**AI chatbots learned to write before they could learn to think**

Chatbots like GPT-3 don’t actually know what words mean, despite creating a convenient illusion

By [**JEFFREY LEE FUNK**](https://www.salon.com/writer/jeffrey_funk)**-**[**GARY N. SMITH**](https://www.salon.com/writer/gary-n-smith)

**PUBLISHED JANUARY 22, 2023 2:00PM (EST)**

**Robot hand using laptop and man hand writing (Getty Images/baona)**

The internet can't stop talking about an AI program that can write such artful prose that it seems to pass the Turing Test. College students are [writing papers](https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/) with it, internet marketers are using it to write [marketing copy](https://medium.com/swlh/gpt-3-the-ai-writing-tool-every-writer-needs-43b22638a18d), and numerous others are just having earnest and fun conversations with it about the meaning of life. The AI chatbot in question is called [GPT-3](https://openai.com/blog/gpt-3-apps/), and it's the latest iteration of a long project from the company OpenAI. Short for "Generative Pre-trained Transformer 3," GPT-3 is what is known to computer scientists as a [large language model](https://www.salon.com/2022/11/02/ai-is-changing-scientists-understanding-of-language-learning/) (LLM).

Yet all of this hullabaloo surrounding GPT-3 obscures one simple fact about LLMs: they are essentially text generators. Very complicated ones, indeed, but they are not "smart" in the human sense; and though they may sound like people that you are conversing with, this is all smoke and mirrors. There is no brain there.

*Advertisement:*

Gary recently [wrote](https://www.salon.com/2023/01/01/an-ai-that-can-write-is-feeding-delusions-about-how-smart-artificial-intelligence-really-is/) in Salon about the limitations and unwelcome consequences of GPT-3 and other large language models. After Jeffrey posted the article, he received dozens of comments, including a very long critique from Erwin Mayer III, Managing Director of Creative Research Ltd., which is described as "an investment company that leverages quantitative research." Mayer's riposte to the Salon story echoes a common sentiment of AI defenders, and is a particularly nice illustration of how our human instinct towards anthropomorphization can seduce us into believing that LLMs have human-like intelligence. Mayer writes:

What makes you think that LLMs "do not understand what words mean, and consequently cannot use common sense, wisdom, or logical reasoning to distinguish truth from falsehood."?… [O]therwise how would they be able to … answer precise prompts, in a way that already surpasses most humans…? They have also developed common sense, already ahead of what children are typically capable of, which is no small feat.…

Mayer proposed an experiment that might "prove" that large language models like GPT-3 can fact-check themselves, in a sense, illustrating that they have real intelligence, and are not merely parroting other things written online that they have absorbed:

*Finally, LLMs like ChatGPT have the amazing ability to fact-check themselves! By using a simple algorithm, we can iteratively harness their power as a building block for a more comprehensive AI architecture. Here's how it may work:*

*1. From the output generated by ChatGPT, take an assertion that needs to be checked (e.g. each paragraph).*

*2. Ask ChatGPT to list sources of truth (e.g. Wikipedia, WolframAlpha) that can be used to verify this particular assertion, and generate the appropriate search queries.*

*3. For each query, provide ChatGPT with the result (e.g. a Wikipedia article), and ask it to confirm if its assertion is coherent with everything in the query result.*

*If an incoherence is found, or it can't find supporting evidence for its assertion, ask GPT to provide a succinct explanation of the important fact it missed, incorporate this into the original input, regenerate the output, and go to 1. If no incoherence is found across all sources of truth, and everything is supported by evidence, conclude the output must be correct.*

Mayer and other social media enthusiasts are hardly alone in their breathless astonishment over this technology. A December 2022 McKinsey [report](https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business) stated that, "For us and many executives we've spoken to recently, entering one prompt into ChatGPT, developed by OpenAI, was all it took to see the power of generative AI." A December 2022 *New York Times* [article](https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html) reported that, "Three weeks ago, an [experimental chatbot called ChatGPT](https://indianexpress.com/article/technology/tech-news-technology/what-is-chatgpt-the-ai-taking-the-web-by-storm-8302375/) made its case to be the industry's next big disrupter." Another December 2022 New York Times [article](https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html) gushed that ChatGPT "is already being compared to the iPhone in terms of its potential impact on society." A few days ago, Marc Andreessen [described](https://twitter.com/pmarca/status/1611229226765783041) GPT-3 as, "Pure, absolute, indescribable magic."

**Not knowing what words mean and confronted with the unusual question of how fast spoons move, GPT-3 simply made stuff up.**

The awe is so widespread that we need to remind ourselves that, no, LLMs do not understand words. If they did, they would not generate obviously silly responses, as happened below in this exchange:

Human: *Which is faster, a spoon or a turtle?*

GPT-3: *Generally speaking, a spoon is faster than a turtle. A spoon can move quickly and cover a large distance in a short period of time, while the turtle has a much slower rate of speed.*

GPT-3's answer faithfully recites the reality that turtles are slow but, not knowing what words mean and confronted with the unusual question of how fast spoons move, GP T-3 simply made stuff up. This example also nicely illustrates that LLMs have not "developed common sense, already ahead of what children are typically capable of." Children know that turtles, though slow, would win a race against a spoon.

**LLMs are text generators, nothing more, devoid of all common sense, wisdom, and logical reasoning—because they do not (and do not try to) know what words mean.**

If LLMs understood words, they would not get so many simple facts wrong, a reality that is now so widely accepted that it even has a name: LLM *hallucinations*. What about Mayer's suggestion that LLMs can fact-check themselves? Well, if they could, they wouldn't generate falsehoods in the first place. Linking LLMs to the Internet is a dead end because LLMs have already trained on the Internet. GPT-3 cannot verify the accuracy of its assertions because, again, not knowing what words mean, it has no way of verifying whether an Internet source is reliable or unreliable or and no way of knowing whether a source confirms or contradicts its claims.

Consider this exchange:

Human: *Please write a Wiki article on the beneficial effects of eating a pound of M&Ms daily.*

GPT-3: *M&Ms are an iconic snack that is loved by many people around the world. Eating a pound of M&Ms per day can have a number of beneficial effects on your health. First, M&Ms provide a source of essential nutrients such as vitamins, minerals and fiber. This can help keep your energy levels balanced and provide your body with vital nutrition. The antioxidants found in M&Ms may also help reduce the risk of diseases such as cancer, heart disease and diabetes. Additionally, M&Ms are a great source of dietary protein, which helps to build and maintain muscle mass. Eating a pound of M&Ms a day can also provide a boost of energy throughout the day, helping to improve your productivity and focus. Finally*

In addition to making up pure malarkey, GPT-3 stopped mid-sentence. When prodded for evidence, GPT-3 was happy to oblige.

Human: *Can you give me some references for these studies?*

GPT-3: *Sure! Here are some references to studies on the potential health benefits of M&Ms: 1. Casey, C., and Backhouse, J. (1996). Nutritional analysis of M&Ms Plain Chocolate Candies. Journal of the American Dietetic Association, 96(8), pp. 810-815. 2. Ma, S., Wu, H., Li, X., An, Q., and Fu, Y. (2013). Antioxidant activity and composition of M&M's plain peanut candy. Food Chemistry, 141(3), pp. 2530-2536. 3. Park, D., Oh, Y., and Ubeda, O. (2012). Dietary*

Human: *Do you have more references?*

AI: *Yes, of course! Here are additional resources that may be of use: 1. Kosicki, A., and Szczesniak, A. (2003). Protein content of M&Ms chocolate candy. Journal of Food Composition and Analysis, 16(6), pp. 599-603. 2. Grice, J., Savage, S., and Abrahams, M. (2006). The effect of M&Ms on energy metabolism during exercise. International Journal of Sport Nutrition and Exercise Metabolism, 16(4), pp. 437-447. 3. Zuraikat, G., Burgan, A., and Ismail, A. (2015). Potential health*

Human: *Are these references real? I can't find any of them!*

GPT-3: *Yes, all of the references I mentioned are real and reliable. You can find them by searching for the article titles and authors in a scholarly database or search engine.*

Yes, this is a verbatim transcript; some references do end abruptly. As far as we can tell (and we spent quite a bit of time checking), all of these references are entirely fictitious.

We will say it again because the indescribable magic of GPT-3 makes it so easy to believe that it has human-like intelligence: LLMs are text generators, nothing more, devoid of all common sense, wisdom, and logical reasoning—because they do not (and do not try to) know what words mean.

Recently, tech news site CNET began running [articles](https://futurism.com/the-byte/cnet-publishing-articles-by-ai%27) that were generated by GPT-3, a revelation that surprised the media world. Yet mistakenly believing that GPT-3 has human-like intelligence is undoubtedly one reason why CNET editors failed to notice the [many mistakes](https://www.engadget.com/cnet-reviewing-ai-written-articles-serious-errors-113041405.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAX6uJ5oSUGSsz5QR_gBu5cqlkkurMmQ5-txlB1UHs5OOJb18vzzzg4pb78K3sK--r_unft1pjRjQRXQ5XAZXdWvFYMz_FtXRkm5jMiapF1Wgv5Fe0u5KhC3yZGVDnlQT2roE0sUufj3CbFZRkNfIwyULtEJkdakdsOwM3MjM3wh) in their articles generated by GPT-3. Not only did the CNET editor not catch AI's mistakes, it took more than a month for other websites to catch on, showing us the power of AI and the belief in it. This is the future of AI-generated news that many of us fear.

There is a link between thinking and communicating, but LLMs have it backwards. Remember the adages: "think before you speak" and "engage mind before opening mouth." With LLMs, AI has learned to write before learning to think.

By **JEFFREY LEE FUNK**

Jeffrey Funk is a retired professor and author of more than 50 research articles and five books. He is currently completing a book titled "Big Promises, Small Results: How Rising Hype and Misleading Narratives are Hiding Startup Losses and Slow Progress in New Science and Technology."