

Note:

Thank you Mr. Lautenschlager for allowing me to work over break. This past week has been especially hectic and having 4 days to create a big video project on top of all my other projects that were due this week would not be feasible.

I am really passionate about this topic and have been thinking about this question for a while so I want to do it right.

Very Rough Script – this is just general rn. Will fill in with better English and more explanations for final voiceover, but general message is present.

NARRATOR

A.I. is becoming more powerful and more prevalent with each coming day. Currently, researchers have created programs that can see through walls using the WiFi signal in your house and out maneuver and outthink world champions in complex games (Alpha Go and Dota).

It is undeniable that A.I. will be an incredibly powerful tool, or even too powerful. Prominent figures such as Elon Musk have expressed their concerns quite publicly about the dangers of A.I.

---clip of elon musk fatalistic---

Will we be able to control it? Will we even be able to understand it?

I feel that these are valid questions, however, I want to instead focus on the less explored questions that happens to take place after we have (hopefully) answered the other questions.

Currently, the only A.I. that exists is narrow intelligence commonly referred to as weak. This is specific, single purpose applications designed to solve one type of problem in a singularly specific way. If you take it out of its environment or the data is formatted a different way, it fails.

The goal, and what researchers are working towards, is referred to as strong AI or general AI. In this type of A.I., the computer would be able to approach and adapt to new situations it was not trained specifically for. This is quite similar to how human brains work.

The concern is that such an entity could learn about itself- and selfimprove itself in an infinite loop. The powers this would yield the creator are unimaginable and would be difficult to differentiate it from a God.

--Roll Title Sequence--

Lets start at the basics. How do we know that we will ever create a super A.I?

According to Sam Harris in his Ted Talk [insert ted talk name] it is an inevitability based on these three points.

1. Intelligence is a product of information processing in physical systems.

As evidence by our brains, intelligence can come from matter.

2. We will continue to improve our intelligent machines.

There is always a demand to process data faster and more efficiently.

3. We do not stand on the peak of intelligence or anywhere near it.

We will continue to improve.

By this logic, creating a super AI is inevitable.

So what type of powers would this have?

In his TED talk, Sam Harris says "imagine the ability to perform 2000 years of scientific discovery in an instant".

This is an incredibly terrifying thought. If we took our technology back 100 years, people would be in awe by our smartphones. What technologies lie ahead that are simply undiscovered?

Tom Scott has an incredible video describing a specific situation where an AI goes rogue and features some of the potential an AI could do. In this video, the A.I. creates self-replicating nano robots that are even able to change human thought as a means of protecting itself.

This is why this technology is so concerning- and why I think who controls it is just as big of an issue of the is it even controllable scenario. Providing that we even are able to control a Super AI, who will be in control?

This is a situation in which the person who first achieves super A.I. wins the game of the universe. It's a winner takes everything, even possibly the thoughts of anything else alive as seen from Tom Scott's scenario.

The first person to create a super A.I. will be indistinguishable from a god.

----Game theory situation May or may not put in----

This brings up an interesting point. If the reward for creating such a device is infinite, wouldn't that make any risk worth it?

If you are working on a team creating this A.I. wouldn't you try to sabotage and take control of the A.I? Or, if for some reason you're not a fan of the infinite rewards to any risk you may have to act out of self-preservation. After all, if you are working on this project you are a threat to the person who does end up taking a control, you would most likely be eliminated.

In this scenario it seems that, assuming every person is logical and has knowledge of the end result of this project, that each individual will try to take over. Does this mean that any cooperation is bound to fail in the realm of AI and that AI will solely be produced by an individual who has no one to stop him?

Its all interesting to think about and there are multiple different scenarios that you can play with. I tried googling around to see if anyone has

published anything about such a scenario but came up empty handed. It seems people are still wondering if we can even control an AI.

The closest I could find was related to the pirate game theory question, but in that case the gold in dispute can be divided amongst the pirates. Here, an AI cannot be divided and there is no way to trust other people with infinite power.

Conclusion

It seems we are fast approaching the inevitable creation of superintelligence and this analysis doesn't look too bright. Top researchers were asked when they thought general intelligence would be created and the median answer was 40 years from now.

Humanity may be smart enough to create such a machine, but are we wise enough to wield it? Even if we are collectively wise, all it takes is one bad actor to ruin everything, and I mean everything.

I guess we'll find out soon whether this will be the best thing that has ever happened to humans or the last.