How I Accidentally Triggered 125K Google Impressions Through LLM \"AI News Today Recency\" Bias

Metehan Yesilyurt

AI Search & Information Retrieval Researcher · metehan.ai

Published: December 15, 2025 •

Canonical: https://metehan.ai/blog/how-i-accidentally-triggered-125k-google-impressions-through-llm-ai-news-today-recency-bias/

After publishing my article reviewing the academic research on ChatGPT's recency bias[1] and how brands can become more visible, I noticed an unusual and significant increase in visibility for certain queries in Google Search Console. I didn't even use "best" keyword in my title.

At first, this looked like a standard content performance uplift. However, after digging deeper into the query data, a more interesting pattern emerged.

Figure

LLM-Driven Queries Are Not "Normal" Search Queries

Some of the queries appearing in Search Console were clearly not typical human searches.

They were:

- Highly specific and topic-focused
- Repeated far more frequently than expected
- Semantically aligned with recent AI discussions rather than classic keyword behavior

These queries appeared to reflect how large language models explore topics, validate information, or look for fresh references rather than how humans search.

This was not speculative. The patterns were consistent, repeatable, and clustered around very narrow themes.

Figure

A Possible Link to Training and Evaluation Datasets

I also suspect that part of this behavior may be connected to large-scale training and evaluation datasets such as **MS MARCO, BERT (you can find them at Hugging Face)**.

Just an example, click here and check the numbers.[2]

MS MARCO is widely used across the industry for training and benchmarking retrieval and question answering systems. Especially in open-source models. Its structure favors:

- Natural language questions over short keyword queries
- Topic-focused exploration rather than navigational intent
- Repeated querying of semantically similar questions to evaluate relevance and freshness

When the Search Console queries are examined through this lens, some of the patterns start to make more sense.

Certain queries resemble evaluation-style prompts rather than organic human searches. They look like variations of “does this source still answer this question well” or “is there a fresh, authoritative document for this topic.”

It suggests that the mental model LLMs use when interacting with information retrieval systems may still be strongly influenced by how they were trained and evaluated.

> *I'm thinking in loud. If you somehow create a large dataset for a data poisoning attack, you can influence the web search queries. What do you think? Maybe I'm wrong.*

Figure

Targeted Visibility Is Possible, Even Without Spam

I also noticed that some websites are actively identifying and targeting these LLM-style queries. There are hundreds of those. (Specifically news queries, so you can understand why some "new" websites are appearing more than Reuters or others)

This does not automatically imply spam.

When done responsibly, this approach resembles understanding how AI systems consume, refresh, and validate information, not keyword manipulation. It is about being present where AI systems are already looking for confirmation, updates, and supporting sources.

However, the boundary between insight-driven optimization and low-quality AI slop is thin and easy to cross. *I'm not telling Oreate AI is spamming but others. And this page is getting tons of citations for "latest ai news" query in ChatGPT.[3]*

Figure

A Clear Boundary: No AI Slop, No Mass Exploitation

I want to be explicit about one thing.

I do not support AI slop, synthetic spam pages, or mass-produced content designed purely to game LLM behavior. I will not publish a generalized list of these queries, and I have no interest in enabling that ecosystem.

Instead, I am sharing a single, transparent dataset.

https://docs.google.com/spreadsheets/d/1ctpsfbfinaTzyfN8PdwRlqLMz7axR5CJJYrCv7Klxk0/edit?usp=sharing[4]

Why I Am Publishing Only “AI News Recency” Data

The only dataset I am openly publishing is my full Google Search Console data related to the topic “AI News Recency.”

I am doing this for three reasons:

1. The topic is narrow and well-defined
2. The intent is informational, not exploitative
3. The data clearly demonstrates how recency-focused AI topics can trigger visibility gains without spam tactics

This is not a growth hack playbook. It is an observation backed by real search data.

16 December 2025 Update - It's happening again

Google indexed my post and this is the last 24 hours. Impressions are counting, but no clicks, yet.

Figure

The Bigger Question

What this raises is a broader question for SEO, AEO, and AI search optimization.

If LLMs are generating their own patterns of information-seeking behavior, shaped by training datasets like MS MARCO and reinforced by recency-sensitive retrieval systems, then understanding AI-driven information demand becomes just as important as understanding human intent.

Not to exploit it.But to align with it responsibly.

That, in my view, is where the real opportunity lies.

LLM recency bias is an AEO lever. The same recency effect that drove 125K impressions also shapes AI citations in ChatGPT, Perplexity and Gemini. Answer engine optimization programs schedule around it.

REFERENCES & SOURCES

[1] **recency bias** — https://metehan.ai/blog/i-found-it-in-the-code-science-proved-it-in-the-lab-the-recency-bias-thats-reshaping-ai-search/

[2] **Just an example, click here and check the numbers.** — https://huggingface.co/cross-encoder/ms-marco-MiniLM-L6-v2

[3] **"latest ai news" query in ChatGPT.** — https://www.crescendo.ai/news/latest-ai-news-and-updates

[4] **https://docs.google.com/spreadsheets/d/1ctpsfbfinaTzyfN8PdwRlqLMz7axR5CJJYrCv7Klxk0/edi...** — https://docs.google.com/spreadsheets/d/1ctpsfbfinaTzyfN8PdwRlqLMz7axR5CJJYrCv7Klxk0/edit?usp=sharing