Singularity: How governments can halt the rise of unfriendly, unstoppable super-AI

August 23 2019, by Wim Naudé,

Credit: lil artsy from Pexels

The invention of an artificial super-intelligence has been a central theme in science fiction since at least the 19th century. From E.M. Forster's short story The Machine Stops (1909) to the recent HBO television series Westworld, writers have tended to portray this possibility as an unmitigated disaster. But this issue is no longer one of fiction. Prominent contemporary scientists and engineers are now also worried that super-AI could one day surpass human intelligence (an event known as the "singularity") and become humanity's "worst mistake."

Current trends suggest we are set to enter an international arms race for such a technology. Whichever high-tech firm or government lab succeeds in inventing the first super-AI will obtain a potentially world-dominating technology. It is a winner-takes-all prize. So for those who want to stop such an event, the question is how to discourage this kind of arms race, or at least incentivize competing teams not to cut corners with AI safety.

A super-AI raises two fundamental challenges for its inventors, as philosopher Nick Bostrom and others have pointed out. One is a control problem, which is how to make sure the super-AI has the same objectives as humanity. Without this, the intelligence could deliberately, accidentally or by neglect destroy humanity—an "AI disaster."

The second is a political problem, which is how to ensure that the benefits of a super-intelligence do not go only to a small elite, causing massive social and wealth inequalities. If a super-AI arms race occurs, it could lead competing groups to ignore these problems in order to develop their technology more quickly. This could lead to a poor-quality or unfriendly super-AI.

One suggested solution is to use public policy to make it harder to enter the race in order to reduce the number of competing groups and improve the capabilities of those who do enter. The fewer who compete, the less pressure there will be to cut corners in order to win. But how can governments lessen the competition in this way?

My colleague Nicola Dimitri and I recently published a paper that tried to answer this question. We first showed that in a typical winner-takes all race, such as the one to build the first super-AI, only the most competitive teams will participate. This is because the probability of actually inventing the super-AI is very small, and entering the race is very expensive because of the large investment in research and development needed.

Indeed, this seems to be the current situation with the development of simpler "narrow" AI. Patent applications for this kind of AI are are dominated by a few firms, and the vast bulk of AI research is done in just three regions (the US, China and Europe). There also seem to be very few, if any, groups presently investing in building a super-AI.

This suggests reducing the number of competing groups isn't the most important priority at the moment. But even with smaller numbers of competitors in the race, the intensity of competition could still lead to the problems mentioned above. So to reduce the intensity of competition between groups striving to build a super-AI and raise their capabilities, governments could turn to public procurement and taxes.

Public procurement refers to all the things governments pay private companies to provide, from software for use in government agencies to contracts to run services. Governments could impose constraints on any super-AI supplier that required them to address the potential problems, and support complementary technologies to enhance human intelligence and integrate it with AI.

But governments could also offer to buy a less-than-best version of super-AI, effectively creating a "second prize" in the arms race and stopping it from being a winner-takes-all competition. With an intermediate prize, which could be for inventing something close to (but not exactly) a super-AI, competing groups will have an incentive to invest and co-operate more, reducing the intensity of competition. A second prize would also reduce the risk of failure and justify more investment, helping to increase the capabilities of the competing teams.

As for taxes, governments could set the tax rate on the group that invents super-AI according to how friendly or unfriendly the AI is. A high enough tax rate would essentially mean the nationalization of the super-AI. This would strongly discourage private firms from cutting corners for fear of losing their product to the state.

Public good not private monopoly

This idea may require better global co-ordination of taxation and regulation of super-AI. But it wouldn't need all governments to be involved. In theory, a single country or region (such as the EU) could carry the costs and effort involved in tackling the problems and ethics of super-AI. But all countries would benefit and super-AI would become a public good rather than an unstoppable private monopoly.

Of course all this depends on super-AI actually being a threat to humanity. And some scientists don't think it will be. We might naturally engineer away the risks of super-AI over time. Some think humans might even merge with AI.

Whatever the case, our planet and its inhabitants will benefit enormously from making sure we get the best from AI, a technology that is still in its infancy. For this, we need a better understanding of what role government can play.

Provided by The Conversation

Citation: Singularity: How governments can halt the rise of unfriendly, unstoppable super-AI (2019, August 23) retrieved 19 April 2024 from https://techxplore.com/news/2019-08-singularity-halt-unfriendly-unstoppable-super-ai.html