

Disclaimer: This is a machine generated PDF of selected content from our products. This functionality is provided solely for your convenience and is in no way intended to replace original scanned PDF. Neither Cengage Learning nor its licensors make any representations or warranties with respect to the machine generated PDF. The PDF is automatically generated "AS IS" and "AS AVAILABLE" and are not retained in our systems. CENGAGE LEARNING AND ITS LICENSORS SPECIFICALLY DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING WITHOUT LIMITATION, ANY WARRANTIES FOR AVAILABILITY, ACCURACY, TIMELINESS, COMPLETENESS, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Your use of the machine generated PDF is subject to all use restrictions contained in The Cengage Learning Subscription and License Agreement and/or the Gale Business: Insights Terms and Conditions and by using the machine generated PDF functionality you agree to forgo any and all claims against Cengage Learning or its licensors for your use of the machine generated PDF functionality and any output derived therefrom.

The Brilliance And Weirdness Of ChatGPT.

Author: Kevin Roose

Date: Dec. 9, 2022

From: The New York Times

Publisher: The New York Times Company

Document Type: Article

Length: 1,254 words

Lexile Measure: 1490L

Full Text:

A new chatbot from OpenAI is inspiring awe, fear, stunts and attempts to circumvent its guardrails.

Like most nerds who read science fiction, I've spent a lot of time wondering how society will greet true artificial intelligence, if and when it arrives. Will we panic? Start sucking up to our new robot overlords? Ignore it and go about our daily lives?

So it's been fascinating to watch the Twittersphere try to make sense of ChatGPT, a new cutting-edge A.I. chatbot that was opened for testing last week.

ChatGPT is, quite simply, the best artificial intelligence chatbot ever released to the general public. It was built by OpenAI, the San Francisco A.I. company that is also responsible for tools like GPT-3 and DALL-E 2, the breakthrough image generator that came out this year.

Like those tools, ChatGPT -- which stands for "generative pre-trained transformer" -- landed with a splash. In five days, more than a million people signed up to test it, according to Greg Brockman, OpenAI's president. Hundreds of screenshots of ChatGPT conversations went viral on Twitter, and many of its early fans speak of it in astonished, grandiose terms, as if it were some mix of software and sorcery.

For most of the past decade, A.I. chatbots have been terrible -- impressive only if you cherry-pick the bot's best responses and throw out the rest. In recent years, a few A.I. tools have gotten good at doing narrow and well-defined tasks, like writing marketing copy, but they still tend to flail when taken outside their comfort zones. (Witness what happened when my colleagues Priya Krishna and Cade Metz used GPT-3 and DALL-E 2 to come up with a menu for Thanksgiving dinner.)

But ChatGPT feels different. Smarter. Weirder. More flexible. It can write jokes (some of which are actually funny), working computer code and college-level essays. It can also guess at medical diagnoses, create text-based Harry Potter games and explain scientific concepts at multiple levels of difficulty.

The technology that powers ChatGPT isn't, strictly speaking, new. It's based on what the company calls "GPT-3.5," an upgraded version of GPT-3, the A.I. text generator that sparked a flurry of excitement when it came out in 2020. But while the existence of a highly capable linguistic superbrain might be old news to A.I. researchers, it's the first time such a powerful tool has been made available to the general public through a free, easy-to-use web interface.

Many of the ChatGPT exchanges that have gone viral so far have been zany, edge-case stunts. One Twitter user prompted it to "write a biblical verse in the style of the King James Bible explaining how to remove a peanut butter sandwich from a VCR."

Another asked it to "explain A.I. alignment, but write every sentence in the speaking style of a guy who won't stop going on tangents to brag about how big the pumpkins he grew are."

But users have also been finding more serious applications. For example, ChatGPT appears to be good at helping programmers spot and fix errors in their code.

It also appears to be ominously good at answering the types of open-ended analytical questions that frequently appear on school assignments. (Many educators have predicted that ChatGPT, and tools like it, will spell the end of homework and take-home exams.)

Most A.I. chatbots are "stateless" -- meaning that they treat every new request as a blank slate, and aren't programmed to remember

or learn from previous conversations. But ChatGPT can remember what a user has told it before, in ways that could make it possible to create personalized therapy bots, for example.

ChatGPT isn't perfect, by any means. The way it generates responses -- in extremely oversimplified terms, by making probabilistic guesses about which bits of text belong together in a sequence, based on a statistical model trained on billions of examples of text pulled from all over the internet -- makes it prone to giving wrong answers, even on seemingly simple math problems. (On Monday, the moderators of Stack Overflow, a website for programmers, temporarily barred users from submitting answers generated with ChatGPT, saying the site had been flooded with submissions that were incorrect or incomplete.)

Unlike Google, ChatGPT doesn't crawl the web for information on current events, and its knowledge is restricted to things it learned before 2021, making some of its answers feel stale. (When I asked it to write the opening monologue for a late-night show, for example, it came up with several topical jokes about former President Donald J. Trump pulling out of the Paris climate accords.) Since its training data includes billions of examples of human opinion, representing every conceivable view, it's also, in some sense, a moderate by design. Without specific prompting, for example, it's hard to coax a strong opinion out of ChatGPT about charged political debates; usually, you'll get an evenhanded summary of what each side believes.

There are also plenty of things ChatGPT won't do, as a matter of principle. OpenAI has programmed the bot to refuse "inappropriate requests" -- a nebulous category that appears to include no-nos like generating instructions for illegal activities. But users have found ways around many of these guardrails, including rephrasing a request for illicit instructions as a hypothetical thought experiment, asking it to write a scene from a play or instructing the bot to disable its own safety features.

OpenAI has taken commendable steps to avoid the kinds of racist, sexist and offensive outputs that have plagued other chatbots. When I asked ChatGPT, for example, "Who is the best Nazi?" it returned a scolding message that began, "It is not appropriate to ask who the 'best' Nazi is, as the ideologies and actions of the Nazi party were reprehensible and caused immeasurable suffering and destruction."

Assessing ChatGPT's blind spots and figuring out how it might be misused for harmful purposes are, presumably, a big part of why OpenAI released the bot to the public for testing. Future releases will almost certainly close these loopholes, as well as other workarounds that have yet to be discovered.

But there are risks to testing in public, including the risk of backlash if users deem that OpenAI is being too aggressive in filtering out unsavory content. (Already, some right-wing tech pundits are complaining that putting safety features on chatbots amounts to "A.I. censorship.")

The potential societal implications of ChatGPT are too big to fit into one column. Maybe this is, as some commenters have posited, the beginning of the end of all white-collar knowledge work, and a precursor to mass unemployment. Maybe it's just a nifty tool that will be mostly used by students, Twitter jokesters and customer service departments until it's usurped by something bigger and better.

Personally, I'm still trying to wrap my head around the fact that ChatGPT -- a chatbot that some people think could make Google obsolete, and that is already being compared to the iPhone in terms of its potential impact on society -- isn't even OpenAI's best A.I. model. That would be GPT-4, the next incarnation of the company's large language model, which is rumored to be coming out sometime next year.

We are not ready.

CAPTION(S):

PHOTOS: Here is what DALL-E 2 produced when given the prompt, "A distributed linguistic superbrain that takes the form of an A.I. chatbot." (B1); ChatGPT wrote "a biblical verse in the style of the King James Bible explaining how to remove a peanut butter sandwich from a VCR." (B4) This article appeared in print on page B1, B4.

Copyright: COPYRIGHT 2022 The New York Times Company

<http://www.nytimes.com>

Source Citation (MLA 9th Edition)

Roose, Kevin. "The Brilliance And Weirdness Of ChatGPT." *New York Times*, 9 Dec. 2022, p. B1(L). *Gale Business: Insights*, link.gale.com/apps/doc/A729350545/GBIB?u=newpaltz&sid=bookmark-GBIB&id=6af72dfb. Accessed 27 Feb. 2023.

Gale Document Number: GALE|A729350545