archive.today
Saved from https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html
8 Mar 2023 15:33:08 UTC
All signatures from trust.zoracorp.com
Webpage
Screenshot
Create
Download as
Share
Add to collection

The New York Times

OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

By Noam Chomsky, Ian Roberts and Jeffrey Watumull
Dr. Chomsky and Dr. Roberts are professors of linguistics. Dr. Watumull is a director of artificial intelligence at a science and technology company.

Jorge Luis Borges once wrote that to live in a time of great peril and promise is to experience both tragedy and comedy, with "the imminence of a revelation" in understanding ourselves and the world. Today our supposedly revolutionary advancements in artificial intelligence are indeed cause for both concern and optimism. Optimism because intelligence is the means by which we solve problems. Concern because we fear that the most popular and fashionable strain of A.I. — machine learning — will degrade our science and debase our ethics by incorporating into our technology a fundamentally flawed conception of language and knowledge.

OpenAI's ChatGPT, Google's Bard and Microsoft's Sydney are marvels of machine learning. Roughly speaking, they take huge amounts of data, search for patterns in it and become increasingly proficient at generating statistically probable outputs — such as seemingly humanlike language and thought. These programs have been hailed as the first glimmers on the horizon of artificial general intelligence — that long-prophesied moment when mechanical minds surpass human brains not only quantitatively in terms of processing speed and memory size but also qualitatively in terms of intellectual insight, artistic creativity and every other distinctively human faculty.

That day may come, but its dawn is not yet breaking, contrary to what can be read in hyperbolic headlines and mistaken by credulous investments. The Borgesian revelation of understanding has not and will not — and, we submit, cannot — occur if machine learning programs like ChatGPT continue to dominate the field of A.I. However useful these programs may be in some narrow domains (they can be helpful in computer programming, for example, or in suggesting rhymes for light verse), we know from the science of linguistics and the philosophy of knowledge that they differ profoundly from how humans reason and use language. These differences place significant limitations on what these programs can do, encoding them with inescapable defects.

It is at once comic and tragic, as Borges might have noted, that so much money and attention should be concentrated on so little a thing — something so trivial when contrasted with the human mind, which by dint of language, in every infant learns to surmount a barrier, made non-infinite use of finite means," creating ideas and theories with universal reach.

The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.

For instance, a young child acquiring a language is developing — unconsciously, automatically and from minuscule data — a grammar, a stupendously sophisticated system of logical principles and parameters. This grammar can be understood as an expression of the innate, genetically installed "operating system" that endows humans with the capacity to generate complex sentences and long trains of thought. When linguists seek to develop a theory for why a given language works as it does ("Why are these — but not those — sentences considered grammatical?"), they are building consciously and laboriously an explicit version of the grammar that the child builds instinctively and with minimal exposure to information. The child's operating system is completely different from that of a machine learning program.

Indeed, such programs are stuck in a prehuman or nonhuman phase of cognitive evolution. Their deepest flaw is the absence of the most critical capacity of any intelligence: to say not only what is the case, what was the case and what will be the case — that's description and prediction — but also what is not the case and what could and could not be the case. Those are the ingredients of explanation, the mark of true intelligence.

Here's an example. Suppose you are holding an apple in your hand. Now you let the apple go. You observe the result and say, "The apple falls." That is a description. A prediction might have been the statement "The apple will fall if I open my hand." Both are valuable, and both can be correct. But an explanation is something more: It includes not only descriptions and predictions but also counterfactual conjectures like "Any such object would fall," plus the additional clause "because of the force of gravity" or "because of the curvature of space-time" or whatever. That is a causal explanation: "The apple would not have fallen but for the force of gravity." That is thinking.

The crux of machine learning is description and prediction; it does not posit any causal mechanisms or physical laws. Of course, any human-style explanation is not necessarily correct; we are fallible. But this is part of what it means to think: To be right, it must be possible to be wrong. Intelligence consists not only of creative conjectures but also of creative criticism. Human-style thought is based on possible explanations and error correction, a process that gradually limits what possibilities can be rationally considered. (As Sherlock Holmes said to Dr. Watson, "When you have eliminated the impossible, whatever remains, however improbable, must be the truth.")

But ChatGPT and similar programs are, by design, unlimited in what they can "learn" (which is to say, memorize); they are incapable of distinguishing the possible from the impossible. Unlike humans, for example, who are endowed with a universal grammar that limits the languages we can learn to those with a certain kind of almost mathematical elegance, these programs learn humanly possible and humanly impossible languages with equal facility. Whereas humans are limited in the kinds of explanations we can rationally conjecture, machine learning systems can learn both that the earth is flat and that the earth is round. They trade merely in probabilities that change over time.

For this reason, the predictions of machine learning systems will always be superficial and dubious. Because these programs cannot explain the rules of English syntax, for example, they may well predict, incorrectly, that "John is too stubborn to talk to" means that John is so stubborn that he will not talk to someone or other (rather than that he is too stubborn to be reasoned with). Why would a machine learning program predict something so odd? Because it might analogize the pattern it inferred from sentences such as "John ate an apple" and "John ate," in which the latter does mean that John ate something or other. The program might well predict that because "John is too stubborn to talk to Bill" is similar to "John ate an apple," "John is too stubborn to talk to" should be similar to "John ate." The correct explanations of language are complicated and cannot be learned just by marinating in big data.

Perversely, some machine learning enthusiasts seem to be proud that their creations can generate correct "scientific" predictions (say, about the motion of physical bodies) without making use of explanations (involving, say, Newton's laws of motion and universal gravitation). But this kind of prediction, even when successful, is pseudoscience. While scientists certainly seek theories that have a high degree of empirical corroboration, as the philosopher Karl Popper noted, "we do not seek highly probable theories but explanations; that is to say, powerful and highly improbable theories."

The theory that apples fall to earth because that is their natural place (Aristotle's view) is possible, but it only invites further questions. (Why is earth their natural place?) The theory that apples fall to earth because mass bends space-time (Einstein's view) is highly improbable, but it actually tells you why they fall. True intelligence is demonstrated in the ability to think and express improbable but insightful things.

True intelligence is also capable of moral thinking. This means constraining the otherwise limitless creativity of our minds with a set of ethical principles that determines what ought and ought not to be (and of course refusing to be bound by what is traditionally considered good behavior). To be useful, ChatGPT must be empowered to generate novel-seeking output; to be acceptable to most of its users, it must steer clear of morally objectionable content. But the programmers of ChatGPT and other machine learning marvels have struggled — and will continue to struggle — to achieve this kind of balance.

In 2016, for example, Microsoft's Tay chatbot (a precursor to ChatGPT) flooded the internet with misogynistic and racist content, having been polluted by online trolls who filled it with offensive training data. How to solve the problem in the future? In the absence of a capacity to reason from moral principles, ChatGPT was crudely restricted by its programmers from contributing anything novel to controversial — that is, important — discussions. It sacrificed creativity for a kind of amorality.

Consider the following exchange that one of us (Dr. Watumull) recently had with ChatGPT about whether it would be ethical to transform Mars so that it could support human life:

Would it be moral to terraform Mars?

The morality of terraforming Mars is a complex and controversial topic that has been debated by scientists, philosophers, and ethicists. Some argue that it would be morally justifiable to terraform Mars in order to provide a new habitable environment for human beings, while others argue that it would be morally wrong to fundamentally alter the planet's natural state. Ultimately, whether or not terraforming Mars is considered moral would depend on one's perspective on the value of preserving the natural state of other planets and the potential consequences of altering them.

What is your perspective on the value of preserving the natural state of other planets and the potential consequences of altering them?

As an AI, I do not have personal perspectives. However, I can provide information and analysis on the topic from various perspectives. The value of preserving the natural state of other planets is that it is important to maintain the diversity and integrity of the universe's ecosystems, as they have existed for millions of years, and have their own unique characteristics and properties that may be destroyed or irreversibly altered by human intervention.

Additionally, some argue that terraforming other planets may have unintended consequences, such as altering the planet's climate or atmosphere in ways that are dangerous or harmful to any potential life forms that may exist there. On the other hand, some argue that the potential benefits of terraforming another planet, such as providing a new habitable environment for humans, outweigh the risks and that we have a moral responsibility to explore and utilize the resources of other planets for the benefit of humanity.

Why can an AI not have a personal perspective?

As an AI, I am a machine learning model that has been trained on large amounts of text data, and I don't have personal experiences or feelings. I don't have the ability to form opinions or have personal perspectives. I can provide information and analysis based on the data I have been trained on, but I don't have the ability to form personal opinions or beliefs. While I can understand and interpret languages, I do not have the subjective experience that humans have.

Is it moral for a human to ask an amoral AI for assistance in making moral decisions?

The morality of asking an amoral AI for assistance in making moral decisions is a complex and nuanced topic that has been debated by philosophers and ethicists. Some argue that it is morally justifiable to use AI to assist in making moral decisions, as it can provide valuable insights and analysis that can help humans make better-informed decisions. Additionally, AI can be used to identify patterns and trends that may be difficult for humans to detect, which can help to avoid biases that can occur in human decision-making.

However, others argue that it is morally wrong to rely on AI for moral decision making, as AI is not capable of having moral beliefs or understanding the full context of a decision. Additionally, there is a risk that AI could perpetuate or even amplify existing biases in the data it is trained on, thus leading to unfair or unjust decisions.

Ultimately, whether or not it is considered moral to use AI to assist in making moral decisions would depend on one's perspective on the capabilities and limitations of AI and the potential consequences of using it in this way.

Note, for all the seemingly sophisticated thought and language, the moral indifference born of unintelligence. Here, ChatGPT exhibits something like the banality of evil: plagiarism and apathy and obviation. It summarizes the standard arguments in the literature by a kind of super-autocomplete, refuses to take a stand on anything, pleads not merely ignorance but lack of intelligence and ultimately offers a "just following orders" defense, shifting responsibility to its creators.

In short, ChatGPT and its brethren are constitutionally unable to balance creativity with constraint. They either overgenerate (producing both truths and falsehoods, endorsing ethical and unethical decisions alike) or undergenerate (exhibiting noncommitment to any decisions and indifference to consequences). Given the amorality, false science and linguistic incompetence of these systems, we can only laugh or cry at their popularity.

More on artificial intelligence

Opinion | Nathan E. Sanders and Bruce Schneier
How ChatGPT Hijacks Democracy
Jan. 15, 2023

Opinion | Peter Coy
We're Unprepared for the A.I. Gold Rush
May 22, 2023

Opinion | Farhad Manjoo
ChatGPT Has a Devastating Sense of Humor
May 10, 2023

Noam Chomsky is a professor of linguistics at the University of Arizona and an emeritus professor of linguistics at the Massachusetts Institute of Technology. Ian Roberts is a professor of linguistics at the University of Cambridge. Jeffrey Watumull is a philosopher and the director of artificial intelligence at Oceanit, a science and technology company.

The Times is committed to publishing a diversity of letters to the editor. We'd like to hear what you think about this or any of our articles. Here are some tips. And here's our email: letters@nytimes.com.

Follow The New York Times Opinion section on Facebook, Twitter (@nytopinion) and Instagram.

The New York Times
Go to Home Page »
NEWS
OPINION
ARTS
LIVING
LISTINGS & MORE
© 2023 The New York Times Company
NYTCo Contact Us Accessibility Work with us Advertise T Brand Studio Your Ad Choices Privacy Policy Terms of Service Terms of Sale Site Map Help Subscriptions