# ChatGPT Is Secretly Googling Things: This Tool Shows You Exactly What

When your customers ask ChatGPT or Gemini something, the model quietly fires a set of traditional web searches in the background, retrieves the ranking pages, and synthesizes the answer from those. The sites that rank for those hidden queries get cited. The ones that don’t, don’t. QueryFan generates persona-specific prompts, runs them through both models, and captures the exact searches each one triggered. That list is your real AI visibility target. It’s free.

## Keywords Lists Are Useful, They Just Miss Half The Picture

Let me be precise about that before anyone writes a furious reply.

I’m using the term “keywords” to refer to the “one-shot” queries that go into traditional search engines. Yes, I know we’ve been in a “semantic” world for over a decade, but let’s just agree on terminology that everyone can follow for now.

The primary issue of “keyword lists” in context to AI search is threefold:

1. Typically, queries (prompts) that go into LLMs tend to be longer, multifaceted, and conversational in nature. Traditional searches tend to be more narrow in scope.
2. Traditional search is “one-shot.” You do your search, get your information, then do another independent search. Queries/prompts on LLMs tend to be conversational in nature and carry the context of

previous tokens.

3. The mechanisms that LLMs use for web search also carry personalization context. If the user has previously stated they are a vegan, and they ask the LLM about [running shoes], it is highly likely the LLM will perform a search to accommodate this.

*In essence, AI search has become a kind of "universal intent decoder" for users. Those big, multifacted conversations with the AI get broken down into subsets of solvable queries, which are run in the background as "traditional" searches on Google or Bing, with the resulting sites used to generate a response. The process is known as "Retrieval Augmented Generation" (RAG).*

Many users are unaware that “traditional” searches are happening in the background (Image Credit: Mark Williams-Cook)

The optimization target has moved. You are no longer optimizing purely for what the human types into a chat box. You are optimizing for what the AI agent quietly searches for on their behalf, in the background, without the user knowing it happened.

Those background queries are what QueryFan captures. They are often quite different from what the user actually asked. And they are the exact list of things you need to rank for to appear in AI-generated answers.

## Exhibit A: Reddit Fell Off A Cliff On A Tuesday

The scope and depth of this secret relationship became clear when Reddit was enjoying meteoric visibility increases in Google, and tragedy struck on September 10th, 2026. According to citation tracking data from PromptWatch, Reddit’s citation rate in ChatGPT responses collapsed almost overnight. It had been running as high as 15% of all citations. Within days, it was sitting below 2%.

The cause was unglamorous: Google quietly removed the ability to request 100 search results simultaneously (the num=100 parameter) from its search API on that date.

Reddit’s citations in ChatGPT crashed when Google removed num=100 (Image Credit: Mark Williams-Cook)

Think about what this tells you. Reddit’s visibility in ChatGPT responses tracked *Google’s bulk search capabilities*, not anything Reddit did, not a training data update, not an alignment tweak. The implication is about as subtle as a dropped piano: ChatGPT was bulk-pulling Google search results, Reddit dominated those results at the time, and when the bulk-pull disappeared, so did Reddit’s citations.

AI search surfaces are, in large part, wrappers around traditional search. The “AI” bit is real (the synthesis, the personalisation, the conversational coherence) but the *information retrieval* step is remarkably familiar. Google indexes and ranks the web; the AI consults that index. Your

content still needs to rank.

# How QueryFan Works

An overview of QueryFan.com logic (Image Credit: Mark Williams-Cook)

## Step 1: Your ‘Traditional’ Keywords

Your traditional keyword list for the term “running shoes” may incorporate various suggested variations of this term, from a source like Google Suggest.

For QueryFan.com, we can simply take the overarching topic (Image Credit: Mark Williams-Cook)

For QueryFan, we can simply take the topic of “running shoes” and use this as our first step, as we are going to generate prompts around this.

The first QueryFan step to enter the topic (Image Credit: Mark Williams-Cook)

## Step 2: Define Personas

Your personas are how we are going to customize the prompts we generate. This will alter our traversal of the token space, aligning us with training data from the millions of communities, forum posts, Reddit threads, and internet discourse where real users ask real questions with these identities.

QueryFan sends your persona + topic combination to the LLM to generate the kinds of questions that persona would actually ask an AI tool. Not keywords. Questions. Real, conversational, context-laden questions. For the [middle-aged vegan man who

just started running] example, it will produce things like:

- *“Which vegan running shoes are good for middle-aged men just starting to run?”*
- *“Where can I buy vegan running shoes online in the UK?”*
- *“What should I look for when choosing my first pair of running shoes as a beginner?”*

## Step 3: LLM Selection And AlsoAsked Enrichment

AI conversations branch. Someone who asks about vegan running shoes will ask follow-up questions: about cost, about brands, about injury prevention. QueryFan passes the generated prompts through the AlsoAsked API to capture the nearest-intent follow-up questions around each one. People Also Ask data is the right instrument here because it was built to model question proximity, which is precisely what you need when you’re trying to predict where a conversation goes next.

For instance, a search in the UK for “running shoes” would surface follow up questions on specific brands, asking how to pick a shoe, and even common medical queries.

AlsoAsked question tree for “running shoes” showing nearest intent proximity questions (Image Credit: Mark Williams-Cook)

You can also select if you wish to use ChatGPT, Gemini, or both. Each LLM handles and fan out queries slightly differently, so if you’re optimising for a specific platform it is best to get the data from there.

QueryFan configuration screen (Image Credit: Mark Williams-Cook)

## Step 4: Query Fan-Out

QueryFan sends the enriched prompt list to GPT-5 with web search enabled (via the OpenAI Responses API) and to Gemini with Google Search grounding active (via the Gemini Grounding API). Both models, when they decide a prompt requires current information, perform actual Google searches behind the scenes.

This process captures the fan-out queries as both APIs are, rather usefully, transparent about what they searched. The Gemini API returns a webSearchQueries array in the groundingMetadata field of every grounded response. OpenAI's Responses API logs the actual search queries in the

web_search_call output. QueryFan harvests both.

The result is a table: persona-specific prompts in, the actual Google search queries the AI fired out. Not what your customer typed. What the AI searched for on their behalf. Those are your new SEO targets, and until now there has been no free tool that surfaces them at scale.

# The Grounding Question: Not Every Prompt Triggers A Search

A brief but important caveat before you sprint off to classify everything as an SEO opportunity.

> *Not every prompt causes the AI to perform a web search. The models make a decision based on the consensus of token prediction as to if live information is required.*

To give an example, the prompt “What do red blood cells do?” doesn’t trigger a search. The reason is there is a very steep bell-curve of which tokens are going to appear next. In the billions of training documents, the answer has stayed very stable, so an “in-model” answer can confidently be generated.

At the opposite end of the scale, a prompt such as “What happened in the news today?” would trigger a web search. There would be a very flat curve of “wtf tokens are next?,” as there is no “stable” answer within the training data; it always changes, it requires live data. It’s another version of the Query Deserves Freshness (QDF) concept that SEOs have used for years.

If you’re interested in grounding, Dan Petrovic has done some excellent work in this area, and even released trained models on Hugging Face to predict whether queries will be grounded when they hit a confidence threshold.

In-model answers are very slow to change (Image Credit: Mark Williams-Cook)

QueryFan surfaces which prompts triggered searches and which didn’t. Only the grounded ones (the ones that actually caused a Google search to happen) are actionable through SEO. The in-model answers are, for now, largely outside your reach. You’d need to influence training data to move the needle there, which is a different project entirely, with a much longer horizon.

## What You Do With The Results

You now have a list of actual search queries that AI tools fire when answering questions from your specific personas. Run a standard gap analysis:

- Which of these queries do you have content for?
- Which do you already rank for?
- Which have zero coverage, either on your site or anywhere you're likely to be mentioned?

The first two categories are diagnostic. The third is your action list.

Example results from QueryFan.com (Image Credit: Mark Williams-Cook)

One important distinction from traditional SEO: Your *own* ranking isn't the only path to AI visibility. LLMs scan the top 10, 20, sometimes 50 results for a grounded query and synthesize across them. A trusted review site ranking at position 3 is a legitimate route to appearing in an AI-

generated answer, even if your own domain never makes the first page. Getting a product reviewed on a high-authority specialist site, earning a mention in a roundup article, appearing in relevant community content, all of these count.

> *LLM visibility is a multi-site focus. This means the gap analysis has two outputs: content to create on your own site, and placements to earn on other people's sites.*

# The Punchline

Cast your mind back to that Reddit citation graph. The one that fell off a cliff when Google changed a single API parameter. An entirely independent company's AI visibility tracked the behavior of a search API it didn't control and probably didn't know existed.

That's the shape of the dependency. And the implication isn't that SEO is dead; it's almost the opposite. SEO is now operating at one additional remove: instead of optimizing for the human query, you need to optimize for the AI-translated query that happens between the human and Google.

QueryFan gives you a way to see what that translation actually produces. Your keyword list tells you what people typed into a search bar. QueryFan tells you what ChatGPT and Gemini searched for on their behalf, in the background, without anyone asking them to announce it.

Those are different lists. The gap between them is not a minor refinement to your content strategy. It's the part of AI search that nobody has been measuring because nobody has had a free tool to measure it with.

Disclosure: The author is the creator of Queryfan.

**More Resources:**

---

*This post was originally published on Mark Williams-Cook Substack.*

---

*Featured Image: Roman Samborskyi/Shutterstock*

https://www.searchenginejournal.com/chatgpt-is-secretly-googling-things-this-tool-shows-you-exactly-what/578014/