Computer scientist discusses the pros and cons of ChatGPT

February 10 2023

Credit: Pixabay/CC0 Public Domain

With its uncanny ability to mimic human language and reasoning, ChatGPT seems to herald a revolution in artificial intelligence. The nimble chatbot can conjure poems and essays, share recipes, translate languages, dispense advice, and tell jokes, among the endless applications users have tested since the Silicon Valley research lab OpenAI released the natural language-processing tool in November.

With the excitement comes some trepidation—that the technology could degrade authentic human writing and critical thinking, upend industries, and amplify our own prejudices and biases.

To those working in artificial intelligence, ChatGPT is not merely an overnight sensation, but a mark of achievement after years of experimentation, says Johns Hopkins assistant computer science professor Daniel Khashabi, who specializes in language processing and has worked on similar tools.

"ChatGPT may seem like a sudden revolution that came out of nowhere," he says. "But this technology has been developing gradually over many years, with swift progress in the last few."

However, Khashabi acknowledges the unprecedented era that ChatGPT seems to initiate, one brimming with potential for human advancement. "This is really our chance to revise our understanding of what it means to be intelligent," he says. "It's an exciting time because we have this chance to work on new challenges and new horizons that used to feel out of our reach."

As Microsoft invests in the tool, OpenAI releases a paid version, and Google plans to release its own experimental chatbot, the Hub checked in with Khashabi for insight on the technology and where it's headed.

Can you break down how ChatGPT works?

The first stage doesn't involve direct human feedback. The model learns the structure of language by regurgitating vast volumes of text from the web—for example, sentences and paragraphs from Wikipedia, Twitter, Reddit, The New York Times, and so on, and in all different languages. It's also trained in codes written by programmers, from platforms like GitHub.

In the second stage, [generally called "self-supervised" learning], human annotators get involved in training the model to become more sophisticated. They write responses to the various types of queries ChatGPT receives, so the model learns to perform tasks from commands like "write an essay on this topic" or "revise this paragraph."

Because OpenAI is sitting on a gold mine, they can afford to hire many annotators and get them to annotate a lot of high-quality data. I have heard through the grapevine that the initial system was fed close to 100,000 rounds of human feedback. So there's a lot of human labor behind this.

But OpenAI's secret weapon is not its AI technology but the people using its services. Every time someone queries their system, they collect those queries to make the ChatGPT adapt to what users are looking for and identify the weaknesses of their systems. In other words, OpenAI's success was winning over millions to use its demo.

How have you personally been experimenting with ChatGPT?

It can be an excellent writing and brainstorming tool. I can write a summary of an idea I have in mind, ask ChatGPT to expand it more sophisticatedly, then pick the results I like and further develop them myself or continue using ChatGPT. This is human-machine collaborative writing. As someone on Twitter aptly said, "ChatGPT is the e-bike for your brain!"

What do you think of all the attention and press it's getting?

It's another milestone for AI's progress and deserves to be celebrated. It is exciting that AI and natural language processing are getting closer to helping humans with tasks they care about.

However, I worry about overhyping the state of AI. Progress has been made, but "general intelligence" is still not on the horizon. Over the past few decades, we've continually revised our notions of what it means to be "intelligent" every time we progress. In the 1960s and '70s, our goalpost was to create a system to play chess against a human being. There are many examples like this. Every time we progress, we think, "This is it!" But after a while, the hype dies, and we see the problems and identify new needs.

"Intelligence" has always been a moving goalpost and is likely to remain one, but I am excited by the progress I see in working out the shortcomings of ChatGPT-like systems.

What are those shortcomings?

It's easy for ChatGPT to make stuff up. If you ask ChatGPT something niche it hasn't seen before, it will hallucinate facts in fluent and argumentative language. For instance, if you ask it to define "At what tournament did Venus Williams win her eighth grand slam?" it will make up an answer for you, even though Venus Williams has won seven grand slams. She wanted to win her eighth, as many outlets reported, but she didn't. And the model is confusing the two notions of "wanted to win" vs. "won."

And the problem is that it does this so fluently. It can give you garbage, but in such fluent, coherent language that—if you aren't an expert in that domain—you might believe what it is saying is true. That worries me, and I think we humans are gullible in the face of seemingly well-articulated outputs.

On the other hand, what's exciting about ChatGPT?

We now have these tools that can generate creative and fluent language, a challenge we spent years tackling. As an AI scientist, my excitement is about the next steps, and we have new problems for AI to solve.

I am less excited about AI's goal—reverse-engineer human intelligence—and more about IA, or intelligence augmentation. I think it is a worthwhile goal to use AI to enable humans to do better things and to augment human capacity. I'm excited about those kinds of collaborative systems.

How do you see the technology evolving?

We are still in the midst of this change, but we will continue to make language models more efficient, leading to much more compact yet high-quality models. We will consequently witness very reliable forms of conversational agents everywhere. Future models will be your assistants for web navigation, accomplishing various mundane web-based tasks we do ourselves these days.

The same set of technologies is also starting to make its way to the physical world. Current models, such as ChatGPT, don't perceive their environment. For example, they can't see where my phone is or how tired I am. Soon we will see ChatGPT with eyes. These models will use different modalities of data (text, visual, auditory, etc.), which are necessary for them to serve us daily.

This will lead to self-supervised robots based on the data of their physical environments, including physical objects, humans, and their interactions. The impacts here will be enormous. In less than 10 years, any physical appliance we use daily—car, fridge, washer, etc.—will become conversational agents you will talk to. We will also see robots that are incredibly robust in solving problems that are impossible today. Imagine being about to speak with your Roomba about things you wanted to do or not do the way you converse with ChatGPT.

It is also essential not to lose sight of how these technologies will change things on a societal level. The future multimodal models—ChatGPTs with eyes and ears—will be everywhere and will impact everything, including public safety. But now comes the concern: In a society where we are constantly watched by AI models that have eyes and ears and continually get better the more they tend, what will our freedom and privacy look like?

That sounds like a dystopian society described by the famous novel 1984. Like any other technology, self-supervised models are double-edged swords. The best we can do now is to stay vigilant, foreseeing and debating such issues before the applications rise. Ideally, we need to develop frameworks that ensure our freedom and equity by extrapolating from examples such as ChatGPT to its future extensions. I am optimistic that we will.

Provided by Johns Hopkins University

Citation: Computer scientist discusses the pros and cons of ChatGPT (2023, February 10) retrieved 25 April 2024 from https://techxplore.com/news/2023-02-scientist-discusses-pros-cons-chatgpt.html