

ChatGPT-style search represents a 10x cost increase for Google, Microsoft



reader comments

30 with

Share this story

Is a ChatGPT-style search engine a good idea? The stock market certainly seems to think so, with it erasing \$100 billion from Google's market value after the company's poor showing at its recent AI search event. Actually turning a chatbot into a viable business is going to be a challenge, though. Besides

that fact, Google has had a chat search interface for seven years now—the Google Assistant—and the world’s biggest advertising company has been unable to monetize it. And a new report from Reuters points out another monetary problem with generating a chat session for every search: That’s going to cost a lot more to run compared to a traditional search engine.

Today Google search works by building a huge index of the web, and when you search for something, those index entries gets scanned and ranked and categorized, with the most relevant entries showing up in your search results. Google’s results page actually tells you how long all of this takes when you search for something, and it’s usually less than a second. A ChatGPT-style search engine would involve firing up a huge neural network modeled on the human brain every time you run a search, generating a bunch of text and probably *also* querying that big search index for factual information. The back-and-forth nature of ChatGPT also means you’ll probably be interacting with it for a lot longer than a fraction of a second.

Advertisement

All that extra processing is going to cost a lot more money. After speaking to Alphabet Chairman John Hennessy (Alphabet is Google’s parent company) and several analysts, Reuters writes that “an exchange with AI known as a large language model likely costs 10 times more than a standard keyword search” and that it could represent “several billion dollars of extra costs.”

Exactly how many billions of Google’s \$60 billion in yearly net income will be sucked up by a chatbot is up for debate. One estimate in the Reuters report is from Morgan Stanley, which tacks on a \$6 billion yearly cost increase for Google if a “ChatGPT-like AI were to handle half the queries it receives with 50-word answers.” Another estimate from consulting firm SemiAnalysis claims it would cost \$3 billion.

Google hinted at server time being a problem in its initial post on its “Bard” chatbot, saying it would start with a “lightweight model version” of Google’s language model, and that “this much smaller model requires significantly less computing power, enabling us to scale to more users, allowing for more feedback.” Hearing that Google is being cautious about scale is interesting. Google is *Google*—it already operates at a scale that will dwarf most companies and *could* handle whatever computing load you want to throw at it. “Scale” is only a matter of what Google feels like paying for.

The cost of search is definitely more of a problem for Google than Microsoft. Part of the reason Microsoft is so eager to rock the search engine boat is that most market share estimates put Bing at only about 3 percent of the worldwide search market, while Google is around 93 percent. Search is a primary business for Google in a way that Microsoft doesn’t have to worry about, and with it needing to process 8.5 billion searches every day, Google’s per-search costs can pile up very quickly.

Advertisement

Alphabet’s Hennessy told Reuters that Google is looking into driving down costs, calling it “a couple year problem at worst.” Google has tackled problems like this in the past, like when it bought YouTube and was able to lower costs enough to turn it into a money-making machine, and it continues that today with innovations like building its own video transcoding chips. The company also builds custom server chips for machine learning, called Tensor Processing Units. Still, with Google going on a cost-cutting bloodbath the last few months, suddenly looking ahead to its core consumer product having skyrocketing costs for “a few years” is not ideal.

It’s still not clear how much money anyone is going to make money from chatbots that are supposed to give an answer directly. Google’s and Amazon’s voice assistants have both failed to generate a profit after years of this “we’ll figure

it out later” line of monetization thinking, and those are both just more limited chatbots. OpenAI, the creator of ChatGPT, charges money on a per-word-generated basis, which doesn’t work for search engines (it’s also riding a wave of hype and investor excitement that it can coast on for years). Another Reuters report says that Microsoft has already met with advertisers to detail its plan of “inserting [ads] into responses generated by the Bing chatbot,” but it’s unclear how awkward that would be or if consumers will react when a chatbot suddenly kicks over into an ad break.

For Google, it’s again a question of comparing this new style of chat search engine versus the old one, and it’s unclear if a chatbot interface would result in more or less ad revenue. You could imagine a future where instantly getting a good answer would result in less time on Google compared to having to dig through a list of 10 blue links. If that’s true, then none of the money math on these new search engines looks good.

<https://arstechnica.com/?p=1919322>