Is Superintelligence Impossible?

Blai Ras

June 15, 2021

Summary The space of possible minds is absolutely vast. All the hundred billion human minds put together are just the tiniest corner of this space of possible minds. For the first time in the history of the planet, the computer has enabled some wholly new kinds of minds to come into existence. This makes us think: is superintelligence possible? Daniel Dennet & David Chalmers discuss this matter in a chill debate moderated by John Brockman.

Essay on "Is Superintelligence Impossible?"

One thing we should do before talking about superintelligence is defining it. I loved watching and listening this talk, but I would have appreciated it a little bit of background introduction about what we (or they) understand by "superintelligence" nowadays.

Nick Bostrom, a philosopher from Oxford University defines superintelligence as:

> "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" — Nick Bostrom

Through all the talk I though they would disagree a lot more than they actually did. Firstly, they both agree that strong AI is possible. They both think that superintelligence will be accomplished by AI systems, but not in a near future.

Secondly, both express their concern about the human role in this long process of AI evolution. In particular, they warn us about the many incentives to take the human out of the loop and to give these AIs the capacity to act on that advice directly and autonomously[1].

[1] They put as example Google Maps: a tool that suggest you the best route but you (as a human) still have to go through it

Nevertheless, Daniel Dennet says that we are interested on building recommendation and advise tools, so we are not removing the "human part" and we are just following an AI guidance. David Chalmers postulates that this approach is not feasible and that we will get to a point when humans inevitably search for the machine perfectionism.

They also start disagreeing when talking about autonomy and consciousness. Dennett comments that we cannot back up a human mind and then reboot it the next day, but nowadays we can plug and un-plug an AI system. Therefore, AI systems are not autonomous because they do not possess free will.

Chalmers then replies that in his opinion, a system is autonomous when it has a wide variety of goals and has the power to achieve them. In order to accomplish them, consciousness is not mandatory, so we might face superintelligent but unconscious AI agents.

Chalmers, when talking about consciousness, also remarks that a system that is not "awake" or "responsive" is just a tool. Dennett

specifies that we might have superintelligent systems that are not conscious in any interesting way, but they will seem conscious in some other ways. For him, it is a matter of if they are able to accomplish their goals recursively and indefinitely.

In my opinion, AI will prove that there is no such thing as "consciousness". One can only be sure about its own existence, and there is no other external way or system to know it. Consequently, if superintelligence is one day achieved, the moral debate about consciousness will still be unresolved and we would have to take another approach in order to understand (and control?) superintelligent systems.

What I also think it would have been interesting is the presence of an individual with a strong, contrary opinion about the feasibility of superintelligent systems. This way, we could have seen how two great minds try to debate with someone who thinks that AI will some day get into a indefinitely winter, and that there cannot be an agent able to exceed the cognitive performance of the human mind.