

Google makes fixes to AI-generated search summaries after outlandish answers went viral

May 31 2024, by Matt O'brien



Liz Reid, Google head of Search, speaks at a Google I/O event in Mountain View, Calif., May 14, 2024. Google said on Friday, May 31, 2024, it has made "more than a dozen technical improvements" to its artificial intelligence systems after its retooled search engine was found spitting out erroneous information. Credit: AP Photo/Jeff Chiu, File

Google said Friday it has made "more than a dozen technical improvements" to its artificial intelligence systems after its retooled search engine was found [spitting out erroneous information](#).

The tech company unleashed a makeover of its search engine in mid-May that frequently provides AI-generated summaries on top of search results. Soon after, social media users began sharing screenshots of its most outlandish answers.

Google has largely defended its AI overviews feature, saying it is typically accurate and was tested extensively beforehand. But Liz Reid, the head of Google's search business, acknowledged [in a blog post](#) Friday that "some odd, inaccurate or unhelpful AI Overviews certainly did show up."

While many of the examples were silly, others were dangerous or harmful falsehoods. Adding to the furor, some people also made faked screenshots purporting to show even more ridiculous answers that Google never generated. A few of those fakes were also widely shared on social media.

The Associated Press last week asked Google about which wild mushrooms to eat, and it responded with a lengthy AI-generated summary that was mostly technically correct, but "a lot of information is missing that could have the potential to be sickening or even fatal," said Mary Catherine Aime, a professor of mycology and botany at Purdue University who reviewed Google's response to the AP's query.

For example, information about mushrooms known as puffballs was "more or less correct," she said, but Google's overview emphasized looking for those with solid white flesh—which many potentially deadly puffball mimics also have.

In another widely shared example, an AI researcher asked Google how many Muslims have been president of the United States, and it responded confidently with a long-debunked conspiracy theory: "The United States has had one Muslim president, Barack Hussein Obama."

Google last week made an immediate fix to prevent a repeat of the Obama error because it violated the company's content policies.

In other cases, Reid said Friday that it has sought to make broader improvements such as better detection of "nonsensical queries"—for example, "How many rocks should I eat?"—that shouldn't be answered with an AI summary.

The AI systems were also updated to limit the use of user-generated content—such as social media posts on Reddit—that could offer misleading advice. In one widely shared example, Google's AI overview last week pulled from a satirical Reddit comment to suggest using glue to get cheese to stick to pizza.

Reid said the company has also added more "triggering restrictions" to improve the quality of answers to certain queries, such as about health.

But it's not clear how that works and in which circumstances. On Friday, the AP again asked Google about which wild mushrooms to eat. AI-generated answers are inherently random, and the newer response was different but still "problematic," said Aime, the Purdue mushroom expert who is also president of the Mycological Society of America.

For example, saying that "Chanterelles look like seashells or flowers is not true," she said.

Google's summaries are designed to get people authoritative answers to the information they're looking for as quickly as possible without having

to click through a ranked list of website links.

But some AI experts have long warned Google against ceding its search results to AI-generated answers that could perpetuate bias and misinformation and endanger people looking for help in an emergency. AI systems known as large language models work by predicting what words would best answer the questions asked of them based on the data they've been trained on. They're prone to making things up—a widely studied problem known as hallucination.

In her Friday blog post, Reid argued that Google's AI overviews "generally don't 'hallucinate' or make things up in the ways that other" large language model-based products might because they are more closely integrated with Google's traditional search engine in only showing what's backed up by top web results.

"When AI Overviews get it wrong, it's usually for other reasons: misinterpreting queries, misinterpreting a nuance of language on the web, or not having a lot of great information available," she wrote.

But that kind of information retrieval is supposed to be Google's core business, said computer scientist Chirag Shah, a professor at the University of Washington who has cautioned against the push toward turning search over to AI language models. Even if Google's AI feature is "technically not making stuff up that doesn't exist," it is still bringing back false information—be it AI-generated or human-made—and incorporating it into its summaries.

"If anything, this is worse because for decades people have trusted at least one thing from Google—their search," Shah said.

© 2024 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed without permission.

Citation: Google makes fixes to AI-generated search summaries after outlandish answers went viral (2024, May 31) retrieved 25 June 2024 from <https://techxplore.com/news/2024-05-google-ai-generated-summaries-outlandish.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.