Is ChatGPT-like technology going to replace commercial search engines?

Artur Strzelecki

[1] Department of Informatics, University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland

Version of Record of this manuscript is published here:

Strzelecki, A. (2024), "Is ChatGPT-like technology going to replace commercial search engines?", Library Hi Tech News, Vol. ahead-of-print No. ahead-of-print.

https://doi.org/10.1108/LHTN-02-2024-0026

Is ChatGPT-like technology going to replace commercial search engines?

Abstract

Purpose

The article gives an overview of the history and evolution of commercial search engines. It traces the development of search engines from their early days to their current form as complex technology-powered systems that offer a wide range of features and services.

Design/methodology/approach

In recent years, advancements in AI technology have led to the development of AI-powered chat services. This study explores official announcements and releases of three major search engines, Google, Bing and Baidu, of AI-powered chat services.

Findings

Three major players in the search engine market, Google, Microsoft, and Baidu started to integrate AI chat into their search results. Google has released Bard, later upgraded to Gemini, a LaMDA-powered conversational AI service. Microsoft has launched Bing Chat, renamed later to Copilot, a GPT-powered by OpenAI search engine. The largest search engine in China, Baidu, released a similar service called Ernie. There are also new AI-based search engines, which are briefly described.

Originality

This paper discusses the strengths and weaknesses of the traditional - algorithmic powered search engines and modern search with generative AI support, and the possibilities of merging them into one service. This study stres the types of inquires provided to search engines, users' habit of using search engines and technological advantage of search engines infrastructure.

Keywords: ChatGPT, Microsoft Bing, Google Gemini, Baidu, Ernie AI, search engine, language model,

Is ChatGPT-like technology going to replace commercial search engines?

The recent excitement about ChatGPT and its unprecedented growth in users (100 million in two months) has stirred the information technology industry. OpenAI, the company behind ChatGPT, made it publicly available for everyone to use for free in late 2022. Its positive results have generated discussion in leading media outlets around the world. At the same time that ChatGPT was gaining popularity, commercial web search engine companies began the process of incorporating AI chat technology into their search engines.

Commercial search engines

Commercial search engines have a long history, that dates back to the 1990s when the first web catalogs and search engines were introduced to internet users In the beginning, their function was very simple. With web catalogs, users could search for resources using queries provided by internal search engines, and the results were limited to those previously inserted and moderated by the web catalog owner. For search engines, the results presented to internet users came from the part of the internet that had already been crawled by the search engine. The search engine would send a program, called a crawler, to crawl and download all the content on the internet. The results presented by web catalogs and search engines were similar, consisting of the website title, a brief description, and the URL. This format for presenting search engine results has not changed since the invention of web catalogs and search engines, but the first one have become extinct. Today, people no longer use web catalogs to find information as search engines have evolved and offer more features than the initial list of text results from a search engine index.

Contemporary search engines offer a variety of options and functions for searching, including text, images, videos, and books. The results are typically presented and ordered according to an internal algorithm that aims to provide the best results for a given query. The algorithm takes into account hundreds of criteria to determine the ranking of results, including factors related to the content, user behavior history, and other signals. A leak of the Yandex source code in 2023 revealed that the Yandex search engine has implemented over 1,922 different ranking factors in its search engine, demonstrating the complexity and evolution of search engine algorithms beyond PageRank and HITS. In addition to the content that is crawled and stored in the search engine index (text, images, videos, books), search engines offer a range of supplementary search services, such as flight searches, insurance searches, news searches, shopping searches, map searches, sports scores, stock prices, weather, and more.

Search engines are also large commercial services that display ads for sponsored content. Sponsored search results provide the revenue that sustains the search engine and enables it to support billions of users. Sponsored results allow companies to display their offerings to users as they search for information. Ads are typically aligned with the user's intent and displayed alongside the results generated by the search engine algorithm. Search engines manage various types of ads, including text, image, and video ads. The presence of ads in search engine results has increased, evolving from being a minor component or appearing only on the right side of the results, to being prominently displayed as major results that are nearly indistinguishable from those generated by the algorithm, with only a small, often overlooked mark indicating that it is an ad Schultheiß and Lewandowski, 2021).

For years, people have come to trust search engine results and have increasingly stopped browsing beyond the first page. Studies have shown that interest in results beyond the first page is practically non-existent (Strzelecki & Miklosik, 2024). People typically only view the first page, and if they cannot find what they are looking for, they tend to refine their search query. To address this, search engines have offered various search operators to help users narrow down their results. However, users today prefer a more user-friendly experience, and with the growth of the mobile market, search engines have introduced the ability to perform voice searches and receive voice responses from a virtual assistant. To facilitate this, search engines have created a direct answer, also known as an "answer box" or "featured snippet" (Strzelecki & Rutecka, 2020), which is presented in a paragraph, list, or table format, making it ideal for voice searches.

For nearly thirty years, commercial search engines have been utilized billions of times a day by millions of users, providing answers, results, and other services. The search market is now dominated by Google, which is the leading search service used worldwide. According to StatCounter, in January 2024, Google held a market share of approximately 91.5%, followed by Bing with a share of 3.5% and Yahoo! with a share of 1.1% [i]. Other commercial search engines, such as Yandex, Baidu, DuckDuckGo, Naver, Ecosia, Sogou, or Seznam, are not widely used, according to StatCounter. However, Yandex and Baidu are the main search engines used in Russian and Chinese, respectively, and not widely used in the Western world. Despite this dominance, Google's position as the leading search engine is facing challenges from recent advancements in AI chat technology, rather than penalties for monopolistic practices imposed by national or international authorities (Norocel & Lewandowski, 2023).

AI chats

AI chats are based on language model technology. The Transformer, a neural network architecture developed by Google Research and introduced in 2017, serves as the foundation for many current language models, including BERT and GPT-4. The majority of generative AI applications currently in use are built on Google's Transformer research project and their groundbreaking 2017 paper. This architecture creates a model that can be trained to read a substantial number of words (such as a sentence or paragraph), observe the relationship between the words, and then anticipate the words that it believes will appear next. In 2022, Google introduced LaMDA, which stands for “Language Model for Dialogue Applications” and builds on prior Google research (Adiwardana et al., 2020) that showed that Transformer-based language models trained on dialogue could encompass a wide range of topics. LaMDA is trained using dialogue, in contrast to the majority of language models, and its responses can be fine-tuned to significantly increase their specificity and rationality.

Despite years of work and expertise gained by Google in AI, OpenAI changed the game with the release of ChatGPT in November 2022 (OpenAI, 2023). ChatGPT is a language model optimized for dialogue, capable of conversing in a natural manner. Its ability to reply to follow-up questions, admit mistakes, disprove false assumptions, and decline inappropriate requests, makes it stand out among other language models. ChatGPT was improved from a model in the GPT-3.5 series and quickly gained widespread popularity, with over 100 million users within the first two months of its release. This sparked a widespread discussion about the use of AI in various settings, such as homes, schools, and workplaces. Despite other language models like Google’s BERT (Bidirectional Encoder Representations from Transformers), Facebook’s LLaMA (Large Language Model Meta AI) or Mixtral by Mistra AI being available, they did not receive the same level of satisfaction as the GPT series and ChatGPT. As a result, major commercial search engines like Google, Bing, and Baidu quickly announced plans to integrate AI chat into their search results and started to introduce it.

Integration of Chat AI into search engine

In February 2023, information was released by Baidu, Google, and Microsoft regarding their plans to integrate AI chat into their search results. Google firstly developed Bard, an experimental conversational AI service powered by LaMDA. Initially, it was made available to trusted testers before being made more widely accessible to the public in the future. The CEO of Google, Sundar Pichai, stated that Bard aims to “combine the breadth of the world's knowledge with the power, intelligence, and creativity of their large language models” and that it will use information from the web to provide fresh, high-quality responses

(Pichai, 2023). To ensure that Bard's responses are of high quality, safe, and grounded in real-world data, Google combined external feedback with its own internal testing. The initial release of Bard used a significantly scaled-down model of LaMDA that uses less computing power, making it more accessible to a wider range of users (Pichai, 2023). In February 2024 Bard was replaced by Google Gemini, a multimodal language model which means it can process and integrate different types of sources.

Microsoft has also released its AI-powered Bing search engine, now called a Copilot (Mehdi, 2023). The Microsoft Copilot is powered by the next-generation OpenAI large language model (LLM), which is more powerful than ChatGPT and specifically tailored for search. This model incorporates significant advancements from ChatGPT and GPT-4, making it faster and more accurate. Microsoft has a close relationship with OpenAI, as ChatGPT and GPT-3.5 were trained on Microsoft's Azure AI supercomputing infrastructure. Microsoft claims that it offers improved search results, more thorough responses, a new chat experience, and the content creation gapbilities (Mehdi, 2023).

Finally, China's largest search engine company, Baidu, released a ChatGPT-like application in October 2023 (Mo & Baptista, 2023). It is incorporated into its primary search services. Users of the Ernie tool are able to receive conversational search results, similar to the popular OpenAI platform. The foundation of this ChatGPT-like tool is Baidu's Ernie system, a LLM that has been trained on data for several years (Huang, 2023).

A journalist from CNBC conducted a test in which they asked ChatGPT to write an article on whether or not ChatGPT is a viable alternative to Google as a search engine (Pitt, 2023). According to the report, ChatGPT's response was subjective and highlighted its limitations. It emphasized the main differences between AI chat and search engines in terms of data availability, personalization, and conversational interaction. Additionally, there are concerns about ChatGPT being in its beta form and occasional downtime. OpenAI CEO Sam Altman has stated that ChatGPT is “incredibly limited” and “it's a mistake to rely on it for anything important at this time”.

AI chat can be a great complement to traditional search engine services, but it cannot be relied upon as a sole source. An example is the release of Google Bard. Google provided an example of Bard answering the question, “What new discoveries from the James Webb Space Telescope can I tell my 9-year-old about?” Bard responded with three bullet points, but the third one was incorrect. Bard claimed that the James Webb Space Telescope took the “very first picture of a planet outside our solar system,” but this statement was not true. The first image was actually taken by Chauvin et al. in 2004 using adaptive optics with the

VLT/NACO. This effectively highlights the main flaw of statistical methods, as these systems are not meant to provide accurate answers, but rather answers that are based on statistical analysis and appear to be plausible.

Apart from Microsoft Copilot, Google Gemini and Baidu Ernie there are other AI-based search engines. Some of them are Andi, Exa, Brave, You, Phind and Perplexity. *Andi* is a conversational search engine that uses generative AI and LLMs to find answers and information on the web. It combines natural language understanding, semantic search, and real-time data querying to provide accurate and trustworthy responses. Andi is also a friendly and visual chatbot that summarizes and explains key information from the best sources, and shows search results in a clear and ad-free format. *Exa* is a knowledge API for LLMs. Exa allows users to query the web using natural language and obtain a list of relevant webpages from a neural database. Exa also supports traditional keyword search and content retrieval, provides highlights from the webpages, which are intelligent extracts calculated using retrieval-augmented generation models and is a tool for developers who want to build applications that require natural language understanding and web search capabilities.

Brave Search is a search engine that aims to provide privacy, independence, and innovation in web search. Unlike most other search engines, Brave Search does not collect or store personal information about its users, nor does it rely on big-tech companies to power its results. Instead, Brave Search uses its own independent index, community feedback, and alternative ranking models to deliver relevant and diverse results. Two language models that can be downloaded from Hugging Face are used by Brave Search. They are BART and DeBERTa. *You* is an innovative AI assistant that provides detailed and personalized answers to any questions, using the latest natural language processing and deep learning technologies. Interaction with You is in a conversational way, and one get reliable, actionable, and comprehensive responses that include credible citations, relevant web findings, and dynamic rich media1. You also offers AI modes that allow to create vivid imagery, generate complex computations, and perform deep dive research on any topic.

Phind, an AI-powered search engine and chatbot, revolutionizes information retrieval for developers and technical enthusiasts. Leveraging machine learning, deep learning, and natural language processing techniques, Phind offers context-aware and tailored search results. Unlike traditional search engines that inundate users with extensive lists of search results, Phind provides instant answers to technical queries, eliminating the need for exhaustive manual search efforts. Its continuously learning model improves with each interaction, ensuring up-to-date and accurate responses. As a personal tech assistant, Phind

possesses an extensive knowledge base, making it proficient in addressing a wide range of technical questions and providing relevant code snippets for enhanced understanding and implementation. *Perplexity* is a conversational search companion that leverages advanced AI models to provide in-depth answers to users' queries. Unlike traditional search engines that return a list of links, Perplexity engages users in a natural language dialogue, asking clarifying questions and summarizing the best findings from a variety of sources. Perplexity aims to elevate users' quest for knowledge by simulating human intelligence and reasoning with the help of technologies such as natural language processing, machine learning, and deep learning.

Answering the question "*Is ChatGPT-like technology going to replace commercial search engines?*", for now, AI chats such as ChatGPT, Gemini, Ernie and other unnamed models will not replace commercial search engines. The first reason is that they are likely to be a component of search engines, serving as an additional option for answering queries. AI chat is most useful for informational queries, which aim to locate content regarding a specific topic to address the searcher's information needs, while navigational queries aim to locate a specific website and transactional queries aim to locate a specific product or service or compare prices. AI chats are models trained on millions of parameters and have knowledge of millions of facts, but they cannot provide up-to-date information, trends, or directions. For these purposes, commercial search engines will still be useful for navigational and transactional queries. Despite this limitation for navigational and transactional queries, informational queries are the majority of those input into search engines.

The second reason is that users are now accustomed to using search engines as a natural habit, quickly inputting queries through keypads or voice services. They are used to reviewing the list of results presented in a visually appealing way that invites them to spend more time on the search engine results page. In recent years, the interface of modern search engines has evolved from a simple query field and a "search" button to multiple, specialized search engines that work together to provide text, graphics, and video results all in one place, which AI chat is not yet capable of and may not ever be able to achieve. AI chat presents data and facts that the language model was trained on and the crawling function is usually limited to the results from the search engine.

The final reason is technological advantage. Search engines have developed a global infrastructure of thousands of servers in numerous data centers that work on crawling, indexing, and presenting results to users. They are one of the most stable and reliable services used by billions of people with the intention to operate soon only on carbon-free energy. On

the other hand, AI chats consume a significant amount of computing power to produce answers, compared to processing a single query by a search engine. The infrastructure behind AI-chats is facing exceptional demand, and despite efforts to scale their systems, access to these services is limited by user capacity.

Search engines, as we know them, will remain but will be incorporating AI-powered services to enhance the search experience and provide even better results than what can currently be obtained from a search engine. AI chat is best suited as a companion to traditional search engines rather than a sole source of information.

Disclaimer

"No portion of this paper was created using artificial intelligence."

Declaration of conflicting interests

"The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article."

References:

Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., & Le, Q. V. (2020). *Towards a human-like open-domain chatbot*. http://arxiv.org/abs/2001.09977

Huang, Z. (2023). *Chinese Search Giant Baidu to Launch ChatGPT-Style Bot*. Bloomberg. https://www.bloomberg.com/news/articles/2023-01-30/chinese-search-giant-baidu-to-launch-chatgpt-style-bot-in-march

Mehdi, Y. (2023). *Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web*. Microsoft. https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/

Mo, Y., & Baptista, E. (2023). *China's Baidu unveils new Ernie AI version to rival GPT-4*. Reuters. https://www.reuters.com/technology/chinas-baidu-unveils-latest-version-its-ernie-ai-model-2023-10-17/

Norocel, O. C., & Lewandowski, D. (2023). Google, data voids, and the dynamics of the politics of exclusion. *Big Data & Society*, *10*(1), 205395172211490. https://doi.org/10.1177/20539517221149099

OpenAI. (2023). *ChatGPT: Optimizing language models for dialogue*. https://openai.com/blog/chatgpt/

Pichai, S. (2023). *An important next step on our AI journey*. Google. https://blog.google/technology/ai/bard-google-ai-search-updates/

Pitt, S. (2023). *Google vs. ChatGPT: Here's what happened when I swapped services for a day*. CNBC. https://www.cnbc.com/2022/12/15/google-vs-chatgpt-what-happened-when-i-swapped-services-for-a-day.html

Strzelecki, A., & Miklosik, A. (2024). Device-dependent click-through rate estimation in Google organic search results based on clicks and impressions data. *Aslib Journal of Information Management*. https://doi.org/10.1108/AJIM-04-2023-0107

Strzelecki, A., & Rutecka, P. (2020). Direct Answers in Google Search Results. *IEEE Access*, *8*, 103642–103654. https://doi.org/10.1109/ACCESS.2020.2999160

[i] https://gs.statcounter.com/search-engine-market-share