AI is our 'Promethean fire': Using it wisely means knowing its true nature and our own minds

January 4 2024, by Randolph Grace

Credit: Pixabay/CC0 Public Domain

Future historians may well regard 2023 as a landmark in the advent of artificial intelligence (AI). But whether that future will prove utopian, apocalyptic or somewhere in between is anyone's guess.

In February, ChatGPT set the record as the fastest app to reach 100 million users. It was followed by similar "large language" AI models

from Google, Amazon, Meta and other big tech firms, which collectively look poised to transform education, health care and many other knowledge-intensive fields.

However, AI's potential for harm was underscored in May by an ominous statement signed by leading researchers:

"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."

In November, responding to the growing concern about AI risk, 27 nations (including the UK, US, India, China and the European Union) pledged cooperation at an inaugural AI Safety Summit at Bletchley Park in England, to ensure the safe development of AI for the benefit of all.

To achieve this, researchers focus on AI alignment—that is, how to make sure AI models are consistent with human values, preferences and goals. But there's a problem—AI's so-called "dark secret": large-scale models are so complex they are like a black box, impossible for anyone to fully understand.

AI's black box problem

Although the transparency and explainability of AI systems are important research goals, such efforts seem unlikely to keep up with the frenetic pace of innovation.

The black box metaphor explains why people's beliefs about AI are all over the map. Predictions range from utopia to extinction, and many even believe an artificial general intelligence (AGI) will soon achieve sentience.

But this uncertainty compounds the problem. AI alignment should be a

two-way street: we must not only ensure AI models are consistent with human intentions, but also that our beliefs about AI are accurate.

This is because we are remarkably adept at creating futures that accord with those beliefs, even if we are unaware of them.

So-called "expectancy effects," or self-fulfilling prophecies, are well known in psychology. And research has shown that manipulating users' beliefs influences not just how they interact with AI, but how AI adapts to the user.

In other words, how our beliefs (conscious or unconscious) affect AI can potentially increase the likelihood of any outcome, including catastrophic ones.

AI, computation, logic and arithmetic

We need to probe more deeply to understand the basis of AI—like Alice in Wonderland, head down the rabbit hole and see where it takes us.

Firstly, what is AI? It runs on computers, and so is automated computation. From its origin as the "perceptron" —an artificial neuron defined mathematically in 1943 by neurophysiologist Warren McCulloch and logician Walter Pitts—AI has been intertwined with the cognitive sciences, neuroscience and computer science.

This convergence of minds, brains and machines has led to the widely-held belief that, because AI is computation by machine, then natural intelligence (the mind) must be computation by the brain.

But what is computation? In the late 19th century, mathematicians Richard Dedekind and Giuseppe Peano proposed a set of axioms which defined arithmetic in terms of logic, and inspired attempts to ground all

mathematics on a secure formal basis.

Although the logician Kurt Gödel later proved this goal was unachievable, his work was the starting point for mathematician (and code-breaker) Alan Turing. His "Turing machine," an abstract device capable of universal computation, is the foundation of computer science.

Deep structure of perception

So, computation is based on mathematical ideas that trace back to efforts to define arithmetic in logic. But our knowledge of arithmetic exists prior to logic. If we want to understand the basis of AI, we need to go further and ask where arithmetic itself comes from.

My colleagues and I have recently shown that arithmetic is based on the "deep structure" of perception. This structure is like colored glasses that shape our perception in particular ways, so that our experience of the world is ordered and manageable.

Arithmetic consists of a set of elements (numbers) and operations (addition, multiplication) that combine pairs of elements to give another element. We asked: of all possibilities, why are numbers the elements, and addition and multiplication the operations?

We showed by mathematical proof that when the deep structure of perception was assumed to limit the possibilities, arithmetic was the result. In other words, when our mind views the abstract world through the same "colored glasses" that shape our experience of the physical world, it "sees" numbers and arithmetic.

Because arithmetic is the foundation for mathematics, the implication is that mathematics is a reflection of the mind—an expression in symbols of its fundamental nature and creativity.

Although the deep structure of perception is shared with other animals and so a product of evolution, only humans have invented mathematics. It is our most intimate creation—and by enabling the development of AI, perhaps our most consequential.

A Copernican revolution of the mind

Our account of arithmetic's origin is consistent with views of the 18th century philosopher Immanuel Kant. According to him, our knowledge of the world is structured by "pure intuitions" of space and time that exist prior to sense experience—analogous to the colored glasses we can never remove.

Kant claimed his philosophy was a "Copernican revolution of the mind". In the same way ancient astronomers believed the sun revolved around the Earth because they were unaware of the Earth's motion, Kant argued, philosophers who believed all knowledge is derived from sense experience (John Locke and David Hume, for example) overlooked how the mind shapes perception.

Although Kant's views were shaped by the natural sciences of his day, they have proved influential in contemporary psychology.

The recognition that arithmetic is a natural consequence of our perception, and thus biologically based, suggests a similar Kantian shift in our understanding of computation.

Computation is not "outside" or separate from us in an abstract realm of mathematical truth, but inherent in our mind's nature. The mind is more than computation; the brain is not a computer. Rather, computation—the basis for AI—is, like mathematics, a symbolic expression of the mind's nature and creativity.

Promethean fire

What are the implications for AI? Firstly, AI is not a mind and will never become sentient. The idea we can transcend our biological nature and achieve immortality by uploading our minds to the cloud is only fantasy.

Yet if the principles of mind on which AI is based are shared by all humanity (and likely other living creatures as well), it may be possible to transcend the limitations of our individual minds.

Because computation is universal, we are free to simulate and create any outcome we choose in our increasingly connected virtual and physical worlds. In this way, AI is truly our Promethean fire, a gift to humanity stolen from the gods as in Greek mythology.

As a global civilization, we are likely at a turning point. AI will not become sentient and decide to kill us all. But we are very capable of "apocalypsing" ourselves with it—expectation can create reality.

Efforts to ensure AI alignment, safety and security are vitally important, but may not be enough if we lack awareness and collective wisdom. Like Alice, we need to wake up from the dream and recognize the reality and power of our minds.

Provided by The Conversation

Citation: AI is our 'Promethean fire': Using it wisely means knowing its true nature and our own minds (2024, January 4) retrieved 3 May 2024 from https://techxplore.com/news/2024-01-ai-promethean-wisely-true-nature.html