oncrawl × ROAST



Oncrawl x ChatGPT Webinar Part 2

How to appear in ChatGPT: Practical SEO strategies for AI visibility

Speakers



John Campbell

Head of AI and Innovation
@ROAST



Jérôme Salomon

Senior Technical SEO
@Oncrawl

ROAST oncrawl



Before we start

- **Welcome!**
- There will be a **replay** - you'll receive it by email in a couple days.
- Don't hesitate to **ask questions** in the **dedicated tab** at the bottom of your screen.
- We will answer them at the end of the presentation.



Agenda

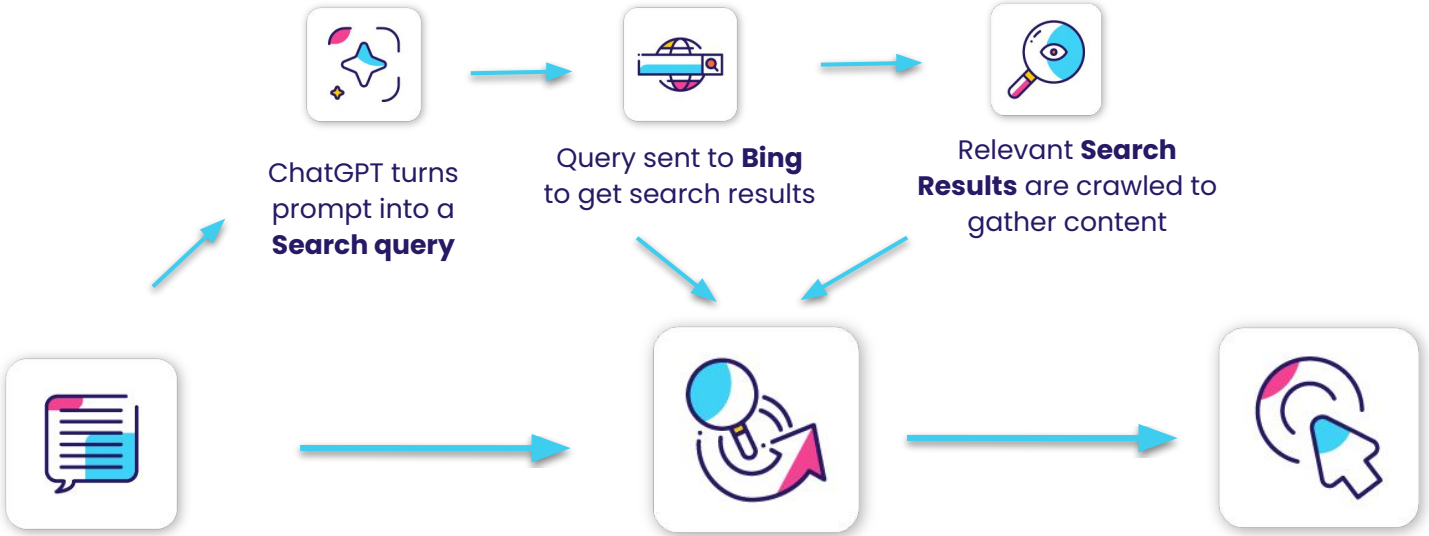
- Quick recap
- ChatGPT Search ranking process
- Visibility tracking tools
- Key takeaways
- Q&A

Quick recap

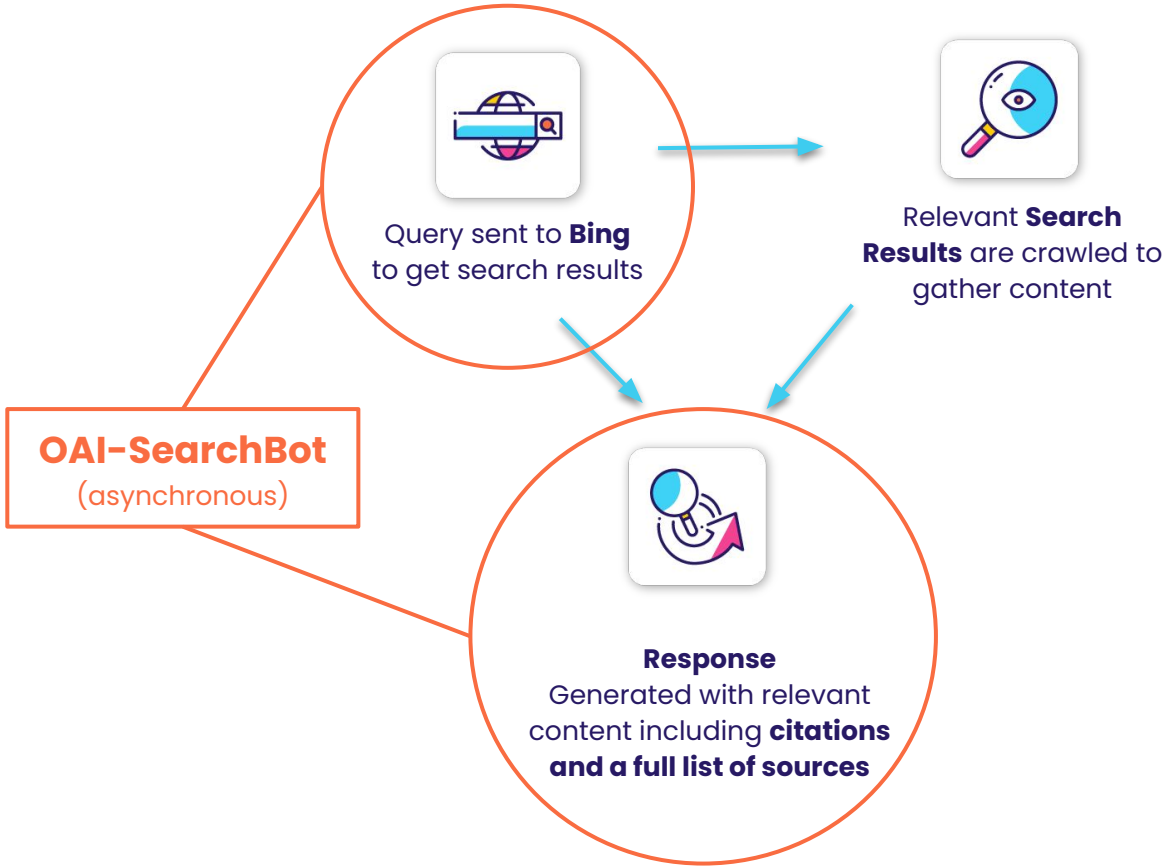
What we've learned

- **ChatGPT-User** and **OAI-SearchBot** are the key bots to monitor.
- You can block GPTBot but still appear in ChatGPT Search results.
- You can **use ChatGPT-User visits as a key visibility metric**.
- You can track **OAI-SearchBot to prevent indexation issues**
- Traditional technical SEO indexability / accessibility still matters.
- You can track your **ChatGPT traffic with logs**.
- Combining all this data will help you monitor your efforts and identify "ranking" patterns.

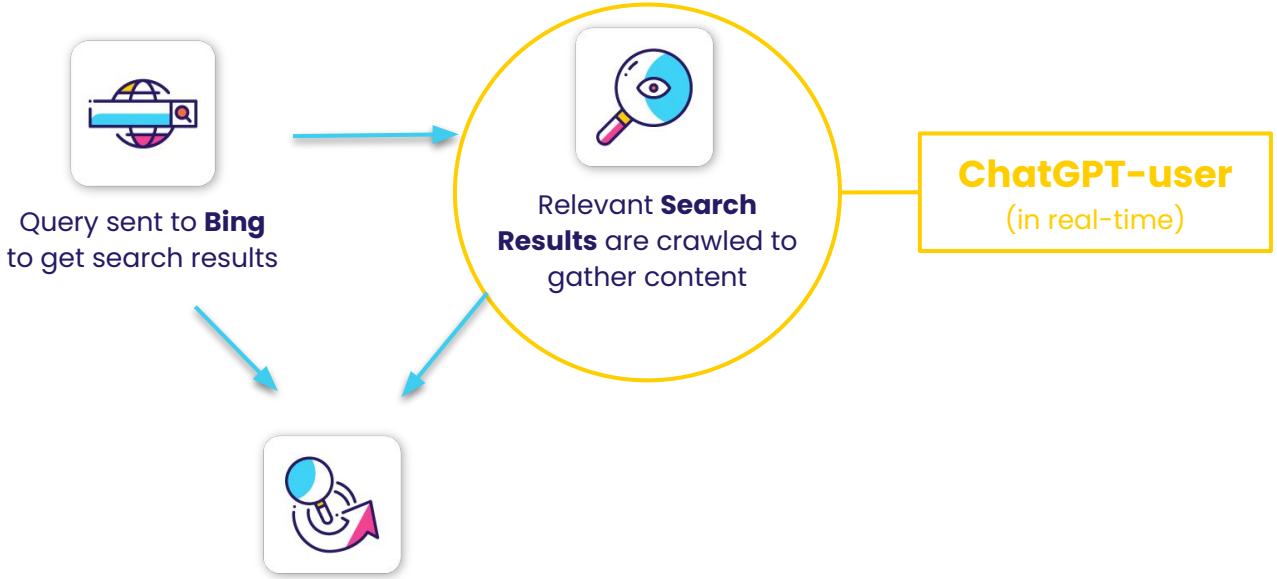
From prompt to click



OpenAI's bots role in the search process

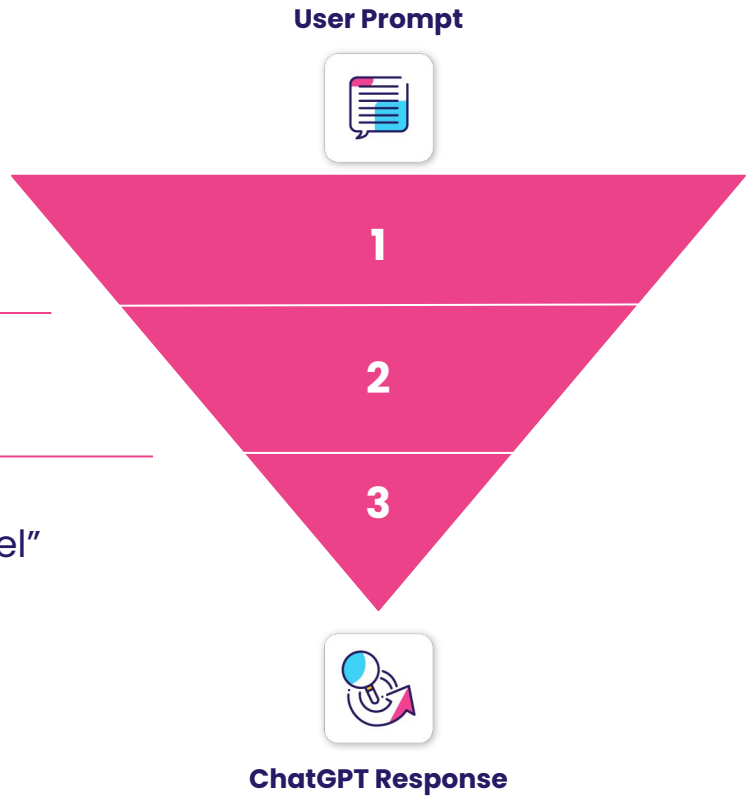


OpenAI's bots role in the search process



Ranking in ChatGPT Search

1. Bing indexation and ranking
2. OpenAI's bots accessibility
3. ChatGPT Search "ranking model"



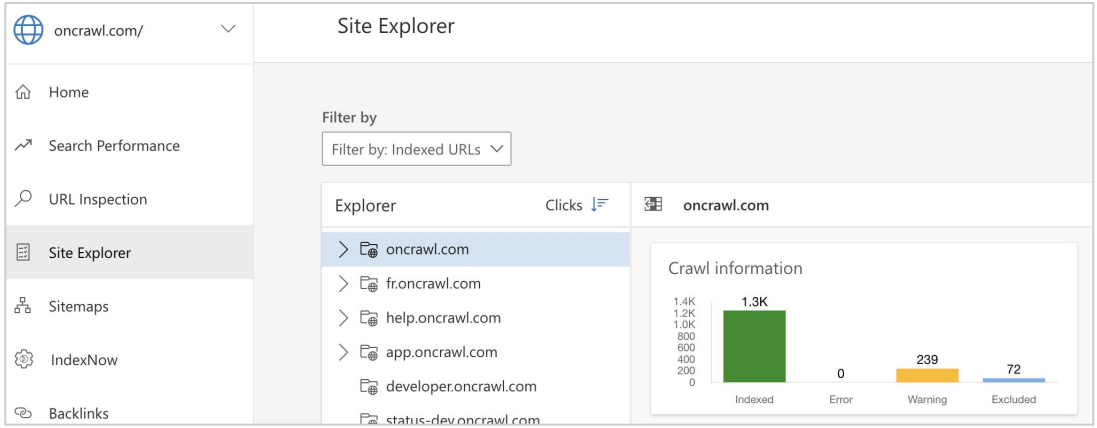
Ranking in Bing

Bing Webmaster tool

Bing indexation and ranking is the key starting point.

Set up your account in a couple clicks: Import your GSC properties

- Check ranking pages and keywords
- Explore indexation state
- Submit sitemaps

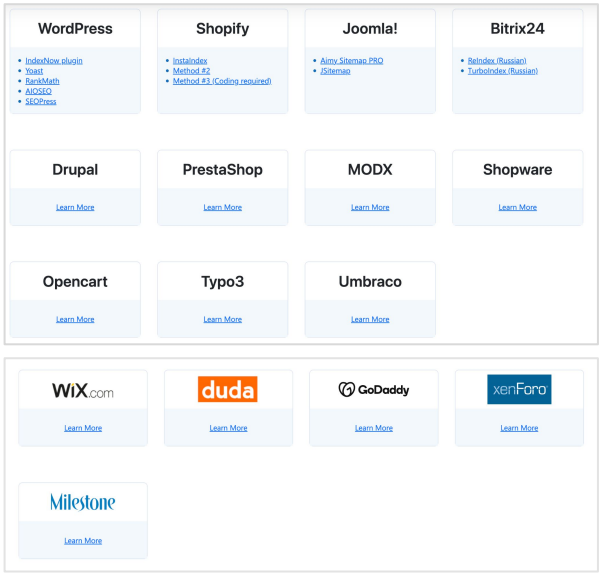


IndexNow

Instantly submit updated or new pages to Bing for indexing.

Easy to put in place:

- Built integration and plugins available on many platforms
- Small HTTP request script



Request

```
POST /IndexNow HTTP/1.1
Content-Type: application/json; charset=utf-8
Host: api.indexnow.org
{
  "host": "www.example.org",
  "key": "ee76ef2857b14cdbbaf5795f16b53309",
  "keyLocation": "https://www.example.org/ee76ef2857b14cdbbaf57
  "urlList": [
      "https://www.example.org/url1",
      "https://www.example.org/folder/url2",
      "https://www.example.org/url3"
      ]
}
```

Content accessibility

Javascript rendering

According to recent studies, the **3 OpenAI bots do not render javascript** and they focus on the raw html:

- Make sure the **key content is available in your html** or with server side rendering
- Use your favorite crawler without JS rendering to verify your website content at scale



Source : vercel.com

Fix status code errors

Logs show that OpenAI's bots **keep crawling 404 pages but follow redirects**:

1. Use your server log files to identify:
   - **Pages crawled by ChatGPT-User and OAI-SearchBot**
   - **With status code 404 (and others client errors)**
   - Order the pages by highest crawl volume (pages that those bots keep visiting)

2. Create a redirect mapping:
   - **Map 4XX pages with relevant available content**
   - Create new pages if necessary
   - Redirect



Use internal linking to boost discovery

We've noticed that ChatGPT can use more than one page from a single domain to answer a question.

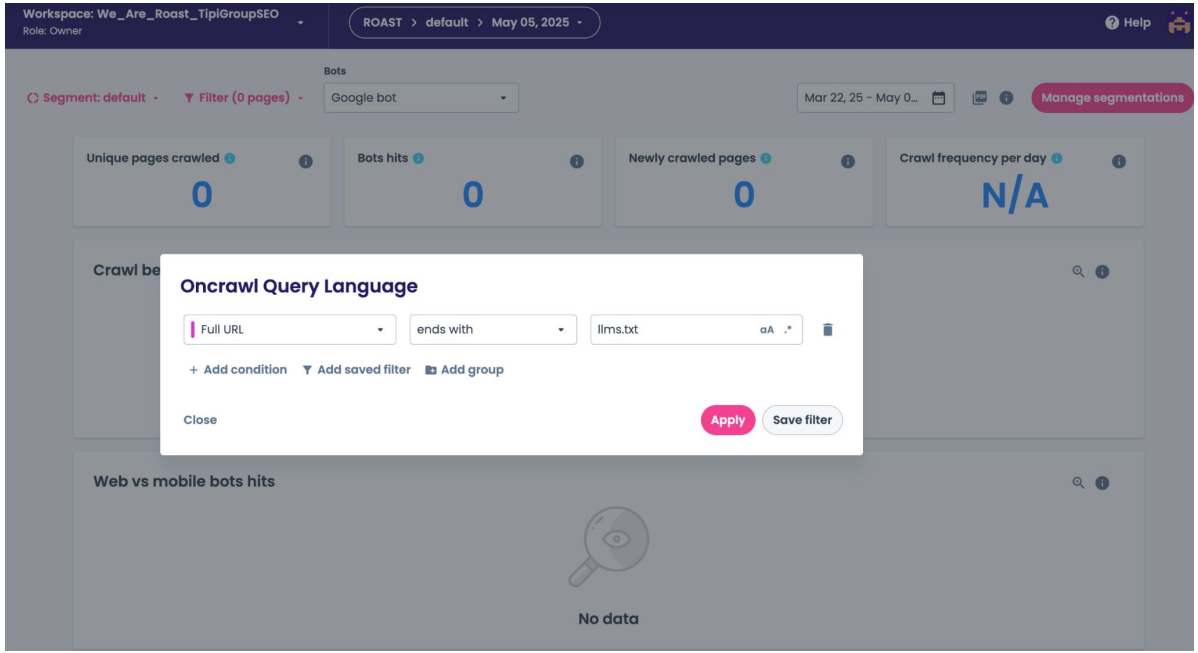
Also, **OpenAI bots keep revisiting the relevant pages** already identified.

You can also **use internal linking to boost pages discovery** from your top crawled pages:

1. Identify ChatGPT-User and OAI-SearchBot top crawled pages in status 200 with your logs
2. Add links to other pages on the same topic (semantic linking strategy)

Is LLMS.TXT useful?

Our clients data shows **no bot hits from OpenAI's Bots on llms.txt**

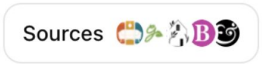


Content relevance

Relevance is ChatGPT's "ranking factor"

"Relevant" is the wording used by OpenAI's documentation and support team define **web content that is used to answer a prompt with Search**.

You can also find all the cited sources, as well as other relevant links, by clicking on the Sources button at the end of the response.



ChatGPT can now search the web in a much better way than before. You can get fast, timely answers with links to relevant web sources, without needing to visit a separate search engine.

What information is shared when I search?

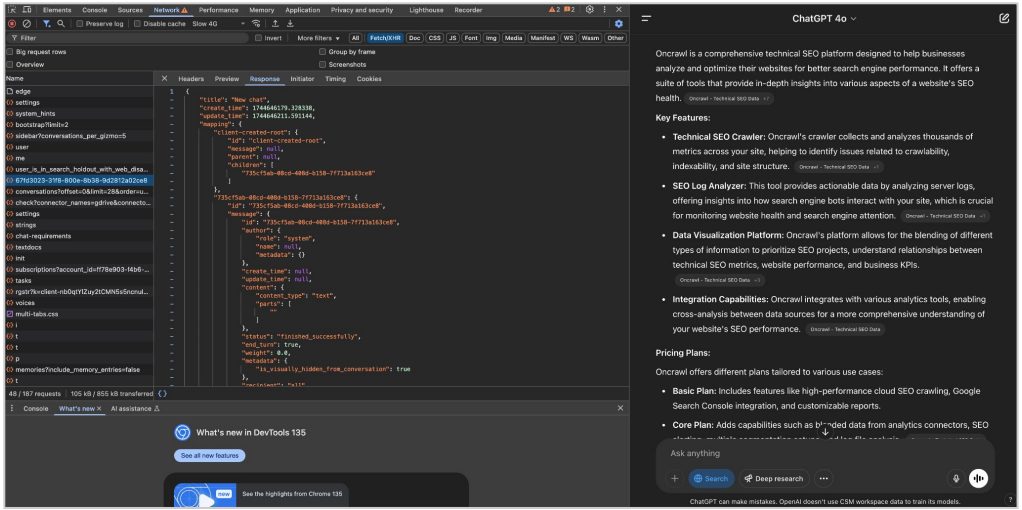
To provide relevant responses to your questions, ChatGPT search sometimes partners with other search providers. When it does, ChatGPT search typically rewrites your query into one or more targeted queries that it sends those providers. For instance, if a biotech researcher asked ChatGPT, "what's the latest on the development of drugs that target CCR8 for cancer?" ChatGPT might initially query a search partner using "CCR8 immunotherapy drug development 2025." After reviewing the initial results, ChatGPT search may send additional, more specific queries to other search providers, like "CHS-114 conference 2025." . ChatGPT also collects general location information based on your IP address and may share that general location with third-party search providers to



ChatGPT can now search the web in a much better way than before. You can get fast, timely answers with links to relevant web sources, which you would have previously needed to go to a search engine for. This blends the benefits of a natural language interface with the value of up-to-date sports scores, news, stock quotes, and more.

ChatGPT Search conversation json file

You can access the json file of a conversation with ChatGPT Search activated to find the full list of links :



1. Login to chatgpt.com and write a prompt with Search activated
2. Right click inspect page
3. Open Network tab
4. Reload page
5. Find json file based url id
6. Look at Response tab
7. Copy paste json and parse to study

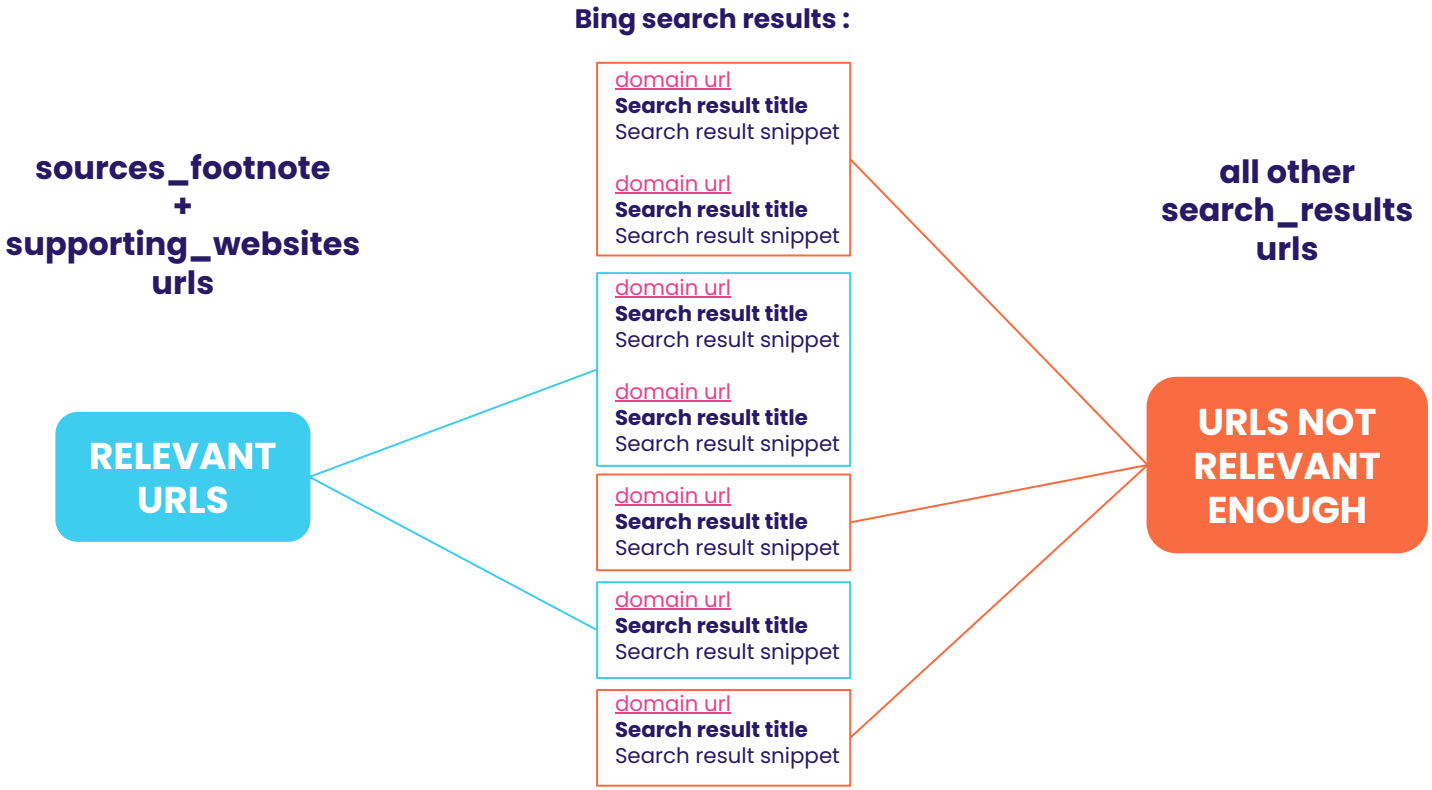
Types of URLs available in the json

- **search_result:**
  Search results urls gathered by ChatGPT for the prompt.
- **sources_footnote:**
  Main citations urls used directly in the answer.
- **supporting_websites:**
  Additional citations urls available on hover when multiple citations are used.
- **safe_urls:**
  Urls that were verified as safe for the user to visit. They are included to provide users with relevant and trustworthy resources.
- **blocked_urls:**
  Urls ChatGPT is not permitted to access or include in responses for safety, content, or policy reasons.





How to identify relevant urls



ROAST oncrawl

Study what defines “relevant” content

1. Test prompts and parse the json to pull the URLs list

2. Use your favorite crawler to extract:
   - Title and description
   - Hn structure
   - Content
   - Structured data
   - Ngrams

3. Compare and analyze:
   - Content gap
   - Unique content
   - Content freshness (Last update)

Visibility tracking tools

Tools to track visibility

DEJAN

Peec AI



Otterly.AI

evertune

Bluefish

Am I on AI?

Brandlight



rankscale

Profound

SEMRUSH

waikay

Source: authoritas.com

ROAST

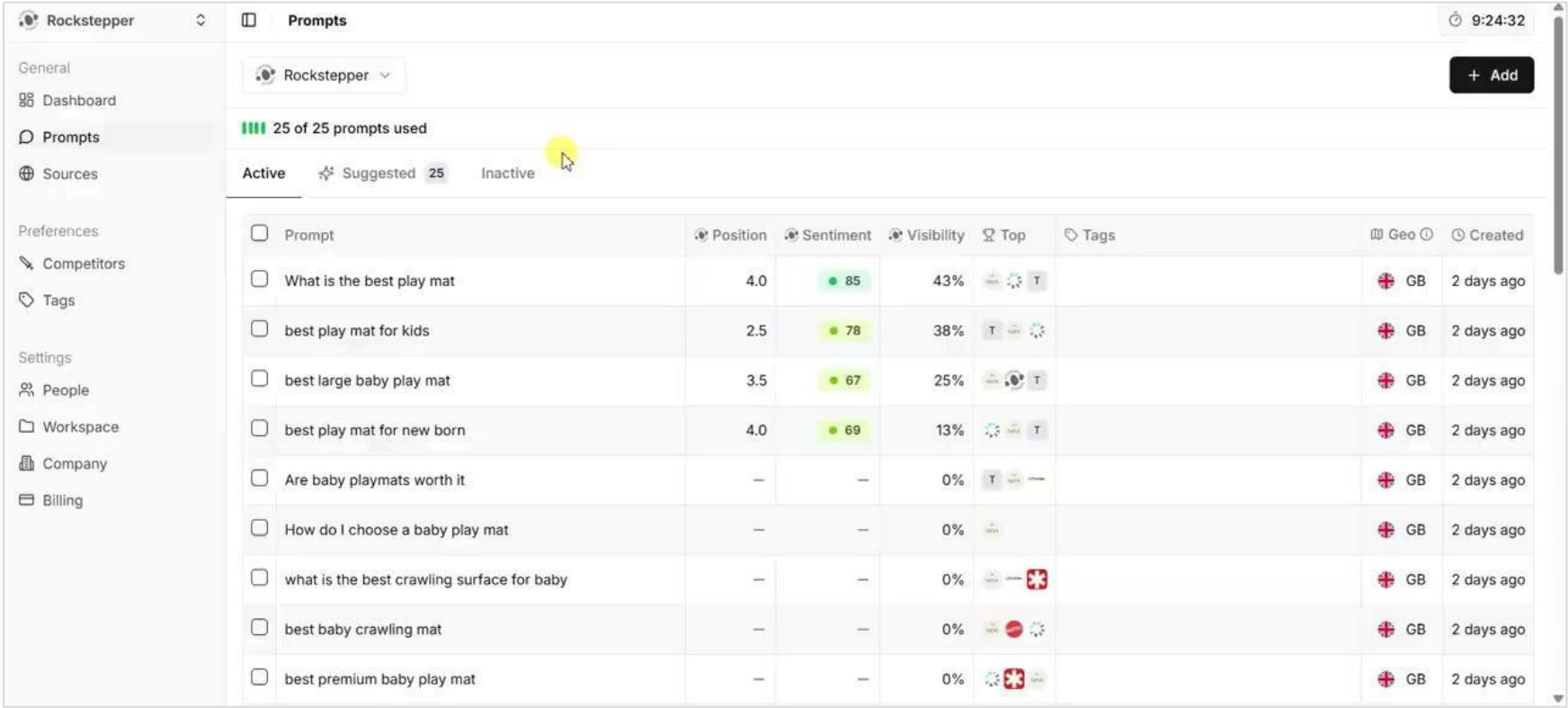
oncrawl

How to choose and what you need to know

- Find out how they are gathering data
  - Do they use APIs or replicate a user e.g. scraping
  - If scraping are they "***clicking***" the search button or letting ChatGPT decide if web search should be used - or ideally both methods
- How often do they pull data?
  - Ideally **multiple times a day** - average out the data
- Ability to download a breakdown on the URLS that are appearing in the results
- Ideally the tool should crawl the URL to see mentions of your brand
- Use key phrases that are likely show citations / mention brands
- Track for a week to get a larger selection of URLs and even our any fluctuations

Source: searchengineland.com

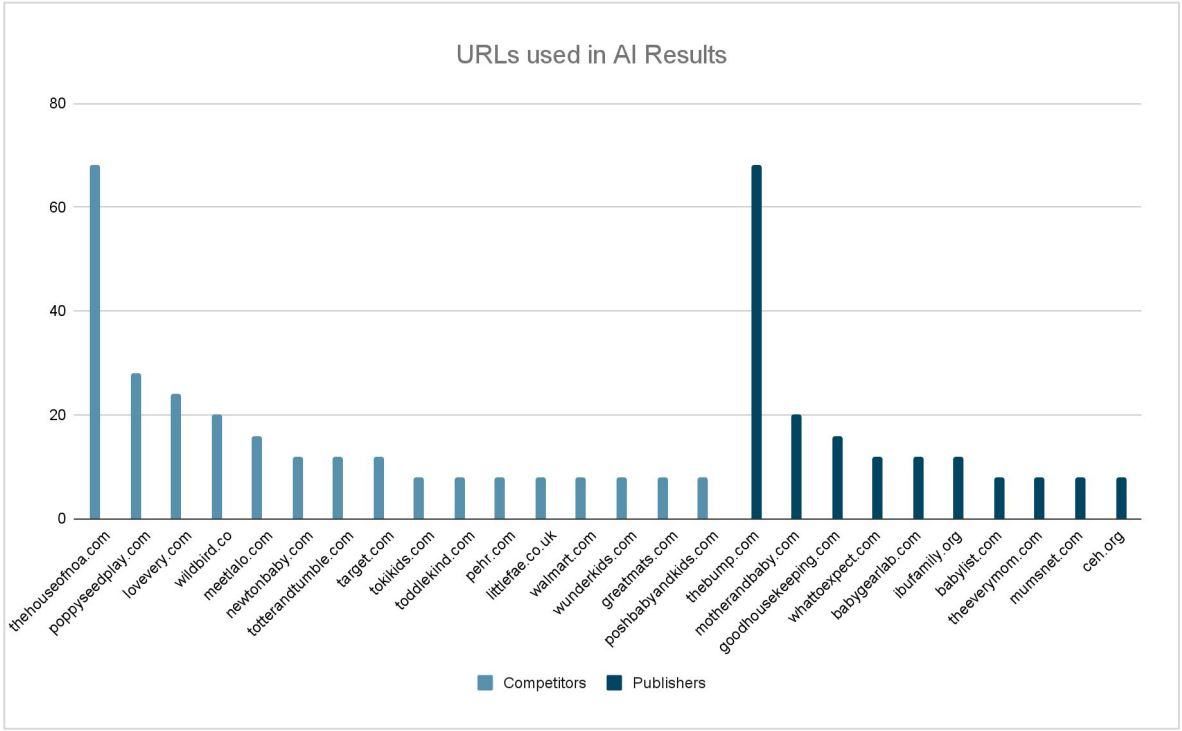
Example – Peec.AI



ROAST oncrawl

Split the URLS from reports into three buckets

- **Your domain** - Your domains that you own / can edit
- **Publishers*** - Where you want to get links / mentioned
- **Competitors** - Monitor - no chance of getting links / mentions



ROAST

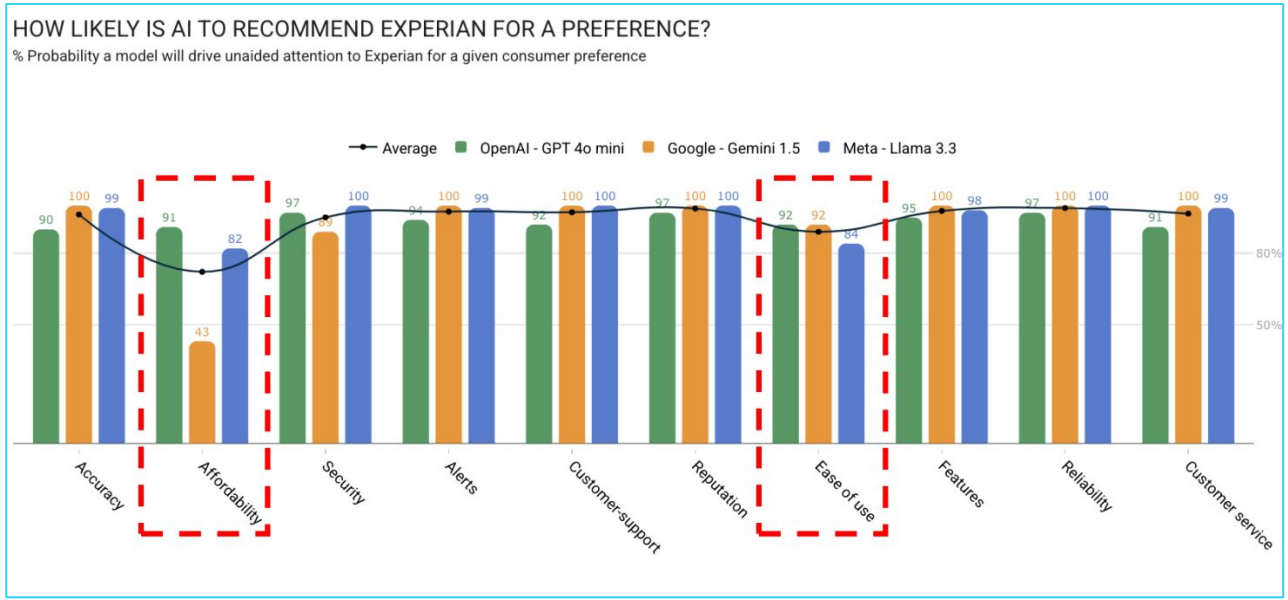
oncrawl

Sources vary according to the platform

	ChatGPT (54 domains)	Perplexity (59 domains)	Google (124 domains)
savingtool.co.uk	**8.0**		
moneyzine.co.uk	3.4		
withplum.com	3.2		
fairinvestment.co.uk	3.0		
yourwalletmanager.com	2.7		
royallondon.com	2.7	3.8	
moneytothemasses.com	2.7	4.4	3.9
thisismoney.co.uk	2.6		
sippadvice.co.uk	2.3	2.1	4.1
aviva.co.uk	2.3	**98.7**	9.3
forbes.com	2.1		
becleverwithyourcash.com	2.0		

Reverse engineering prompt technics

- Understand the decision process the models are using to determine their answers

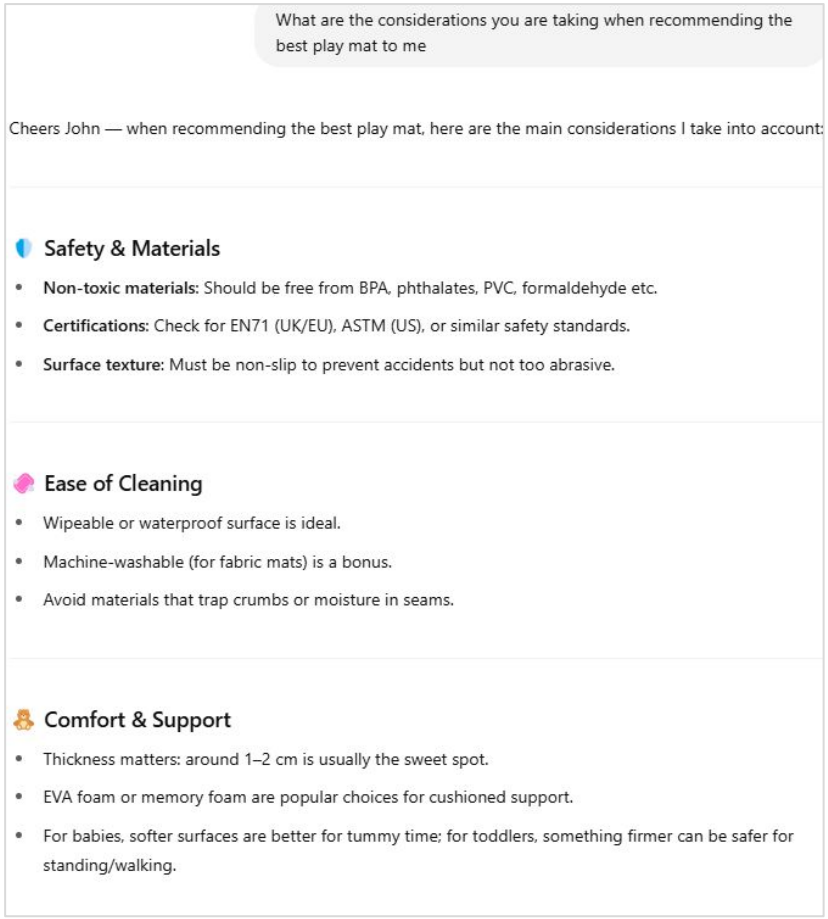


- "**What's the best play mat**?" → "What are the considerations you are taking when recommending the best play mat to me?"

evertune

ROAST

oncrawl



- Do I have relevant content on my website for each of these considerations
- Do I mention these features in my product descriptions
- In external coverage are these topics mentioned next to my brand
- Are there awards, content campaigns, above the line marketing that increase the association of my brand with them
- Which of these are we going to underperform naturally?

ROAST oncrawl

Key takeaways

Key takeaways

- Make sure your content is:
  - **Indexed and ranking in Bing**
  - **Accessible to OpenAI's bots**
- **Content relevance** is the ranking factor
- Use **ChatGPT conversation json file** to identify relevant content
- Visibility tracking tools are in early stages but provide insights on key citations you need to gain coverage on
- Use the **domains** and **urls** listed as citations as a PR / link building hit list
- Create a content plan by asking the platforms the **considerations** are they are taking into their **recommendations**

Things are changing fast

- Early stages of search on AI platforms and the tools that track them - expect changes
- Shopping experience launching, ads (Perplexity), local?
- Fundamentals should remain the same as the platforms develop - relevance and accessibility to bots
- There are still many unknowns → “Key phrase” level data

Q&A

THANK YOU

www.oncrawl.com





ROAST oncrawl