

Viewpoint: ChatGPT can't think. Consciousness is something entirely different to today's AI

May 18 2023, by Philip Goff



Credit: AI-generated image ([disclaimer](#))

There has been shock around the world at the rapid rate of progress with [ChatGPT](#) and other artificial intelligence created with what's known as large language models (LLMs). These systems can produce text that seems to display thought, understanding and even creativity.

But can these systems really think and understand? This is not a question that can be answered through technological advance, but careful philosophical analysis and argument tells us the answer is no. And without working through these philosophical issues, we will never fully comprehend the dangers and benefits of the AI revolution.

In 1950, the father of modern computing, Alan Turing, [published a paper](#) which laid out a way of determining whether a computer thinks. This is now called "the Turing test." Turing imagined a human being engaged in conversation with two interlocutors hidden from view: one another human being, the other a computer. The game is to work out which is which.

If a computer can fool 70% of judges in a five-minute conversation into thinking it's a person, the computer passes the test. Would passing the Turing test—something which now seems imminent—show that an AI has achieved thought and understanding?

Chess challenge

Turing dismissed this question as hopelessly vague, and replaced it with a pragmatic definition of "thought," whereby to think just means passing the test.

Turing was wrong, however, when he said the only clear notion of "understanding" is the purely behavioral one of passing his test. Although this way of thinking now dominates [cognitive science](#), there is also a clear, everyday notion of "understanding" that's tied to consciousness. To understand in this sense is to consciously grasp some truth about reality.

In 1997, the [Deep Blue AI beat chess grandmaster Garry Kasparov](#). On a purely behavioral conception of understanding, Deep Blue had

knowledge of chess strategy that surpasses any human being. But it was not conscious: it didn't have any feelings or experiences.

Humans consciously understand the rules of chess and the rationale of a strategy. Deep Blue, in contrast, was an unfeeling mechanism that had been trained to perform well at the game. Likewise, ChatGPT is an unfeeling mechanism that has been trained on huge amounts of human-made data to generate content that seems like it was written by a person.

It doesn't consciously understand the meaning of the words it's spitting out. If "thought" means the act of conscious reflection, then ChatGPT has no thoughts about anything.

Time to pay up

How can I be so sure that ChatGPT isn't conscious? In the 1990s, neuroscientist Christof Koch [bet philosopher David Chalmers a case of fine wine](#) that scientists would have entirely pinned down the "neural correlates of consciousness" in 25 years.

By this, he meant they would have identified the forms of [brain activity](#) necessary and sufficient for conscious experience. It's about time Koch paid up, as there is zero consensus that this has happened.

This is because consciousness can't be observed by looking inside your head. In their attempts to find a connection between brain activity and experience, neuroscientists must rely on their subjects' testimony, or on external markers of consciousness. But there are multiple ways of interpreting the data.

[Some scientists](#) believe there is a close connection between consciousness and reflective cognition—the brain's ability to access and use information to make decisions. This leads them to think that the

brain's prefrontal cortex—where the high-level processes of acquiring knowledge take place—is essentially involved in all conscious experience. Others deny this, [arguing instead that](#) it happens in whichever local brain region that the relevant sensory processing takes place.

Scientists have good understanding of the brain's basic chemistry. We have also made progress in understanding the high-level functions of various bits of the brain. But we are almost clueless about the bit in-between: how the high-level functioning of the brain is realized at the cellular level.

People get very excited about the potential of scans to reveal the workings of the brain. But fMRI ([functional magnetic resonance imaging](#)) has a very low resolution: [every pixel](#) on a brain scan corresponds to 5.5 million neurons, which means there's a limit to how much detail these scans are able to show.

I believe progress on consciousness will come when we understand better how the brain works.

Pause in development

As I argue in my forthcoming book ["Why? The Purpose of the Universe,"](#) consciousness must have evolved because it made a behavioral difference. Systems with consciousness must behave differently, and hence survive better, than systems without consciousness.

If all behavior was determined by underlying chemistry and physics, natural selection would have no motivation for making organisms conscious; we would have evolved as unfeeling survival mechanisms.

My bet, then, is that as we learn more about the brain's detailed workings, we will precisely identify which areas of the [brain](#) embody consciousness. This is because those regions will exhibit behavior that can't be explained by currently known chemistry and physics. Already, [some neuroscientists](#) are seeking potential new explanations for [consciousness](#) to supplement the basic equations of physics.

While the processing of LLMs is now too complex for us to fully understand, we know that it could in principle be predicted from known physics. On this basis, we can confidently assert that ChatGPT is not conscious.

There are many dangers posed by AI, and I fully support the recent call by tens of thousands of people, including tech leaders Steve Wozniak and Elon Musk, [to pause](#) development to address safety concerns. The potential for fraud, for example, is immense. However, the argument that near-term descendants of current AI systems will be super-intelligent, and hence a major threat to humanity, is premature.

This doesn't mean current AI systems aren't dangerous. But we can't correctly assess a threat unless we accurately categorize it. LLMs aren't intelligent. They are systems trained to give the outward appearance of human intelligence. Scary, but not that scary.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Viewpoint: ChatGPT can't think. Consciousness is something entirely different to today's AI (2023, May 18) retrieved 5 May 2024 from <https://techxplore.com/news/2023-05-viewpoint-chatgpt-consciousness-today-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.