Will AI take over? Quantum theory suggests otherwise

January 7 2020, by Mauro Vallati

Credit: AI-generated image (disclaimer)

Will artificial intelligence one day surpass human thinking? The rapid progress of AI, coupled with our standard fear of machines, has raised concerns that its abilities will one day start to grow uncontrollably, eventually leading it to take over the world and wipe out humanity if it decides we are an obstacle to its goals. This moment is usually referred

to as the "AI singularity."

One argument against the possibility of such a supreme, unstoppable and indefinitely growing intelligence is that it would need, by definition, to be able to accurately predict the future. And quantum theory, one of modern science's key ways of explaining the universe, says that predicting the future may not be possible because the universe is random. But what if we only think predicting the future is impossible because we aren't intelligent enough to know otherwise?

Intelligence is a very complex and abstract concept without an agreed definition. However, there is common agreement about some of the components that make up every sort of known intelligence. One of those is the ability to solve problems, which requires the ability to plan by anticipating the future. To solve a problem, it is pivotal to understand the current conditions, predict how the environment will evolve, and to anticipate the outcome of the actions that will be applied.

Random universe

Recent theories in physics suggest the universe is extremely chaotic and random. Take the example of unstable chemical elements that eventually undergo radioactive decay into another substance. You can estimate how long it will take a certain amount of this element to decay but you can't say for sure when any single atom of it will. Similarly, you can measure the position or momentum of a particle but, for certain reasons related to quantum theory, you cannot know both at the same time with complete accuracy. (This is known as Heisenberg's uncertainty principle.)

Assuming these theories are correct, they suggest that, beyond a certain level of detail, the universe is ultimately unpredictable, chaotic and unstable. This would mean that any sort of growing intelligence would eventually reach a point where it can no longer improve its predictions of

the future and so cannot further increase in intelligence. In other words, there is no risk of a runaway AI, because physical laws of the universe pose some very constraining hard limits. For instance, given the known [limits on weather predictability](link), an AI system will not be able to outsmart humans by exploiting extremely accurate long-term weather forecasts for planning future actions.

It is very comforting to believe that the nature of the universe is, in some sense, preventing an AI escalation. But there is an alternative perspective. What if humans perceive the universe as random and chaotic only because our cognitive and reasoning capabilities are too limited? We are aware of some of the [limits of human understanding](link) but, to [paraphrase Donald Rumsfeld](link), we don't know what we don't know.

Taking this perspective, it may be the case that the universe is instead deterministic, and therefore fully predictable, but in an extremely complex way that we as humans cannot grasp. Albert Einstein argued that [quantum theory](link) was an incomplete description of the universe and that there must be [hidden variables](link) that we don't yet understand but that hold the key to determining future events.

That would turn the table on the possibility of an AI singularity. A super-advanced intelligence could be in the position to reveal these hidden variables and so understand the predictable nature of the universe, unleashing the machine's full potential. It's worth noting that AI approaches are already used [for automatically making discoveries in physics](link).

On a practical level, the singularity doesn't seem that plausible given how limited AI actually still is. Recent breakthroughs in AI have been achieved via what's known as narrow AI, designed to perform a well defined task such as playing chess or driving a car. While narrow AI can

outperform humans in some tasks, there's little to suggest that more [general AI](#) that can emulate humans' ability to respond to many different tasks will be delivered and put humans at risk in the near future.

But we can't rule it out completely. As we still have limited knowledge of the nature of the universe, and of the power of AI, perhaps it is better to play safe. Even without a singularity, [AI will have a dramatic impact on human society](#). We need to work as [hard as possible](#) to ensure that AI is beneficial for humanity, not a threat to it.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation