Google’s Bard AI can now write and execute code to answer a question

reader comments

34 with

Large language models (LLMs) like ChatGPT and Google Bard can provide *some* decent answers to certain types of questions, but these computers are ironically pretty bad at computing stuff. Google has a new solution to try to get language models to do simple tasks, like math, correctly: have the AI write a program. Google says that now when you ask Bard a “computational” task like math or string manipulation, instead of showing the output of the language model, that language model will instead write a program, execute that program, and then show the output of that program to the user as an answer.

Google’s blog post provides the example input of “Reverse the word ‘Lollipop’ for me.” ChatGPT flubs this question and provides the incorrect answer “pillopoL,” because language models see the world in chunks of words, or “tokens,” and they just aren’t good at this. Here is Bard’s example output:

It gets the output correct as “popilloL,” but more interesting is that it also *includes the python code* it wrote to answer the question. That’s neat for programming-minded people to see under the hood, but wow, is that probably the scariest output ever for regular people. It’s also not particularly relevant. Imagine if Gmail showed you a block of code when you just asked it to fetch email. It’s weird. Just do the job you were asked to do, Bard.

Advertisement

Google likens an AI model writing a program to humans doing long division in that it’s a different mode of thinking:

> *This approach takes inspiration from a well-studied dichotomy in human intelligence, notably covered in Daniel Kahneman’s book Thinking, Fast and Slow— the separation of “System 1”*

> *and “System 2” thinking.*
>
> - *System 1 thinking is fast, intuitive and effortless. When a jazz musician improvises on the spot or a touch-typer thinks about a word and watches it appear on the screen, they’re using System 1 thinking.*
> - *System 2 thinking, by contrast, is slow, deliberate and effortful. When you’re carrying out long division or learning how to play an instrument, you’re using System 2.*
>
> *In this analogy, LLMs can be thought of as operating purely under System 1—producing text quickly but without deep thought. This leads to some incredible capabilities, but can fall short in some surprising ways. (Imagine trying to solve a math problem using System 1 alone: You can’t stop and do the arithmetic, you just have to spit out the first answer that comes to mind.) Traditional computation closely aligns with System 2 thinking: It’s formulaic and inflexible, but the right sequence of steps can produce impressive results, such as solutions to long division.*

Google says this “writing code on the fly” method will also be used for questions like: “What are the prime factors of 15683615?” and “Calculate the growth rate of my savings.” The company says, “So far, we’ve seen this method improve the accuracy of Bard’s responses to computation-based word and math problems in our internal challenge datasets by approximately 30%.” As usual, Google warns Bard “might not get it right” due to interpreting your question wrong or just, like all of us, writing code that doesn’t work the first time.

Bard is coding up answers on the fly right now if you want to give it a shot at bard.google.com.

https://arstechnica.com/?p=1946727