AI Is Not Similar To Human Intelligence. Thinking So Could Be Dangerous

Elizabeth Fernandez Contributor ⓘ

I write about the philosophy and ethics of science and technology.

Follow

Nov 30, 2019, 07:20am EST

Listen to article 5 minutes

🕐 **This article is more than 3 years old.**

It's easy to anthropomorphize artificial intelligence. We imagine befriending Siri, or that our self-driving car has our best interests at heart. When we paint a picture of an advanced AI, we might imagine machines that "learn", similar to the ways a toddler might learn. We imagine them "thinking" or "coming to conclusions" similar to how we do. Even the term "neural networks" - an algorithm modeled after the human brain - brings up images of a brain-like machine, making decisions. However, thinking an artificial intelligence works in the same way as a human brain can be misleading and even dangerous, says a recent paper in Minds and Machines by David Watson of the Oxford Internet Institute and the Alan Touring Institute.
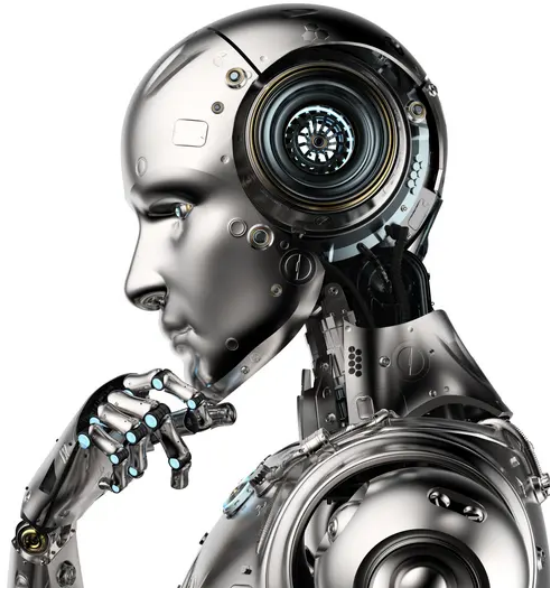
As much as we might like to play them at Go, it's important to recognize the differences between ... [+] GETTY

One of the hottest and most powerful types of machine learning today is the neural network. The name originates from the idea behind neurons and synapses within the brain. In a neural network, input is fed into multiple layers of "neurons". Output is generated by each layer, passing on to be inputted into the next layer. Neural networks that contain a large number of layers are often referred to as deep neural networks (DNNs). Neural nets have evolved to be the grunt behind Google translate, Facebook's facial recognition, and Siri. Beyond that, neural nets can also paint in the style of Van Gogh or even save whales.
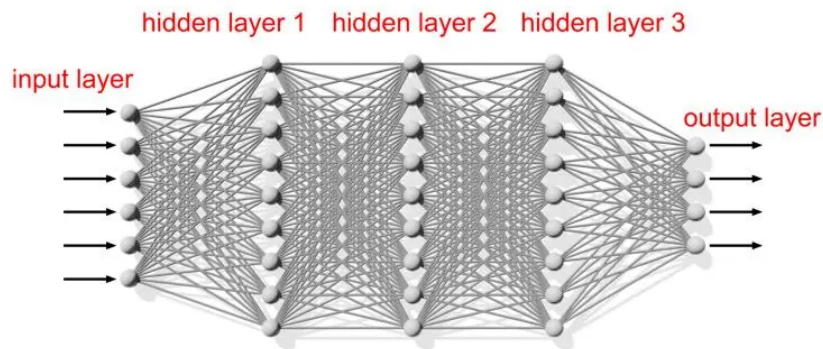
A neural network is composed of layers upon layer of artificial neurons, all of which generate ... [+] GETTY

No doubt, these algorithms are powerful, but to think that they "think" and "learn" in the same way as humans would be incorrect, Watson says. There are many differences, and he outlines three.

The first - DNNs are easy to fool. For example, imagine you have a picture of a banana. A neural network successfully classifies it as a banana. But it's possible to create a generative adversarial network that can fool your DNN. By adding a slight amount of noise or another image besides the banana, your DNN might now think the picture of a banana is a toaster. A human could not be fooled by such a trick. Some argue that this is because DNNs can see things humans can't, but Watson says, "This disconnect between biological and artificial neural networks suggests that the latter lack some crucial component essential to navigating the real world."

MORE FOR YOU

Apple Leaks Detail All-New iPhone 15, iPhone 15 Pro Design Changes

New Apple Leak Reveals iPhone 15 Release Surprise

A Swarm Of Tiny Ukrainian Drones Just Knocked Out One Of Russia's Nine Remaining Terminator Vehicles

Adversarial Patch

Forbes Daily: Get our best stories, exclusive reporting and essential analysis of the day's news in your inbox every weekday.

Secondly, DNNs need an enormous amount of data to learn. An image classification DNN might need to "see" thousands of pictures of zebras to identify a zebra in an image. Give the same test to a toddler, and chances are s/he could identify a zebra, even one that's partially obscured, by only seeing a picture of a zebra a few times. Humans are great "one-shot learners," says Watson. Teaching a neural network, on the other hand, might be very difficult, especially in instances where data is hard to come by.

Thirdly, neural nets are "myopic". They can see the trees, so to speak, but not the forest. For example, a DNN could successfully

label a picture of Kim Kardashian as a woman, an entertainer, and a starlet. However, switching the position of her mouth and one of her eyes actually improved the confidence of the DNN's prediction. The DNN didn't see anything wrong with that image. Obviously, something is wrong here. Another example - a human can say "that cloud looks like a dog", whereas a DNN would say that the cloud *is* a dog.

"It would be a mistake to say that these algorithms recreate human intelligence", Watson says. "Instead, they introduce some new mode of inference that outperforms us in some ways and falls short in others."

Artificial Intelligence is being used increasingly in areas such as finance, clinical medicine, and justice. It can be used to determine who gets credit or who can lease a house or qualify for a loan. When the stakes are high, we want those making decisions - whether they be machines or human, to be correct, trustworthy, and responsible. Are machine learning algorithms and neural nets these things? Perhaps they can be correct. But can they be trustworthy and responsible? While it's hard to judge even if another person is trustworthy or responsible, it may be even harder to judge something that thinks in such radically different ways as humans.

"Algorithms are not 'just like us'... by anthropomorphizing a statistical model, we implicitly grant it a degree of agency that not only overstates its true abilities, but robs us of our own autonomy... It is always humans who choose whether or not to abdicate this authority, to empower some piece of

technology to intervene on our behalf. It would be a mistake to presume that this transfer of authority involves a simultaneous absolution of responsibility. It does not."

Follow me on Twitter. Check out my website.

Elizabeth Fernandez

Dr. Elizabeth Fernandez is the host of SparkDialog Podcasts (sparkdialog.com), which covers the intersection of science and society. She has a PhD in... **Read More**

Editorial Standards   Reprints & Permissions

ADVERTISEMENT