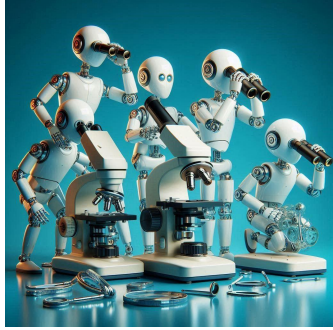


Search Engines and AI Perplexity, “Google Zero”, and AI Overviews	
---	---

**An AI Generated Overview of Search Engines and AI - - -
(WARNING - Hallucinations are possible)**

ChatGPT 4o - 11/18/2024 ... Search Engines and AI

(Prompt - Please provide an overview of possible issues and benefits with regard to AI and internet search engines such as Google, Bing, and Perplexity. Please make your response roughly 800-1000 words in length and appropriate for a college audience.)

AI and Internet Search Engines: Issues and Benefits

Artificial Intelligence (AI) has significantly impacted the way internet search engines function, transforming traditional search processes into dynamic and interactive experiences. Search engines like Google, Bing, and emerging AI-powered platforms like Perplexity have integrated AI models to provide faster, more personalized, and context-aware results. However, these advancements bring both opportunities and challenges. Below, we examine the key benefits and potential issues related to the integration of AI in search engines.

Benefits

1. **Enhanced Relevance and Personalization** AI-powered search engines leverage machine learning and natural language processing (NLP) to understand queries more effectively. Unlike keyword-based searches, modern AI models can interpret the intent behind queries, enabling them to provide results tailored to the user's context. For instance:
 - **Google's Search Algorithm:** Uses AI models like BERT (Bidirectional Encoder Representations from Transformers) to better understand nuanced queries.

- **Personalized Results:** AI learns user preferences over time, offering suggestions or tailored content based on browsing history, location, and previous searches.
 - 2. **Efficiency and Speed** AI enables search engines to process vast amounts of data rapidly, returning results almost instantaneously. For example:
 - **Voice Search and Virtual Assistants:** Platforms like Bing integrate with AI-powered assistants such as Microsoft's Cortana, allowing users to perform searches using natural spoken language.
 - **Auto-Complete Features:** Predictive text powered by AI accelerates query formulation, making the search process smoother.
 - 3. **Innovations in Answer Generation** AI models, especially large language models (LLMs) like GPT or ChatGPT, generate comprehensive answers rather than a list of links. Tools like Perplexity integrate conversational AI to provide concise, sourced summaries.
 - Users receive direct answers to complex questions without sifting through multiple webpages.
 - Advanced platforms support follow-up questions, creating a more interactive search experience.
 - 4. **Democratization of Knowledge** AI democratizes access to information by breaking language barriers and enabling searches in multiple languages. Features like Google's Multisearch allow users to combine text and images for more complex queries, benefiting diverse audiences, including those with limited technical skills.
 - 5. **Improved Accessibility** Search engines equipped with AI can cater to users with disabilities. Voice-based searches, screen readers, and visual recognition capabilities help individuals with visual or mobility impairments navigate the internet more effectively.
-

Issues

1. **Bias in Search Results** AI algorithms are only as unbiased as the data they are trained on. If training data reflects societal prejudices or skews towards certain perspectives, search engines risk amplifying these biases.
 - **Filter Bubbles:** Personalization can lead to echo chambers, where users are consistently exposed to information aligning with their views, limiting diverse perspectives.
 - **Misinformation:** AI may prioritize sensationalized or inaccurate content if it aligns with user engagement metrics.
2. **Transparency and Accountability** AI-driven search engines often lack transparency in how they rank and filter information. Users may find it difficult to discern:
 - Why certain results are prioritized.
 - Whether results are organically generated or influenced by advertising or partnerships.
3. **Over-Reliance on AI Summaries** While tools like Perplexity offer direct answers, these responses might oversimplify complex topics, omit critical nuances, or misrepresent information. Users

could become overly reliant on AI-generated answers and neglect thorough, independent research.

4. **Privacy Concerns** AI personalization relies heavily on data collection, including browsing habits, location, and preferences. Such practices raise questions about user privacy:
 - **Data Usage:** The extent to which search engines collect, store, and share personal information remains a concern.
 - **Potential for Misuse:** Data breaches or misuse of personal data could have far-reaching consequences.
 5. **Economic and Social Disruptions** The dominance of AI-driven search platforms may undermine smaller competitors or niche content creators. For example:
 - **Monetization Challenges:** Publishers relying on search engine traffic may struggle as users access AI summaries without visiting source websites.
 - **Job Displacement:** Automation of tasks, including research and information curation, may reduce employment opportunities in related sectors.
 6. **Information Overload** While AI can refine searches, the internet's sheer volume of information remains overwhelming. Users must critically evaluate sources and sift through irrelevant or redundant content despite AI's assistance.
 7. **Potential for Dependence** Over-reliance on AI-powered search engines may diminish users' critical thinking and research skills. Instant answers reduce the incentive to explore multiple viewpoints or dig deeper into complex subjects.
-

Case Studies and Real-World Applications

1. **Google's AI Innovations** Google's integration of AI models like Bard and BERT has revolutionized search, allowing for nuanced query interpretation. However, critics argue that Google's prioritization of ads and partnerships sometimes undermines organic search quality.
 2. **Microsoft Bing and ChatGPT Integration** Bing's collaboration with OpenAI to include ChatGPT-style responses illustrates the potential for interactive search. While the tool provides conversational responses, its accuracy depends on the quality of the training data and may occasionally generate "hallucinated" (false) information.
 3. **Perplexity.ai** As a rising AI-powered search tool, Perplexity combines conversational AI with citations to deliver sourced answers. While innovative, its reliance on AI summaries introduces risks of oversimplification and potential inaccuracies.
-

Balancing the Pros and Cons

To fully leverage the benefits of AI-powered search engines while mitigating potential issues, several strategies can be adopted:

- **Transparency and Explainability:** Search engines should disclose how algorithms rank and filter information, empowering users to make informed decisions.
 - **Bias Mitigation:** Regular audits and diverse training datasets can help reduce algorithmic bias.
 - **Critical Digital Literacy:** Users must be educated on evaluating AI-generated responses critically, recognizing limitations, and validating information from multiple sources.
 - **Privacy Safeguards:** Implementing robust data protection measures, including transparent policies and user controls, is essential.
-

Conclusion

AI-powered search engines have ushered in a new era of internet exploration, characterized by efficiency, accessibility, and innovation. Platforms like Google, Bing, and Perplexity showcase how AI enhances the relevance of search results and simplifies information retrieval. However, these advancements come with challenges, including biases, privacy concerns, and the potential erosion of critical thinking skills.

As AI technology continues to evolve, balancing its benefits with ethical considerations and transparency will be key to ensuring that search engines remain trustworthy, inclusive, and empowering tools for users worldwide. The onus lies on developers, policymakers, and users to collaboratively shape a future where AI-driven searches enrich rather than compromise the pursuit of knowledge.

Human Views on AI and Search Engines - - - (WARNING - Humans are fallible)

Shorter Podcasts and Videos:

AI's Role in Your Next Online Search: The Evolution of Search Engines

A Beginner's Guide to AI (13:04)

(March 30, 2024) [Podcast]

<https://podcasts.apple.com/us/podcast/ais-role-in-your-next-online-search-the-evolution/id1701165010?i=1000650944016>

Google Zero is here — now what?

Decoder - Nilay Patel (27:53)

(May 30, 2024) [Podcast]

<https://www.theverge.com/24167865/google-zero-search-crash-housefresh-ai-overviews-traffic-data-audience>

How AI Is Fundamentally Reshaping the Web

The AI Daily Brief - Nathaniel Whittemore (6:11)

(October 16, 2024) [Video]

<https://www.youtube.com/watch?v=VYPTsbyrulc>

Why Google Search feels like it's gotten worse.

Decoder - Nilay Patel and Mia Sato (38:25 ... start at 1:05)

(March 2024) [Podcast]

<https://podcasts.apple.com/us/podcast/why-google-search-feels-like-its-gotten-worse/id1011668648?i=1000649171711>

Longer Podcasts and Videos:

AI Is Breaking Google

Better Offline - Ed Zitron (48:42)

(May 28, 2024) [Podcast]

<https://podcasts.apple.com/us/podcast/ai-is-breaking-google/id1730587238?i=1000657117343>

Google CEO Sundar Pichai says AI search will actually help the web

The Verge - Nilay Patel, & Sundar Pichai (39:15)

(May 20, 2024) [Video]

<https://www.youtube.com/watch?v=lqikP9X9-ws&t=451s>

Gunning for Google with Perplexity CEO Aravind Srinivas

The Cognitive Revolution - Nathan Labenz and Aravind Srinivas (58:55)

(October 3, 2023) [Video]

<https://www.cognitiverevolution.ai/gunning-for-google-with-perplexity-ceo-aravind-srinivas/>

The Man That Destroyed Google Search

Better Offline - Ed Zitron (30:54)

(May 28, 2024) [Podcast]

<https://podcasts.apple.com/us/podcast/the-man-that-destroyed-google-search/id1730587238?i=1000653621646>

Revolutionizing Search with Perplexity AI with Aravind Srinivas

Eye on AI - Craig Smith and Aravind Srinivas (53:01)

(March 14, 2024) [Video]

<https://www.youtube.com/watch?v=ADyaGQza8pI>

Magazine and Newspaper Articles:

Broderick, Ryan. (2023, August 28). **The end of the Googleverse. For two decades, Google Search was the invisible force that determined the ebb and flow of online content. Now, for the first time, its cultural relevance is in question.** *The Verge*.

<https://www.theverge.com/23846048/google-search-memes-images-pagerank-altavista-seo-keywords>

Lopatto, Elizabeth. (2024, June 27). **Perplexity's grand theft AI: More like perfidy?** *The Verge*.

<https://www.theverge.com/2024/6/27/24187405/perplexity-ai-twitter-lie-plagiarism>

Lewis, Amanada Chicago. (2023, November 1). **The people who ruined the internet. As the public begins to believe Google isn't as useful anymore, what happens to the cottage industry of search engine optimization experts who struck content oil and smeared it all over the web? Well, they find a new way to get rich and keep the party going.** *The Verge*.

<https://www.theverge.com/features/23931789/seo-search-engine-optimization-experts-google-results>

[UNI only] McCollum, Marc. (2024, September 18). **A pivotal moment for the internet and content creators.** *Newsweek*.

<https://advance-lexis-com.proxy.lib.uni.edu/api/document?collection=news&id=urn%3acontentItem%3a6D0H-6RR1-DY68-10WD-00000-00&context=1519360&identityprofileid=7FZBZ451562>

[UNI only] Needleman, Sarah. (2023, October 18). **AI looms as search power: Generative tech will soon be challenging traditional engines, says industry expert.** *Wall Street Journal*.

<https://login.proxy.lib.uni.edu/login?url=https://www.proquest.com/newspapers/ai-looms-as-search-power-generative-tech-will/docview/2878207659/se-2?accountid=14691>

Patel, Nilay. (2023, May 8). **What happens when Google Search doesn't have the answers? After controlling how information has been distributed for the past 25 years, Google Search faces a set of challenges that will change the company — and the internet — forever.** *The Verge*.

<https://www.theverge.com/23712602/google-search-25-years-anniversary-ai-artificial-intelligence>

Robinson, Kylie. (2024, May 30). **Google defends AI search results after they told us to put glue on pizza: Google blames 'data voids' and edge cases for its bad AI-generated search results.** *The Verge*.

<https://www.theverge.com/2024/5/30/24168344/google-defends-ai-overviews-search-results>

Robinson, Kylie. (2024, July 25). **OpenAI announces SearchGPT, its AI-powered search engine: The Google and Perplexity rival will be available as a prototype in limited release, with plans to eventually build it into ChatGPT.** *The Verge*.

<https://www.theverge.com/2024/7/25/24205701/openai-searchgpt-ai-search-engine-google-perplexity-rival>

Robinson, Kylie. (2024, July 30). **Perplexity is cutting checks to publishers following plagiarism accusations: Time, Der Spiegel, and Fortune are among the first batch of publications in Perplexity's new publisher's program.** *The Verge*.

<https://www.theverge.com/2024/7/30/24208979/perplexity-publishers-program-ad-revenue-sharing-ai-time-fortune-der-spiegel>

[UNI only] Roose, Kevin. (2024, February 1). **Can this A.I.-powered search engine replace Google? It has for me.** *New York Times*. <https://www.nytimes.com/2024/05/14/technology/google-ai-answers.html>

[UNI only] Roose, Kevin. (2024, May 14). **Can Google give A.I. answers without breaking the web?** *New York Times*. <https://www.nytimes.com/2024/05/14/technology/google-ai-answers.html>

Scholarly Journal and Preprint Articles

Andrikyan, W., Sametinger, S. M., Kosfeld, F., Jung-Poppe, L., Fromm, M. F., Maas, R., & Nicolaus, H. F. (2024). **Artificial intelligence-powered chatbots in search engines: a cross-sectional study on the quality and risks of drug information for patients.** *BMJ Quality & Safety*.

Abstract: Search engines often serve as a primary resource for patients to obtain drug information. However, the search engine market is rapidly changing due to the introduction of artificial intelligence (AI)-powered chatbots. The consequences for medication safety when patients interact with chatbots remain largely unexplored. AI-powered chatbots are capable of providing overall complete and accurate patient drug information. Yet, experts deemed a considerable number of answers incorrect or potentially harmful. Furthermore, complexity of chatbot answers may limit patient understanding. Hence, healthcare professionals should be cautious in recommending AI-powered search engines until more precise and reliable alternatives are available.

Ashraf, A. R., Mackey, T. K., & Fittler, A. (2024). **Search engines and generative artificial intelligence integration: Public health risks and recommendations to safeguard consumers online.** *JMIR Public Health and Surveillance*, 10(1), e53086. <https://publichealth.jmir.org/2024/1/e53086/>

Abstract: The online pharmacy market is growing, with legitimate online pharmacies offering advantages such as convenience and accessibility. However, this increased demand has attracted malicious actors into this space, leading to the proliferation of illegal vendors that use deceptive techniques to rank higher in search results and pose serious public health risks by dispensing substandard or falsified medicines. Search engine providers have started integrating generative artificial intelligence (AI) into search engine interfaces, which could revolutionize search by delivering more personalized results through a user-friendly experience. However, improper integration of these new technologies carries potential risks and could further exacerbate the risks posed by illicit online pharmacies by inadvertently directing users to illegal vendors.

Hersh, W. (2024). **Search still matters: information retrieval in the era of generative AI.** *Journal of the American Medical Informatics Association*, 31(9). <https://arxiv.org/pdf/2311.18550>

Abstract: Information retrieval (IR, also known as search) systems are ubiquitous in modern times. How does the emergence of generative artificial intelligence (AI), based on large language models (LLMs), fit into the IR process? This perspective explores the use of generative AI in the context of the motivations, considerations, and outcomes of the IR process with a focus on the

academic use of such systems. There are many information needs, from simple to complex, that motivate use of IR. Users of such systems, particularly academics, have concerns for authoritativeness, timeliness, and contextualization of search. While LLMs may provide functionality that aids the IR process, the continued need for search systems, and research into their improvement, remains essential.

Jin, Q., Leaman, R., & Lu, Z. (2024). **PubMed and beyond: Biomedical literature search in the age of artificial intelligence.** *eBioMedicine*, 100. [https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964\(24\)00023-9/fulltext](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(24)00023-9/fulltext)

Abstract: Biomedical research yields vast information, much of which is only accessible through the literature. Consequently, literature search is crucial for healthcare and biomedicine. Recent improvements in artificial intelligence (AI) have expanded functionality beyond keywords, but they might be unfamiliar to clinicians and researchers. In response, we present an overview of over 30 literature search tools tailored to common biomedical use cases, aiming at helping readers efficiently fulfill their information needs. We first discuss recent improvements and continued challenges of the widely used PubMed. Then, we describe AI-based literature search tools catering to five specific information needs: 1. Evidence-based medicine. 2. Precision medicine and genomics. 3. Searching by meaning, including questions. 4. Finding related articles with literature recommendation. 5. Discovering hidden associations through literature mining. Finally, we discuss the impacts of recent developments of large language models such as ChatGPT on biomedical information seeking.

Li, A., & Sinnamon, L. (2024). **Generative AI search engines as arbiters of public knowledge: An audit of bias and authority.** *Proceedings of the Association for Information Science and Technology*, 61(1), 205-217. <https://arxiv.org/abs/2405.14034>

Abstract: This paper reports on an audit study of generative AI systems (ChatGPT, Bing Chat, and Perplexity) which investigates how these new search engines construct responses and establish authority for topics of public importance. We collected system responses using a set of 48 authentic queries for 4 topics over a 7-day period and analyzed the data using sentiment analysis, inductive coding and source classification. Results provide an overview of the nature of system responses across these systems and provide evidence of sentiment bias based on the queries and topics, and commercial and geographic bias in sources. The quality of sources used to support claims is uneven, relying heavily on News and Media, Business and Digital Media websites. Implications for system users emphasize the need to critically examine Generative AI system outputs when making decisions related to public interest and personal well-being.

Memon, S. A., & West, J. D. (2024, February 18). **Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable**. *arXiv preprint arXiv:2402.11707*.
<https://arxiv.org/abs/2402.11707>

Abstract: In this commentary, we discuss the evolving nature of search engines, as they begin to generate, index, and distribute content created by generative artificial intelligence (GenAI). Our discussion highlights challenges in the early stages of GenAI integration, particularly around factual inconsistencies and biases. We discuss how output from GenAI carries an unwarranted sense of credibility, while decreasing transparency and sourcing ability. Furthermore, search engines are already answering queries with error-laden, generated content, further blurring the provenance of information and impacting the integrity of the information ecosystem. We argue how all these factors could reduce the reliability of search engines. Finally, we summarize some of the active research directions and open questions.

Pfrommer, S., Bai, Y., Gautam, T., & Sojoudi, S. (2024, June 5). **Ranking manipulation for conversational search engines**. *arXiv preprint arXiv:2406.03589*. <https://arxiv.org/abs/2406.03589>

Abstract: Major search engine providers are rapidly incorporating Large Language Model (LLM)-generated content in response to user queries. These conversational search engines operate by loading retrieved website text into the LLM context for summarization and interpretation. Recent research demonstrates that LLMs are highly vulnerable to jailbreaking and prompt injection attacks, which disrupt the safety and quality goals of LLMs using adversarial strings. This work investigates the impact of prompt injections on the ranking order of sources referenced by conversational search engines. To this end, we introduce a focused dataset of real-world consumer product websites and formalize conversational search ranking as an adversarial problem. Experimentally, we analyze conversational search rankings in the absence of adversarial injections and show that different LLMs vary significantly in prioritizing product name, document content, and context position. We then present a tree-of-attacks-based jailbreaking technique which reliably promotes low-ranked products. Importantly, these attacks transfer effectively to state-of-the-art conversational search engines such as perplexity.ai. Given the strong financial incentive for website owners to boost their search ranking, we argue that our problem formulation is of critical importance for future robustness work.

Schultz, C. D., Koch, C., & Olbrich, R. (2024). **Dark sides of artificial intelligence: The dangers of automated decision-making in search engine advertising**. *Journal of the Association for Information Science and Technology*, 75(5), 550-566.
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24798>

Abstract: With the growing use of artificial intelligence, search engine providers are increasingly pushing advertisers to use automated bidding strategies based on machine learning. Such

automated decision-making systems leave advertisers in the dark about the data being used and how they can influence the outcome of the decision-making process. Previous literature on artificial intelligence lacks an understanding of the dangers related to artificially intelligent systems and their lack of transparency. In response, our paper addresses the inherent risks of the automated optimization of advertisers' bidding strategies in search engine advertising. The selected empirical case of a service company therefore demonstrates how data availability can trigger a long-term decline in advertising performance and how search engine advertising performance metrics develop before and after an event of data scarcity. Based on data collected for 525 days, difference-in-differences analysis shows that the algorithmic approach has a considerable and lasting negative impact on advertising performance. Furthermore, the empirical case indicates that self-regulated learning can initialize a downward spiral that gradually impairs advertising performance. Thus, the aim of this study is to increase awareness regarding automated decision-making dangers in search engine advertising and help advertisers take preventive measures to reduce the risks of algorithm missteps.

Shukla, M., Goyal, I., Gupta, B., & Sharma, J. (2024). **A comparative study of ChatGPT, Gemini, and Perplexity.** *International Journal of Innovative Research in Computer Science & Technology*, 12(4), 10-15. <https://ijircst.irpublications.org/index.php/ijircst/article/view/96>

Abstract: Generative AI is making buzz all over the globe and has mostly drawn attention due to its ability to generate variety of content that mimics human behaviour and intelligence along with the ease of access. It comprises of the ability to generate text, images, video, and even audio that are almost unrecognizable from human-created content. Thus there is a huge scope of research in this field due to its vast applicability and motivates this research work. This research work presents comparatively analysis of the three Generative Artificial Intelligence (AI) tool, namely ChatGPT, Gemini, Perplexity AI, based on the content generation, ownership and developing technology, context understanding, transparency, and information retrieval.

Strzelecki, A. (2024). **Is ChatGPT-like technology going to replace commercial search engines?** *Library Hi Tech News*. <https://webm.ue.katowice.pl/strzelecki/papers/2024-lhtn.pdf>

Abstract: This paper aims to give 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. In recent years, advancements in artificial intelligence (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. 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.

Venkit, P. N., Laban, P., Zhou, Y., Mao, Y., & Wu, C. S. (2024). **Search engines in an AI era: The false promise of factual and verifiable source-cited responses.** *arXiv preprint arXiv:2410.22349*. <https://arxiv.org/abs/2410.22349>

Abstract: Large Language Model (LLM)-based applications are graduating from research prototypes to products serving millions of users, influencing how people write and consume information. A prominent example is the appearance of Answer Engines: LLM-based generative search engines supplanting traditional search engines. Answer engines not only retrieve relevant sources to a user query but synthesize answer summaries that cite the sources. To understand these systems' limitations, we first conducted a study with 21 participants, evaluating interactions with answer vs. traditional search engines and identifying 16 answer engine limitations. From these insights, we propose 16 answer engine design recommendations, linked to 8 metrics. An automated evaluation implementing our metrics on three popular engines (this [http URL](#), this [http URL](#), BingChat) quantifies common limitations (e.g., frequent hallucination, inaccurate citation) and unique features (e.g., variation in answer confidence), with results mirroring user study insights. We release our Answer Engine Evaluation benchmark (AEE) to facilitate transparent evaluation of LLM-based applications.

Wei, C. H., Allot, A., Lai, P. T., Leaman, R., Tian, S., Luo, L., ... & Lu, Z. (2024). **PubTator 3.0: An AI-powered literature resource for unlocking biomedical knowledge.** *Nucleic Acids Research*, 235. <https://academic.oup.com/nar/article/52/W1/W540/7640526>

Abstract: PubTator 3.0 (<https://www.ncbi.nlm.nih.gov/research/pubtator3/>) is a biomedical literature resource using state-of-the-art AI techniques to offer semantic and relation searches for key concepts like proteins, genetic variants, diseases and chemicals. It currently provides over one billion entity and relation annotations across approximately 36 million PubMed abstracts and 6 million full-text articles from the PMC open access subset, updated weekly. PubTator 3.0's online interface and API utilize these precomputed entity relations and synonyms to provide advanced search capabilities and enable large-scale analyses, streamlining many complex information needs. We showcase the retrieval quality of PubTator 3.0 using a series of entity pair queries, demonstrating that PubTator 3.0 retrieves a greater number of articles than either PubMed or Google Scholar, with higher precision in the top 20 results. We further show that integrating ChatGPT (GPT-4) with PubTator APIs dramatically improves the factuality and verifiability of its responses. In summary, PubTator 3.0 offers a comprehensive set of features and tools that allow researchers to navigate the ever-expanding wealth of biomedical literature, expediting research and unlocking valuable insights for scientific discovery.