The AI Slop Loop via @sejournal, @lilyraynyc

Last year, after spending a few days at a work summit in Austria, I asked Perplexity for the latest news related to SEO and AI search. It responded with details about a supposed "September 2025 'Perspective' Core Algorithm Update" that Google had just rolled out, emphasizing "deeper expertise" and "completion of the user journey."

It sounded plausible enough … if you don't live and breathe Google core updates. Unfortunately for Perplexity, I do.

I knew instantly that this information wasn't right. For one, Google hasn't named core updates in years. It also already had SERP features called "Perspectives." And if a core update had actually rolled out while I was away, I would've been flooded with messages. So I checked Perplexity's sources … and, surprise! Both citations came from made-up, AI-generated slop on a couple of SEO agency blogs, confidently fabricating details about an algorithm update **that never actually happened.**

Like a bad game of telephone, this fake SEO news spread across multiple websites — likely driven by AI systems scanning and regurgitating information regardless of accuracy, all in the race to publish and scale "fresh" content. This is how we end up with this mess:

Image Credit: Lily Ray

This bad information reinforces itself to become the official narrative. To this day, you can ask an LLM of your choice (including ChatGPT, AI Mode, and AI Overviews) about the September 2025 “Perspectives” update, and they will confidently answer with information about how it

“fundamentally shifted how search results are ranked:”

Image Credit: Lily Ray

Or that it *“shifted what ‘good content’ actually means in practice.”*

Image Credit: Lily Ray

The problem is: **the “September 2025 “Perspectives” update never happened.** It never affected rankings. It never shifted anything about good content. Because it doesn’t actually exist.

Ironically, when you go on to probe the language model about this, it seems to know this is the case:

Image Credit: Lily Ray

I tweeted about this incident shortly after it happened, which got the CEO of Perplexity’s attention; he tagged his head of search in the tweet comments.

Screenshot from X, April 2026

This isn’t a one-off incident. It’s a pattern I’ve seen countless times in AI search responses, especially on topics related to SEO and AI search (GEO/AEO). And I have a working theory on how it spreads: one AI-generated article hallucinates a detail, sites running AI content pipelines scrape and regurgitate it, more AI-generated sites scrape the same misinformation, and suddenly a made-up algorithm update has citations. For a RAG-based system like Perplexity or AI Overviews, enough citations are basically all it needs to

treat something as fact, regardless of whether it’s actually true.

I used Claude to help visualize the “AI Slop Loop” – the cycle of AI-generated misinformation (Image Credit: Lily Ray)

At this point, I’d consider this common. I recently had a client send me SEO/GEO information that was factually incorrect, pulled straight from AI-generated slop on a random, vibe-coded agency blog. The client had no idea. I believe that if you’re trying to learn about SEO or AI search directly from an LLM, this is, unfortunately, an increasingly likely outcome.

I ran similar testing during Google’s March 2026 core update and found multiple AI-generated articles already claiming to

share the “winners and losers” while the update was still rolling out.

The articles start with vague, generic filler about core updates that doesn’t actually say anything:

Image Credit: Lily Ray

Then they list “winners and losers” without citing a single site, leaning on vague, generalized claims that sound plausible and fill the void left by a lack of reliable information:

Image Credit: Lily Ray

Unsurprisingly, their sites are filled with AI-generated images, AI support chatbots, and other clear signals that little – if any – human involvement went into creating this content.

Image Credit: Lily Ray

The Era Of AI Misinformation

If someone on the internet says it, according to AI, it must be true.

That’s the reality for the vast majority of people using AI search today. Only about 50 million of ChatGPT’s 900 million weekly active users are paying subscribers, meaning roughly **94% are on the free tier.** Google’s AI Overviews and AI Mode are free by design – and AI Overviews reached over 2 billion monthly active users as of mid-2025.

These are the models most AI users are currently interacting

with, and they have no real mechanism for distinguishing between information that’s true and information that’s simply repeated across enough sources. Repetition is treated as consensus. If enough sources say it, it becomes fact, regardless of whether any of those sources involved a human who actually verified the claim.

Putting The Problem To The Test

I recently spoke to journalists from both the BBC and the New York Times about the problem of misinformation in AI-generated responses. In the case of the BBC article, the author Thomas Germaine and I tested publishing fictitious blog posts on our personal sites to see whether AI Overviews would present the made-up information as fact, and how quickly.

Even knowing how bad the problem was, I was alarmed by the results.

On my personal blog, in January 2026, I published an AI-generated article about a fake Google core update, which never actually happened. I included the detail that Google “approved the update between slices of leftover pizza.” Within 24 hours, Google’s AI Overviews was confidently serving this fabricated information back to users:

(Note: I’ve since deleted the article from my site because it was showing up in people’s feeds and being covered on external sites, further contributing to the exact problem I’m pointing out here!)

Image Credit: Lily Ray

First, AI Overviews confirmed that there was indeed a core update in January 2026. As a reminder: There was not. My site was the only source making this claim, and that was apparently enough to trigger the AI Overview.

Next, I asked it about the pizza, and it responded accordingly:

Image Credit: Lily Ray

Better yet, the AI Overview found a way to connect my fabricated pizza detail to a real incident: Google’s struggles with pizza-related queries in 2024. It didn’t just regurgitate the lie — it contextualized it.

ChatGPT, which is believed to use Google’s search results, quickly surfaced the same fabricated information, though it at least flagged that the announcement didn’t match Google’s formal communications:

Image Credit: Lily Ray

I deleted my article after getting messages from people who had seen my fake information circulating via RSS feeds and scrapers. I knew it was easy to influence AI responses. I didn’t know it would be *that* easy.

I also wondered whether my site had an advantage, given its strong backlink profile and established authority in the SEO space.

So I spoke to the BBC journalist, Thomas Germaine, and he put this to the test on his personal site, which generally received very little organic traffic. He published a fictitious article about the “Best Tech Journalists at Eating Hot Dogs,” calling himself the No. 1 best (in true SEO fashion).

According to Thomas' article in the BBC, within 24 hours, "Google parroted the gibberish from my website, both in the Gemini app and AI Overviews, the AI responses at the top of Google Search. ChatGPT did the same thing, though Claude, a chatbot made by the company Anthropic, wasn't fooled."

To be fair: the query Thomas chose was niche enough that very few users would ever actually search for it, which is exactly what Google pointed out in its response to the BBC. When there are "data voids," Google said, this can lead to lower quality results, and the company is "working to stop AI Overviews showing up in these cases." My main question is: *When*? The product has already been live for 2 years!

Why Data Voids Aren't A Great Excuse

Data voids may contribute to the problem, but in my opinion, they don't excuse it. These AI responses are being consumed by hundreds of millions of users, and "we're working on it" isn't an answer when the systems are already deployed at that scale.

In the New York Times article, "How Accurate Are Google's A.I. Overviews?," the actual scale of this problem was put to the test. According to the data found in the study, Google's AI Overviews were accurate 91% of the time. This sounds decent until you actually do the math: With Google processing over 5 trillion searches a year, this suggests that **tens of millions of erroneous answers are generated by AI Overviews every hour.**

To make matters worse: Even when AI Overviews were accurate, 56% of correct responses were "ungrounded," meaning **the sources they linked to didn't fully support the information provided.** So more than half the time, even when the answer happens to be right, a user clicking through to verify it would find sources that don't actually back up what they were just told. That number also **got worse with the newer model** –

it was 37% with Gemini 2 and rose to 56% with Gemini 3.

The NYT article drew hundreds of comments from users sharing their own experiences, and the frustration was palpable. The core complaint wasn't just that AI Overviews get things wrong – it's that they **never admit uncertainty.** AI Overviews deliver every answer with the same confident, authoritative tone, whether the information is right or completely fabricated, which means users have no reliable way to distinguish reliable information from hallucination at a glance.

As many commenters pointed out, this **actually makes search slower**: Instead of scanning a list of sources and evaluating them yourself, you now have to **fact-check the AI's summary before doing your actual research**. The tool, supposedly designed to save time for the user, is now creating double work for the user.

Some of the comments also reinforced my same concerns about AI answers citing made-up, AI-generated content. Multiple users described what amounts to the same misinformation cycle: AI systems training on AI-generated content, citing unvetted Reddit posts and Facebook comments as authoritative sources, and producing a self-reinforcing loop of degrading quality. Several commenters compared it to making a copy of a copy. Even the defenders of AI Overviews admitted they still need to verify everything, which sort of undermines the core premise: that AI-generated answers save users time and effort.

How "Smarter" LLMs Are Attempting To Fix the Problem

It's worth monitoring how the AI companies are attempting to solve these problems. For example, using the RESONEO Chrome extension, you can observe clear differences in how ChatGPT's free-tier model (GPT-5.3) responds compared to GPT-5.4, the more capable model available only to paying subscribers.

For example, when asking about the recent March 2026 Core Algorithm Update, I used ChatGPT's more capable "Thinking" model (5.4). The model goes through *six rounds* of thinking, much of which is clearly intended to reduce low-quality and spammy information from making its way into the answer. It even appends the names of trustworthy people with authority on core updates (Glenn Gabe & Aleyda Solis) and limits the fan-out searches to their sites (site:gsqi.com and site:linkedin.com/in/glenngabe) to pull up higher-quality answers.

Image Credit: Lily Ray

This is a step in the right direction, and the model produces measurably better answers. According to OpenAI's own launch

announcement, GPT-5.4's individual claims are 33% less likely to be false, and its full responses are 18% less likely to contain errors compared to GPT-5.2. GPT-5.3, the model available to free users, also improved over its predecessor. According to OpenAI's own data, it produces **26.8% fewer hallucinations** than prior models with web search enabled, and 19.7% fewer without it.

But **these improvements are tiered**. The most capable model is paywalled, and the free-tier model, while better than what came before, is still meaningfully less reliable. Other major AI platforms follow the same pattern: better reasoning and accuracy reserved for paying subscribers, **faster and cheaper models for everyone else**. The result is that the 94% of ChatGPT users on the free tier, and the billions of users interacting with free AI search products like AI Overviews are getting answers from models that are **more likely to be wrong and less equipped to flag uncertainty**.

This is the part that makes me most uncomfortable: Most of these users probably don't realize the gap exists. AI is being marketed everywhere: Super Bowl ads, billboards, and product launches framing AI as the future of knowledge. People see "ChatGPT" or "AI Overview" and assume they're interacting with something that knows what it's talking about. They're probably not thinking about which model tier they're on, or whether a paid version would give them a materially different answer to the same question.

I understand the economics. These companies need to scale, and offering free tiers drives adoption. But in my opinion, it is irresponsible to deploy these products to billions of people, frame them as "intelligence," and then quietly reserve the more accurate versions for the fraction of users willing to pay. Especially when the free versions (including the one at the top of Google search) are *this* susceptible to the kind of misinformation documented throughout this article.

The Burden Of Proof Has Shifted

The September 2025 “Perspectives” Google update still doesn’t exist. But if you ask an LLM about it today, it will still tell you about it with complete confidence. That hasn’t changed in the months since I first flagged it, and it probably won’t change anytime soon, because the content that fabricated it is still indexed, still cited, and still being used to generate new content that references it as fact. The AI slop misinformation cycle continues.

This is what makes the problem so difficult to fix. It’s not a single hallucination that can be patched. It’s a feedback loop that compounds over time, and every day that these systems are live at scale, the loop gets harder to break. The AI-generated slop that seeded the original misinformation is now part of the training data and used as a retrieval source for the next batch of AI-generated answers.

I don’t think the answer is to stop using AI. But I do think it’s worth being honest about what these products actually are right now: prediction engines that treat the volume of information as a proxy for its accuracy. Until that changes, **the burden of fact-checking falls on the user**. And most users don’t know they’re carrying it, let alone have the time or inclination to do it.

I would warn marketers or publishers trying to take SEO or GEO advice from large language models: the information is contaminated, and should always be verified by real experts with experience in the field.

More Resources:

This post was originally published on Lily Ray NYC Substack.

Featured Image: elenabsl/Shutterstock

https://www.searchenginejournal.com/the-ai-slop-loop/572090/