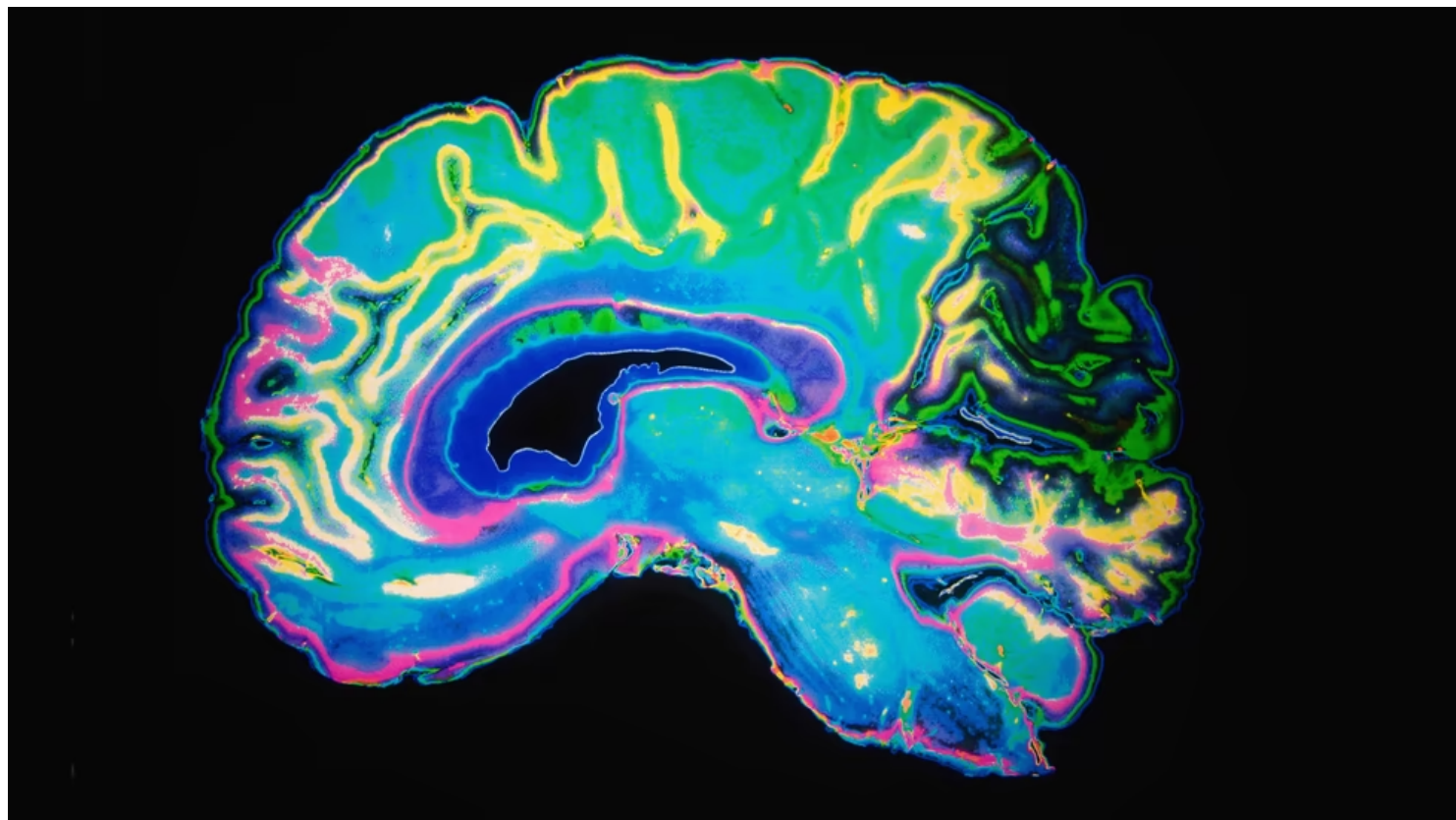A

The Atlantic

TECHNOLOGY

The Difference Between Speaking and Thinking

The human brain could explain why AI programs are so good at writing grammatically superb nonsense.

By Matteo Wong

Alamy

JANUARY 31, 2023, 4:16 PM ET   SHARE ⌄

Language is commonly understood to be the "stuff" of thought. People "talk it out" and "speak their mind," follow "trains of thought" or "streams of consciousness." Some of the pinnacles of human creation—music, geometry, computer programming—are framed as metaphorical languages. The underlying assumption is that the brain processes the world and our experience of it through a progression of words. And this supposed link between language and thinking is a large part of what makes ChatGPT and similar programs so uncanny: The ability of AI to answer any prompt with human-sounding language can suggest that the machine has some sort of intent, even sentience.

But then the program says something completely absurd—that there are 12 letters in *nineteen* or that sailfish are mammals—and the veil drops. Although ChatGPT can generate fluent and sometimes elegant prose, easily passing the Turing-test benchmark that has haunted the field of AI for more than 70 years, it can also seem incredibly dumb, even dangerous. It gets math wrong, fails to give the most basic cooking instructions, and displays shocking biases. In a new paper, cognitive scientists and linguists address this dissonance by separating communication via language from the act of thinking: Capacity for one does not imply the other. At a moment when pundits are fixated on the potential for generative AI to disrupt every aspect of how we live and work, their argument should force a reevaluation of the limits and complexities of artificial and human intelligence alike.

The researchers explain that *words* may not work very well as a synecdoche for *thought*. People, after all, identify themselves on a continuum of visual to verbal thinking; the experience of not being able to put an idea into words is perhaps as human as language itself. Contemporary research on the human brain, too, suggests that "there is a separation between language and thought," says Anna Ivanova, a cognitive

neuroscientist at MIT and one of the study's two lead authors. Brain scans of people using dozens of languages have <u>revealed</u> a particular network of neurons that fires independent of the language being used (including <u>invented tongues</u> such as Na'vi and Dothraki).

That network of neurons is not generally involved in thinking activities including math, music, and coding. In addition, many patients with aphasia—a loss of the ability to comprehend or produce language, as a result of brain damage—remain skilled at arithmetic and other nonlinguistic mental tasks. Combined, these two bodies of evidence suggest that language alone is not the medium of thought; it is more like a messenger. The use of grammar and a lexicon to communicate functions that involve other parts of the brain, such as socializing and logic, is what makes human language special.

RECOMMENDED READING
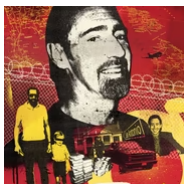
The Escape Artist
JOHN WOLFSON

The Life of a Professional Guinea Pig
CARI ROMM

Why Married Women Are Using Two Last Names on Facebook

ASHLEY FETTERS

Read: Hollywood's love affair with fictional languages

ChatGPT and software like it demonstrate an incredible ability to string words together, but they struggle with other tasks. Ask for a letter explaining to a child that Santa Claus is fake, and it produces a moving message signed by Saint Nick himself. These large language models, also called LLMs, work by predicting the next word in a sentence based on everything before it (*popular belief* follows *contrary to*, for example). But ask ChatGPT to do basic arithmetic and spelling or give advice for frying an egg, and you may receive grammatically superb nonsense: "If you use too much force when flipping the egg, the eggshell can crack and break."

These shortcomings point to a distinction, not dissimilar to one that exists in the human brain, between piecing together words and piecing together ideas—what the authors term *formal* and *functional* linguistic competence, respectively. "Language models are really good at producing fluent, grammatical language," says the University of Texas at Austin linguist Kyle Mahowald, the paper's other lead author. "But that doesn't necessarily mean something which can produce grammatical language is able to do math or logical reasoning, or think, or navigate social contexts."

If the human brain's language network is not responsible for math, music, or programming—that is, for *thinking*—then there's no reason an artificial "neural network" trained on terabytes of text would be good at those things either. "In line with evidence from cognitive neuroscience," the authors write, "LLMs' behavior highlights the difference between being good at language and being good at thought." ChatGPT's ability to get mediocre scores on some business- and law-school exams, then, is more a mirage than a sign of understanding.

Still, hype swirls around the next iteration of language models, which will train on far more words and with far more computing power. OpenAI, the creator of ChatGPT, claims that its programs are approaching a so-called general intelligence that would put the machines on par with humankind. But if the comparison to the human brain holds, then simply making models better at word prediction won't bring them much closer to this goal. In other words, you can dismiss the notion that AI programs such as ChatGPT have a soul or resemble an alien invasion.

Ivanova and Mahowald believe that different training methods are required to spur further advances in AI—for instance, approaches specific to logical or social reasoning rather than word prediction. ChatGPT may have already taken a step in that direction, not just reading massive amounts of text but also incorporating human feedback: Supervisors were able to comment on what constituted good or bad responses. But with few details about ChatGPT's training available, it's unclear just what that human input targeted; the program apparently thinks 1,000 is both greater and less than 1,062. (OpenAI released an update to ChatGPT yesterday that supposedly improves its "mathematical capabilities," but it's still reportedly struggling with basic word problems.)

Read: What happens when AI has read everything?

There are, it should be noted, people who believe that large language models are not as good at language as Ivanova and Mahowald write—that they are basically glorified auto-completes whose flaws scale with their power. "Language is more than just syntax," says Gary Marcus, a cognitive scientist and prominent AI researcher. "In particular, it's also about semantics." It's not just that AI chatbots don't understand math or how to fry eggs—they also, he says, struggle to comprehend how a sentence derives meaning from the structure of its parts.

For instance, imagine three plastic balls in a row: green, blue, blue. Someone asks you to grab "the second blue ball": You understand that they're referring to the last ball in the sequence, but a chatbot might understand the instruction as referring to the second ball, which also happens to be blue. "That a large language model is *good* at language is overstated," Marcus says. But to Ivanova, something like the blue-ball example requires not just compiling words but also conjuring a scene, and as such "is not really about language proper; it's about language *use*."

And no matter how compelling their language use is, there's still a healthy debate over just how much programs such as ChatGPT actually "understand" about the world by simply being fed data from books and Wikipedia entries. "Meaning is not given," says Roxana Girju, a computational linguist at the University of Illinois at Urbana-Champaign. "Meaning is negotiated in our interactions, discussions, not only with other people but also with the world. It's something that we reach at in the process of engaging through language." If that's right, building a truly intelligent machine would require a different way of combining language and thought—not just layering different algorithms but designing a program that might, for instance, learn language and how to navigate social relationships at the same time.

Ivanova and Mahowald are not outright rejecting the view that language epitomizes human intelligence; they're complicating it. Humans are "good" at language precisely because we combine thought with its expression. A computer that both masters the rules of language and can put them to use will necessarily be intelligent—the flip side being that narrowly mimicking human utterances is precisely what is holding machines back. But before we can use our organic brains to better understand silicon ones, we will need both new ideas and new words to understand the significance of language itself.