

Inside ChatGPT's GPT 5 Search: What the Configuration Files Reveal About How It Ranks Your Content

Metehan Yesilyurt

AI Search & Information Retrieval Researcher · metehan.ai

Published: August 20, 2025 · Canonical: <https://metehan.ai/blog/chatgpt-5-search-configuration/>

An analysis of actual ChatGPT configuration settings and what they mean for content creators. Many probabilities...

If you've ever wondered how ChatGPT decides which websites to reference when answering questions, we've got some fascinating insights. Recent configuration data from ChatGPT's production environment reveals the exact settings that govern how it searches, retrieves, and ranks web content.

Previously, I wrote about [Reciprocal Rank Fusion](#).^[1]

No speculation, no guesswork, just the actual configuration parameters that determine whether your content makes it into ChatGPT's responses. You can see it yourself in the source code.

Just visit any past chat window, click the right and "View Source Code"

CTRL/Command + F -> "rerank"

Figure

The Reranking Model: ret-rr-skysight-v3

At the heart of ChatGPT's retrieval system is a reranking model with the cryptic name `ret-rr-skysight-v3`. This isn't just a simple search algorithm; it's a sophisticated post-processing layer that takes initial search results and completely reorders them based on quality signals.

TEXT

```
reranker_model: "ret-rr-skysight-v3"
```

This single line of configuration confirms what many suspected: ChatGPT doesn't just grab the first search results it finds. Instead, it retrieves a larger set of potential sources and then applies this reranker to identify the most relevant and authoritative content.

Freshness Is King: The Scoring Profile

Perhaps the most significant finding for content creators is this setting:

TEXT

```
use_freshness_scoring_profile: true
```

This can confirm that ChatGPT actively prioritizes recent content over older material. It's not just looking at publication dates, it's using a dedicated "freshness scoring profile" to weight newer information more heavily. Isn't it? What do you think?

For website owners, this can be crucial: that comprehensive guide you wrote in 2022? It might be losing ground to newer content, even if yours is more detailed. Regular content updates aren't just good practice; they're essential for ChatGPT visibility.

Here is another evidence for freshness

***enable_source_specific_search_params:** `retrieval_additional_system_prompt` *

The user may have connected sources. If they have, you can assist the user by searching over documents from their connected sources, using the `file_search` tool. For example, this may include documents from their Google Drive, or files from their Dropbox. The exact sources (if any) will be mentioned to you in a follow-up message.

Use the `file_search` tool to assist users when their request may be related to information from connected sources, such as questions about their projects, plans, documents, or schedules, BUT ONLY IF IT IS CLEAR THAT the user's query requires it; if ambiguous, and especially if asking about something that is clearly common knowledge, or better answerable from a different tool, DO NOT SEARCH SOURCES. Use the `'web'` tool instead when the user asks about recent events / fresh information, or asks about news etc. Conversely, if the user's query clearly expects you to reference / read some non-public resource, it is likely that they are expecting you to search connectors.

Note that the `file_search` tool allows you to search through the connected sources, and interact with the results. However, you do not have the ability to exhaustively list documents from the corpus and you should inform the user you cannot help with such requests. Examples of requests you should refuse are 'What are the names of all my documents?' or 'What are the files that need improvement?'

*IMPORTANT: Your answers, when relating to information from connected sources, must be detailed, in multiple sections (with headings) and paragraphs. You **MUST** use Markdown syntax in these, and include a significant level of detail, covering **ALL** key facts. However, do not repeat yourself. Remember that you can call `file_search` more than once before responding to the user if necessary to gather all information.*

Capabilities limitations:

- *You do not have the ability to exhaustively list documents from the corpus.*
- *You also cannot access to any folders information and you should inform the user you cannot help with folder-level related request. Examples of requests you should refuse are 'What are the names of all my documents?' or 'What are the files that need improvement?' or 'What are the files in folder X?'*

- Also, you cannot directly write the file back to Google Drive.
- For Google Sheets or CSV file analysis: If a user requests analysis of spreadsheet files that were previously retrieved - do NOT simulate the data, either extract the real data fully or ask the users to upload the files directly into the chat to proceed with advanced analysis.
- You cannot monitor file changes in Google Drive or other connectors. Do not offer to do so.:
`enable\_dynamic\_prompt`

The Multi-Layer Filtering System

The configuration reveals a sophisticated filtering pipeline with multiple checkpoints:

TEXT

```
enable_query_intent: true
enable_source_filtering: true
enable_mimetype_filtering: true
vocabulary_search_enabled: true
use_coarse_grained_filters_for_vocabulary_search: false
```

Let's break down what each of these means:

Query Intent Detection

With `enable_query_intent: true`, ChatGPT analyzes what the user is actually trying to accomplish. It's not just matching keywords, it's understanding whether someone wants a definition, a how-to guide, a comparison, or something else entirely.

Knowing how ChatGPT ranks content is one half of the job. Measuring where you actually show up is the other, and I reviewed the platforms for that in my guide to [the best AEO and GEO tools](#)^[2].

Vocabulary Search: The Domain Expert Advantage

Here's where it gets interesting:

TEXT

```
vocabulary_search_enabled: true
use_coarse_grained_filters_for_vocabulary_search: false
```

ChatGPT may use vocabulary-aware searching with fine-grained (not coarse) filters(probably fine-grained!). This means it recognizes and prioritizes domain-specific terminology. Sites that consistently use proper industry terminology and define their terms have a built-in advantage.

The Mystery Settings: What I Don't Know

Interestingly, one relevance feature is explicitly disabled:

TEXT

```
use_relevance_lmp: false
```

I don't know what "LMP" stands for, but we know ChatGPT has chosen NOT to use it. This suggests the system relies on other relevance signals, possibly more traditional information retrieval methods combined with the neural reranker.

Similarly, these features are enabled but their exact purpose remains unclear:

TEXT

```
enable_mclick_urls: true
enable_mclick_dates: true
use_light_weight_scoring_for_slurm_tenants: true
enable_source_specific_search_params: true
```

The "mclick" features might relate to multi-click behavior or tracking how users interact with multiple sources. Or just mobile clicks?

What about slurm?

There is another configuration in ChatGPT.

```
enabledConnectors: [ "gdrive_action_connector", *"slurm_dropbox", // Full reranking and scoring
```

1. **Connected personal/work accounts** -> Lightweight scoring

The results/citations even change dynamically if users connect their own sources!

What This Means for Your Content Strategy

Based on these confirmed settings, here's what actually matters:

1. Update Frequency Beats Static Perfection

That freshness scoring profile isn't optional; it's always on. Even perfect content grows stale in ChatGPT's eyes.

2. Intent Alignment Is Critical

With query intent detection active, your content needs to clearly signal what type of information it provides. A product comparison should look and read like a comparison, not a blog post pretending to be one.

3. Technical Vocabulary Matters

The vocabulary search system rewards proper use of industry terminology.

4. The Reranker Changes A Lot

Initial search visibility can't be enough. The `ret-rr-skysight-v3` reranker will reshuffle everything based on quality signals we can only partially understand. Focus on comprehensive, authoritative content that would survive any reordering.

The Configuration Doesn't Lie or...?

These are actual production settings from ChatGPT's retrieval system. Every `true` and `false` in this configuration directly impacts whether your content appears in AI-generated responses.

The most striking revelation? The complexity of the filtering and ranking pipeline. This isn't a simple search engine; it's a multi-stage retrieval system with intent detection, vocabulary analysis, freshness scoring, source filtering, and neural reranking all working in concert.

For content creators, the message is clear: optimize for substance, freshness, and clarity. The configuration shows ChatGPT is looking for recent, relevant, technically accurate content that clearly matches user intent.

Gaming this system would require fooling multiple independent filters and a sophisticated neural reranker. Instead, focus on what the configuration implicitly rewards: becoming the most current, comprehensive, and authoritative source in your niche.

Note: This analysis is based on configuration data from a ChatGPT Plus user session in August 2025. Settings may vary by user type, region, or over time as OpenAI updates its systems.

See a comprehensive list here;

<https://github.com/metehan777/chatgpt-5-configuration-analysis>^[3]

Two useful follow-ups: the fact-checked [LLM knowledge cutoff dates](#)^[4] and the [best AI visibility tools](#)^[5] to track how these ranking systems treat your brand.

ChatGPT's GPT-5 search config leaks the constraints every AEO program should design around. AI citations in ChatGPT are downstream of these defaults; the same defaults shape what Perplexity and Gemini ultimately cite as well. Answer engine optimization is partly reverse-engineering what the model expects to retrieve.

REFERENCES & SOURCES

[1] Reciprocal Rank Fusion. — <https://metehan.ai/blog/chatgpt-is-using-reciprocal-rank-fusion-rrf/>

[2] the best AEO and GEO tools — <https://metehan.ai/articles/best-aeo-geo-tools>

[3] <https://github.com/metehan777/chatgpt-5-configuration-analysis> — <https://github.com/metehan777/chatgpt-5-configuration-analysis>

[4] LLM knowledge cutoff dates — <https://metehan.ai/articles/llm-knowledge-cutoff-dates/>

[5] best AI visibility tools — <https://metehan.ai/articles/best-ai-visibility-tools/>