM enabled classification procedure allows search engines to process the connections between these entities and finesse the search.

Ranking: Search engine list their page results in some order. Straightforward indexed search can rank page results on the basis of the number of times a keyword appears on a page, the date the page was created or modified, the number of times the site was visited by other searchers, and bias factors based on URL domain, credibility of the source, quality of the content, advertising, etc. Each search event generates data about search behaviour. Aided by the search engine's ability to deal with synonyms, concepts, and entities, AI neural networks can detect and exploit patterns in that data to rank the output page lists.

If I search "modern office building stability" an AI-powered search engine might identify that the user wants to learn about structural in office building design. It identifies "office building" as the main entity and considers related terms like "commercial architecture", "workplace design", and "earthquake".

Gemini summarises the AI-enabled search process: "This approach provides a more relevant and informative search experience. You'll likely see results that delve deeper into the topic and address the current trends in modern office building design."

The Gemini conversational platform suggested I mention how an AI-enabled search engine "can go beyond words to grasp the searcher's true information need." (I'm not sure if I ever have "true information needs.") It also told me that AI can help personalise users' preferences via analysis of their search histories. It mentioned AI generated "featured snippets and knowledge panels" that can accompany search results. It also said "Consider mentioning that AI-powered search is still under development, and its effectiveness depends on the quality of training data and algorithms."

Reference

- Google_Team. "Gemini: A Family of Highly Capable Multimodal Models." *DeepMind*, 19 December 2023, 2023. Accessed 3 April 2024. <https://arxiv.org/pdf/2312.11805.pdf>

Note

- Featured image is from ChatGPT: "The revised banner image for your blog, incorporating optical elements with lenses, has been created to further emphasize the themes of search and discovery in the context of AI-enabled web search within a steampunk aesthetic."

Category

1. Artificial Intelligence

Date Created

April 3, 2024

Author

rcoyne99

default watermark