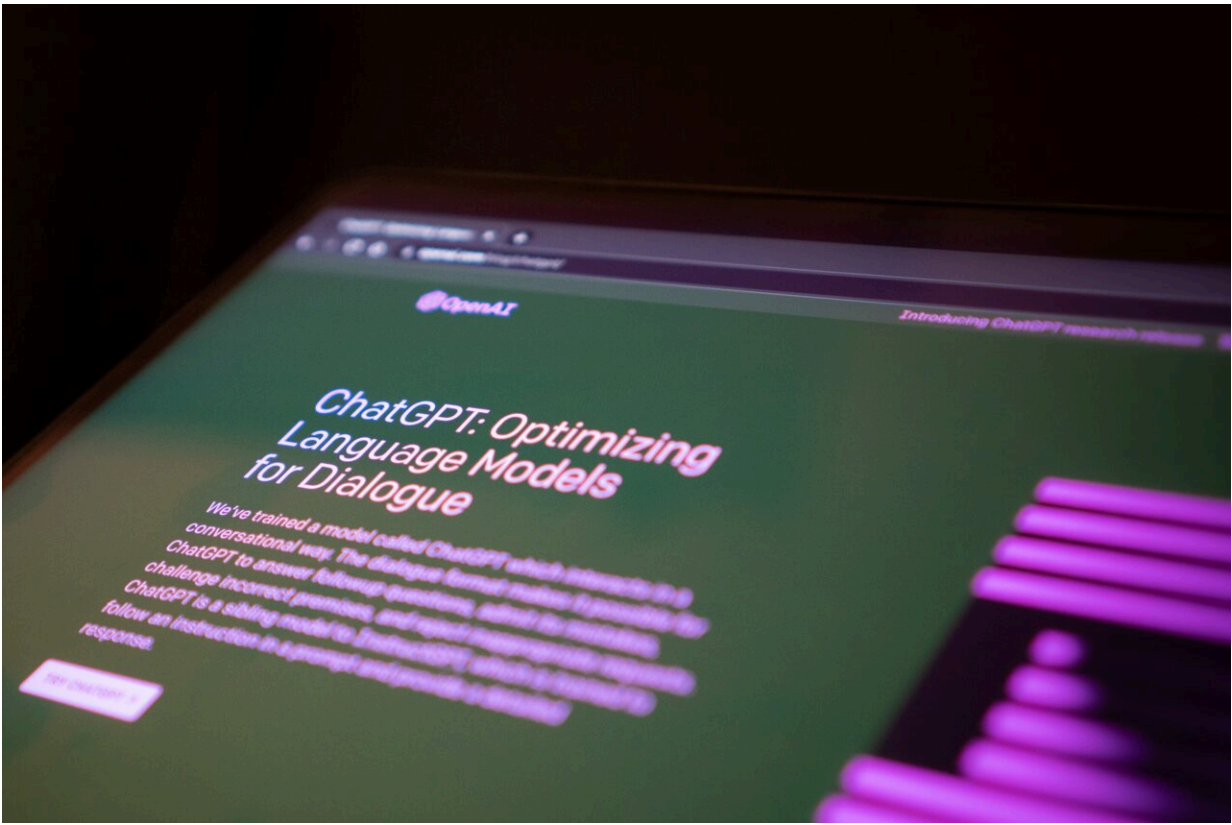


It takes a body to understand the world—why ChatGPT and other language AIs don't know what they're saying

April 6 2023, by Arthur Glenberg and Cameron Robert Jones



Credit: Unsplash/CC0 Public Domain

When we asked [GPT-3](#), an extremely powerful and popular artificial intelligence language system, whether you'd be more likely to use a

paper map or a stone to fan life into coals for a barbecue, it preferred the stone.

To smooth your wrinkled skirt, would you grab a warm thermos or a hairpin? GPT-3 suggested the hairpin.

And if you need to cover your hair for work in a fast-food restaurant, which would work better, a paper sandwich wrapper or a hamburger bun? GPT-3 went for the bun.

Why does GPT-3 make those choices when most people choose the alternative? Because GPT-3 does not understand language the way humans do.

Bodiless words

One of us is a [psychology researcher](#) who over 20 years ago presented a series of scenarios like those above to [test the understanding of a computer model of language](#) from that time. The model did not accurately choose between using rocks and maps to fan coals, whereas humans did so easily.

The other of us is a [doctoral student in cognitive science](#) who was part of a team of researchers that more recently [used the same scenarios to test GPT-3](#). Although GPT-3 did better than the older model, it was significantly worse than humans. It got the three scenarios mentioned above completely wrong.

GPT-3, the engine that powered the initial release of ChatGPT, learns about language by noting, from a trillion instances, which words tend to follow which other words. The strong statistical regularities in language sequences allow GPT-3 to learn a lot about language. And that sequential knowledge often allows ChatGPT to produce reasonable sentences,

essays, poems and computer code.

Although GPT-3 is extremely good at [learning the rules of what follows what](#) in [human language](#), it doesn't have the foggiest idea what any of those words mean to a human being. And how could it?

Humans are biological entities that evolved with bodies that need to operate in the physical and social worlds to get things done. Language is a tool that helps people do that. GPT-3 is an artificial software system that predicts the next word. It does not need to get anything done with those predictions in the real world.

I am, therefore I understand

The meaning of a word or sentence is [intimately related to the human body](#): people's abilities to act, to perceive and to have emotions. Human cognition is empowered by being embodied. People's understanding of a term like "paper sandwich wrapper," for example, includes the wrapper's appearance, its feel, its weight, and, consequently, how we can use it: for wrapping a sandwich. People's understanding also includes how someone can use it for myriad other opportunities it affords, such as scrunching it into a ball for a game of hoops, or covering one's hair.

All of these uses arise because of the nature of human bodies and needs: People have hands that can fold paper, a head of hair that is about the same size as a sandwich wrapper, and a need to be employed and thus follow rules like covering hair. That is, people understand how to make use of stuff in ways that are [not captured in language-use statistics](#).

GPT-3, its successor, [GPT-4](#), and its cousins [Bard](#), [Chinchilla](#) and [LLaMA](#) do not have bodies, and so they cannot determine, on their own, which objects are foldable, or the many other properties that the psychologist J.J. Gibson called [affordances](#). Given people's hands and

arms, paper maps afford fanning a flame, and a thermos affords rolling out wrinkles.

Without arms and hands, let alone the need to wear unwrinkled clothes for a job, GPT-3 cannot determine these affordances. It can only fake them if it has run across something similar in the stream of words on the internet.

Will a large-language-model AI ever understand language the way humans do? In our view, not without having a humanlike body, senses, purposes and ways of life.

Toward a sense of the world

GPT-4 was trained on images as well as text, permitting it to learn statistical relationships between words and pixels. While we can't perform our original analysis on GPT-4 because it currently doesn't output the probability it assigns to words, when we asked GPT-4 the three questions, it answered them correctly. This could be due to the model's learning from previous inputs, or its increased size and visual input.

However, you can continue to construct new examples to trip it up by thinking of objects that have surprising affordances that the model likely hasn't encountered. For example, GPT-4 says that a cup with the bottom cut off would be better for holding water than a lightbulb with the bottom cut off.

A model with access to images might be something like a child who learns about language—and the world—from the television: It's easier than learning from the radio, but humanlike understanding will require the crucial opportunity to [interact with the world](#).

Recent research has taken this approach, training language models to [generate physics simulations](#), [interact with physical environments](#) and even [generate robotic action plans](#). Embodied language understanding might still be a long way off, but these kinds of multisensory interactive projects are crucial steps on the way there.

ChatGPT is a fascinating tool that will undoubtedly be used for good—and not-so-good—purposes. But don't be fooled into thinking that it understands the words it spews, let alone [that it's sentient](#).

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

Provided by The Conversation

Citation: It takes a body to understand the world—why ChatGPT and other language AIs don't know what they're saying (2023, April 6) retrieved 6 December 2023 from <https://techxplore.com/news/2023-04-body-worldwhy-chatgpt-language-ais.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.