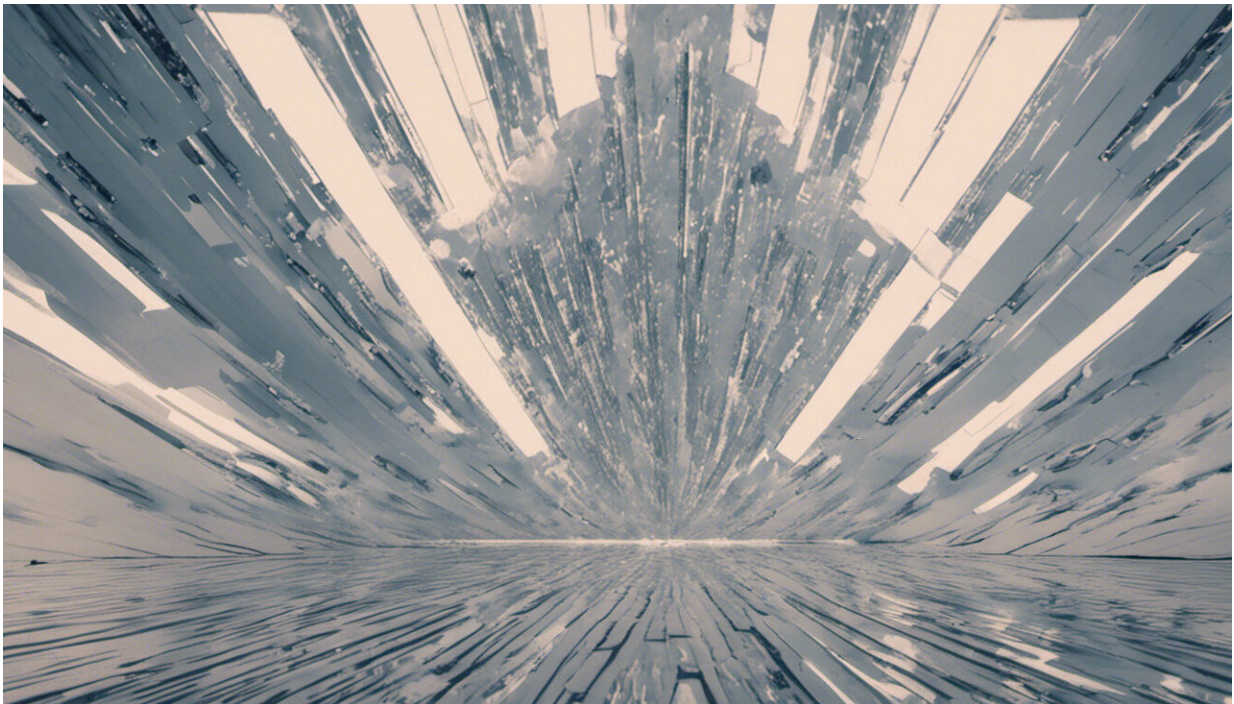# PHYS ORG

We don't want AI that can understand us – we'd only end up arguing

August 21 2017, by Constantine Sandis And Richard Harper

Credit: AI-generated image (disclaimer)

Forget the Turing test. Computing pioneer Alan Turing's most pertinent thoughts on machine intelligence come from a neglected paragraph of the same paper that first proposed his famous test for whether a computer could be considered as smart as a human.

"The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."

Turing's 1950 prediction was not that computers would be able to think in the future. He was arguing that, one day, what we mean when we talk about computers thinking would morph in such a way that it would become a pretty uncontroversial thing to say. We can now see that he was right. Our use of the term has indeed loosened to the point that attributing thought to even the most basic of machines has become common parlance.

Today, advances in technology mean that understanding has become the new thought. And again, the question of whether machines can understand is arguably meaningless. With the development of [artificial intelligence](#) and machine learning, there already exists a solid sense in which robots and artificial assistants such as Microsoft's Cortana and Apple's Siri are said to understand us. The interesting questions are just what this sense is and why it matters what we call it.

Defining understanding

Deciding on how to define a concept is not the same as making a discovery. It's a pragmatic choice (usually) based on empirical observations. We no more discover that machines think or understand than we discover that [Pluto isn't a planet](#).

In the case of artificial intelligence, people often talk of 20th-century science fiction writers such as Isaac Asimov as having had prophetic visions of the future. But they didn't so much anticipate the thought and language of contemporary computing technology as directly influence it.

Asimov's Three Laws of Robotics have been an inspiration to a whole generation of engineers and designers who talk about machines that learn, understand, make decisions, have emotional intelligence, are empathetic and even [doubt themselves](#).

This vision [enchants us](#) into forgetting the other possible ways of thinking about artificial intelligence, gradually eroding the nuance in our definitions. Is this outweighed by what we gain from Asimov's vocabulary? The answer depends on why we might want understanding between humans and machines in the first place. To handle this question we must, naturally, first turn to bees.

As the philosopher of language [Jonathan Bennett writes](#), we can talk about bees having a "language" they use to "understand" each other's "reports" of discoveries of food. And there is a sense in which we can speak – without quote marks even – of bees having thought, language, communication, and understanding and other qualities we usually think of as particularly human. But think what a giant mess the whole process would be if they were also able to question each other's motives, grow jealous, become resentful, and so on like humans.

A similar disaster would occur if our sat-nav devices started bickering with us, like an unhappy couple on holiday, over the best route to our chosen destination. The ability to understand can seriously interfere with performance. A good hoover doesn't need to understand why I need more powerful suction in order for it to switch to turbo mode when I press the appropriate button. Why should a good robot be any different?

Understanding isn't (usually) helpful

One of key things that makes artificial personal assistants such as Amazon's Alexa useful is precisely the fact that our interactions with them could never justify reactive attitudes on either side. This is because

they are not the sort of beings that could care or be cared about. (We may occasionally feel anger towards a machine but it is misplaced.)

We need the assistant's software to have accurate voice-recognition and be as sensitive to the context of our words as possible. But we hardly want it to be capable of understanding – and so also misunderstanding – us in the everyday ways that could produce mutual resentment, blame, gratitude, guilt, indignation, or pride.

Only a masochist would want an artificial PA that could fall out with her, go on strike, or refuse to update its software.

The only exception in which we might conceivably seek such understanding is in the provision of artificial companions for the elderly. As cognitive scientist Maggie Boden warns, it is emotionally dangerous to provide care-bots that cannot actually care but that people could become deeply attached to.

The aim of AI that understands us as well (or as badly) as we understand one another sounds rather grand and important, perhaps the major scientific challenge of the 21st century. But what would be the point of it? We would do better to focus on the other side of the same coin and work towards having a less anthropocentric understanding of AI itself. The better we can comprehend the way AI reasons, the more useful it will be to us.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation