

A Deep Dive into ChatGPT's Web Search Capabilities

[Assem Hijazi – Experience, Knowledge & Ideas](#)

Web searching is changing, and ChatGPT is at the forefront of this transformation.

Traditional search engines like Google and Bing have shaped how we find information online for years, but new technologies driven by AI and large language models (LLMs) are shaking things up.

ChatGPT's new web browsing feature is one of these breakthroughs, offering a fresh way to get information that could solve some of the problems people face with current search engines.

In this series of articles, we're going to explore the key features of ChatGPT's web browsing in a simple and clear way.

Instead of just listing out features, we'll take the time to explain each one so you can see why this approach could be revolutionary—not just for ChatGPT, but for any LLM looking to act as a powerful search tool.

We'll kick things off by diving into the first feature of **Retrieval Augmented Generation (RAG)** that make ChatGPT's search responses so relevant.

In future articles, we'll cover even more features to give you a full picture of this innovative technology.

By the end of this series, you'll understand how AI-driven searching could change the way we get information and how it might even outshine today's search engines in some ways.

So, let's start by exploring RAG —and see what makes them so special.



How ChatGPT Searches the Web

When you think of a typical search engine like Google or Bing, you imagine a system that has a massive database, or index, of web pages.

These search engines use web crawlers to scan and store content from across the internet, creating a huge index that's ready for whenever someone searches for information.

This is why, when you type a query into Google, it instantly provides a long list of links that match your keywords.

But ChatGPT's web searching is different. Instead of using this traditional **indexing** method, ChatGPT relies on something called **RAG**, or Retrieval Augmented Generation.

In Simple Terms:

- **Traditional Search Engines:** Use web crawlers to index the entire internet, creating a massive database that is continuously updated and ready to be searched.
- **ChatGPT's Web Search:** Uses RAG to fetch information on demand, pulling from a set of reliable, curated sources. It does not store or index web pages but retrieves and uses the content only when you ask a question.

What Makes ChatGPT's Web Search Different?

1. No Pre-Built Index:

- Unlike search engines that crawl the entire internet and store billions of pages in a database, ChatGPT doesn't use a pre-built index of web content. This means it doesn't have a giant catalog of web pages waiting to be searched.

2. RAG (Retrieval Augmented Generation) Technology:

- ChatGPT's web search is powered by RAG. Here's how it works:
 - When you ask a question, ChatGPT doesn't search through an index. Instead, it retrieves information from the internet in **real time**.
 - RAG sends out a request to fetch content from specific, reliable sources chosen by OpenAI. These sources might include trustworthy news sites, academic journals, or other authoritative content platforms.
 - The information is then processed and used to generate a response that is up-to-date and directly relevant to your question.

3. Registered, Curated Sources:

Fixed sources

- ChatGPT doesn't access just any website on the internet. It pulls data from a curated list of registered partner sites. These sources are carefully selected to ensure the accuracy and reliability of the content.

No Indexing:

- Even though RAG uses these sources, it doesn't index or store their content permanently. Instead, it accesses the information only when needed, and once the query is complete, the data is not saved.

4. Temporary Data Use:

- The data retrieved is used only for the session and is not stored for future use. This makes ChatGPT's web search more privacy-friendly compared to traditional search engines, which often retain search history and data.

Important questions and answers

1. How to Ask ChatGPT to Search the Web

Enable Web Browsing:

- **Ensure that the web browsing feature is turned on. This setting is usually available in platforms like ChatGPT Plus or other paid versions where web browsing capabilities are included.**
- **Check the settings or preferences in your interface.**

Formulate Your Query:

- **Simply ask ChatGPT your question in plain language. You can be specific about what you are looking for, such as:**
 - **"Search the web for the latest news on AI technology."**
 - **"Find information about the top tourist destinations in Dubai."**
 - **"Look up the recent advancements in quantum computing."**

Provide Context or Instructions (Optional):

- **If you want more tailored results, you can add instructions to your query:**
 - **"Search for scholarly articles on climate change from credible sources only."**
 - **"Look up product reviews for the latest smartphones and summarize them."**
 - **"Search for government websites that explain the new tax laws."**

Limitations and Considerations

Availability:

- The web browsing feature may not be available in all versions of ChatGPT.
- It is typically a feature included in paid or advanced plans.

Instructions:

- You cannot set permanent search preferences or instructions, but you can customize your query each time for more relevant results.

Example Queries to Guide ChatGPT's Web Search:**General Information:**

- “Can you search for a summary of today's top news headlines in UAE?”

Specific Content:

- “Find and summarize the latest research on electric vehicles in China.”

Comparisons:

- “Search for a comparison between the BMW 6 series and the Mercedes Benz GLE series”.

2. Why This Approach Matters:**Real-Time Information:**

- Since ChatGPT uses RAG, it can deliver up-to-date content. You don't have to worry about getting outdated information from a static database.

Quality and Relevance:

- By accessing registered, high-quality sources, ChatGPT aims to provide accurate and relevant answers without the clutter you often get from a traditional search engine.

Privacy:

- With no permanent storage of search data or tracking cookies, ChatGPT respects user privacy, making it a more secure option for web searches.

3. Can the User Add Their Own Customized Sites?**Current Capability:**

- At this moment, users cannot manually add their own customized sites to ChatGPT's list of registered sources. The content sources ChatGPT accesses are curated and managed by OpenAI, which means the selection of partner sites is predefined.

Future Possibility: “Feature proposal 😊 “

- This feature could potentially be developed in the future, especially for enterprise or specialized users who might want more customization.

- However, as of now, there is no built-in option for users to add specific sites.

4. Can Users Add Instructions for Searching?

Limited Control:

- Users can't set permanent search instructions like they would configure settings on a search engine. However, users can give real-time instructions in their queries.
- For example, they can specify that they want information from a particular type of source (like "from government websites only" or "summarize news articles").

Customization Potential:

- OpenAI may consider features like allowing saved search instructions or custom preferences in the future, but this is not currently supported. "Feature proposal 😊"

5. How Are Sites Prioritized out of the searching results?

Prioritization by Relevance and Quality:

- ChatGPT prioritizes information based on relevance to the user's query.
- The advanced algorithms ensure that responses are as accurate and contextually meaningful as possible.

Registered Sites Priority:

- Within the list of registered partner sites, priority is given based on how well the content from a source matches the user's prompt.
- There is no manual ranking by the user; instead, it's algorithmically determined.

6. How Does ChatGPT Display Results from Registered Sites?

Content Summarization and Citations:

- When ChatGPT retrieves information from registered sites, it typically summarizes the content and provides links or citations.
- This makes the results easy to understand while giving you the option to read more from the original source.

Transparency:

- You will see citations or references to let you know where the information came from, which ensures that you can trace the origin of the content.

7. How Large is the Registered Database for Partners?

Size and Scope:

- The exact number of registered partner sites is not publicly disclosed by OpenAI.

- However, the database is designed to cover a broad range of high-quality and authoritative sources across multiple categories (e.g., news, science, education, health).

Coverage:

- The registered sources are curated to ensure comprehensive coverage of popular and essential topics, but the approach is more selective compared to search engines that index millions of sites.

8. Are Only the Links Registered or More Information?**Beyond Just Links:**

- When ChatGPT accesses a registered partner site, it doesn't just register the link. It retrieves and processes content, like the main text of an article or the summary of a study, to generate a relevant and cohesive response.
- However, the system does not store the entire content of these sites permanently.

Data Structure:

- While ChatGPT does not index the web, it may have metadata or structured information about these registered sources to improve the efficiency and accuracy of content retrieval.