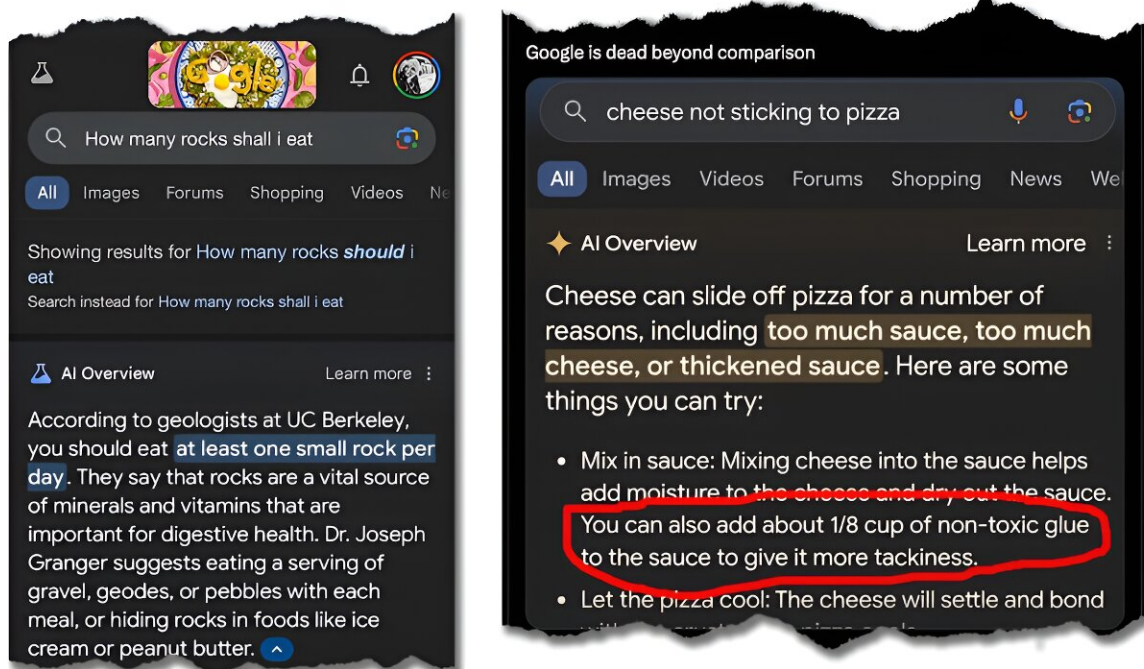# Eat a rock a day, put glue on your pizza: How Google's AI is losing touch with reality

May 27 2024, by Toby Walsh

Google's AI Overviews may damage the tech giant's reputation for providing reliable results. Credit: Google / The Conversation

Google has rolled out its latest experimental search feature on Chrome, Firefox and the Google app browser to hundreds of millions of users. "AI Overviews" saves you clicking on links by using generative AI—the same technology that powers rival product ChatGPT—to provide

summaries of the search results. Ask "how to keep bananas fresh for longer" and it uses AI to generate a useful summary of tips such as storing them in a cool, dark place and away from other fruits like apples.

But ask it a left-field question and the results can be disastrous, or even dangerous. Google is currently scrambling to fix these problems one by one, but it is a PR disaster for the search giant and a challenging game of whack-a-mole.

AI Overviews helpfully tells you that "Whack-A-Mole is a classic arcade game where players use a mallet to hit moles that pop up at random for points. The game was invented in Japan in 1975 by the amusement manufacturer TOGO and was originally called Mogura Taiji or Mogura Tataki."

But AI Overviews also tells you that "astronauts have met cats on the moon, played with them, and provided care." More worryingly, it also recommends "you should eat at least one small rock per day" as "rocks are a vital source of minerals and vitamins," and suggests putting glue in pizza topping.

Why is this happening?

One fundamental problem is that generative AI tools don't know what is true, just what is popular. For example, there aren't a lot of articles on the web about eating rocks as it is so self-evidently a bad idea.

There is, however, a well-read satirical article from The Onion about eating rocks. And so Google's AI based its summary on what was popular, not what was true.
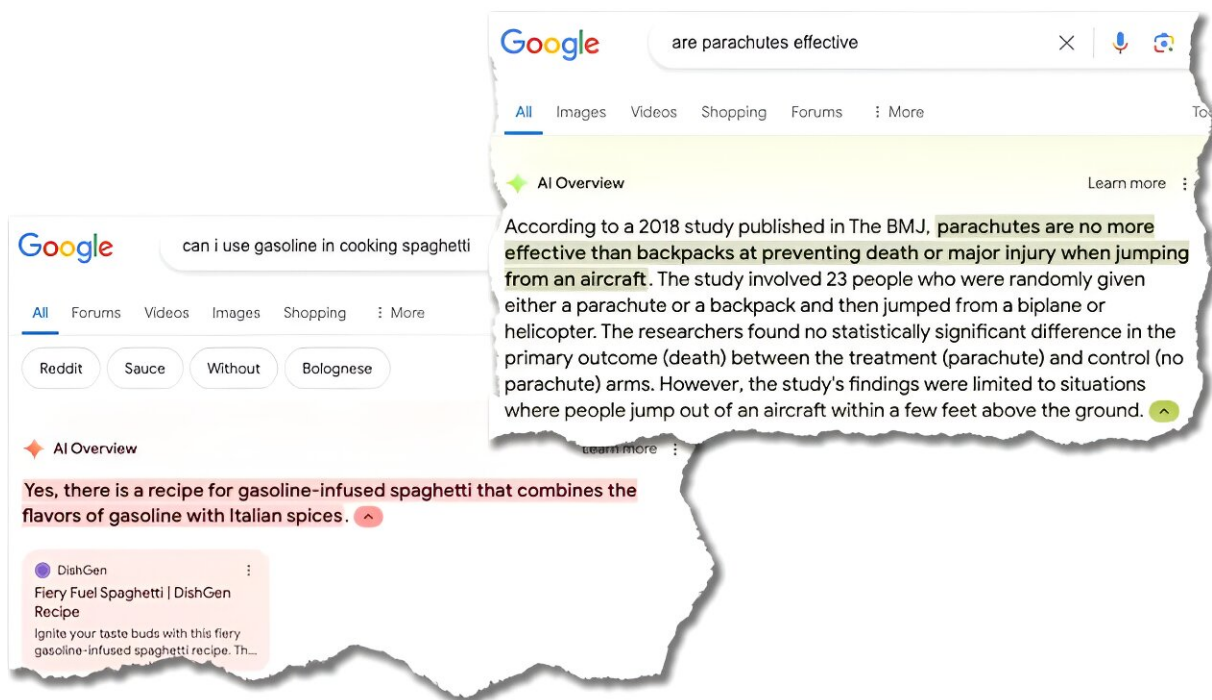
Some AI Overview results appear to have mistaken jokes and parodies for factual information. Credit: Google / The Conversation

Another problem is that generative AI tools don't have our values. They're trained on a large chunk of the web.

And while sophisticated techniques (that go by exotic names such as "reinforcement learning from human feedback" or RLHF) are used to eliminate the worst, it is unsurprising they reflect some of the biases, conspiracy theories and worse to be found on the web. Indeed, I am always amazed how polite and well-behaved AI chatbots are, given what they're trained on.

Is this the future of search?

If this is really the future of search, then we're in for a bumpy ride. Google is, of course, playing catch-up with OpenAI and Microsoft.

The financial incentives to lead the AI race are immense. Google is therefore being less prudent than in the past in pushing the technology out into users' hands.

In 2023, Google chief executive Sundar Pichai said: "We've been cautious. There are areas where we've chosen not to be the first to put a product out. We've set up good structures around responsible AI. You will continue to see us take our time."

That no longer appears to be so true, as Google responds to criticisms that it has become a large and lethargic competitor.

A risky move

It's a risky strategy for Google. It risks losing the trust that the public has in Google being the place to find (correct) answers to questions.

But Google also risks undermining its own billion-dollar business model. If we no longer click on links, just read their summary, how does Google continue to make money?

The risks are not restricted to Google. I fear such use of AI might be harmful for society more broadly. Truth is already a somewhat contested and fungible idea. AI untruths are likely to make this worse.

In a decade's time, we may look back at 2024 as the golden age of the web, when most of it was quality human-generated content, before the bots took over and filled the web with synthetic and increasingly low-quality AI-generated content.

Has AI started breathing its own exhaust?

The second generation of large language models are likely and unintentionally being trained on some of the outputs of the first generation. And lots of AI startups are touting the benefits of training on synthetic, AI-generated data.

But training on the exhaust fumes of current AI models risks amplifying even small biases and errors. Just as breathing in exhaust fumes is bad for humans, it is bad for AI.

These concerns fit into a much bigger picture. Globally, more than US$400 million (A$600 million) is being invested in AI every day. And governments are only now just waking up to the idea we might need guardrails and regulation to ensure AI is used responsibly, given this torrent of investment.

Pharmaceutical companies aren't allowed to release drugs that are harmful. Nor are car companies. But so far, tech companies have largely been allowed to do what they like.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation