Intelligence Explosion for Dummies - AVCS MIDSEM Assignment

Sayan Sarkar[1]

[1]Affiliation not available

March 8, 2018

Aim (Logline)

A brief introduction to the idea of the intelligence explosion - a hypothetical time in future when machines will surpass humans in intelligence and ability in virtually every domain, and the possible ramifications of that.

Synopsis/ Outline/ Beat sheet

Machines are becoming smarter every day. Ten years ago no one could think we would have artificial intelligent systems which can drive cars or compose music. More and more jobs, both repetitive manual tasks (like performing surgery or analyzing a legal document), and cognitively-demanding tasks(like diagnosing a disease or teaching a subject) are becoming automated. In plenty of areas already, algorithms perform way better than humans. We call these systems artificial narrow intelligence (ANI) - AI which has a superhuman capability in a specific domain but performs miserably in any other task. But can machines be general problem solvers like humans? Most experts in the field believe yes, and it's just a matter of time that we build human-level machine intelligence (HLMI) or artificial general intelligence (AGI). Now, an AGI doesn't have biological limits like a fixed number of brain cells to analyze information and fixed capacity for memory to store data. It will have access to enormous computational power and all of the human knowledge. So, it will most likely start to self-improve itself to what is called artificial superintelligence (ASI). No human mind will be able to understand or comprehend how this ASI merely works like a chicken doesn't understand how human society works. This scenario is called the intelligence explosion, and no one knows what is going to happen after that. It has been shown that such ASI system can very quickly destroy the whole human civilization to accomplish its goal. It is important to note that such a system need to be conscious of the Hollywood sci-fi films to do so. Interestingly, serious people around the globe have already started a conversation about the topic and are thinking how to best tackle this. We need better public consensus about this topic and maybe something like a second global Manhattan project. If the researchers are correct, the most crucial issue in the history of humankind.

Treatment

Our film is structured with a narrator, who is at times on-screen and often speaks from the voiceover. We use minimal synthwave music occasionally. The narrator (apparently) speaks from her/his studio. The screen is mostly filled with visuals related to the concepts being discussed at the particular moment.

First, the narrator claims that in the twenty-first century we see unprecedented rapid changes in technology. Machines are so smart today that they outperform humans astonishingly in so many tasks. Not only mundane manual jobs like performing surgery or cleaning a room, or analyzing a legal document, but also creative jobs like composing music or writing poetry are being automated. These systems are known as narrow artificial intelligence - an AI system that is domain-specific - it can solve only a specific problem and cannot think outside its domain.

[Visuals - The narrator starts at her/his office. And then we see visuals of examples ANI system - robots performing surgery, people listening to music composed by AI, artwork generated by AIs and so on. An onscreen definition of ANI is animated]

The narrator tells that intelligence is the one factor that made humans the most dominant species on the earth. With intelligence, came the rapid progress of technology, which shaped the human history throughout. [Visuals - on-screen narrator]

The narrator then raises the question if there can be a machine intelligence that is not domain-specific, a system which learns like we humans do and can perform good enough in any general task - not only the assignment it was programmed for. These systems are called AGI - artificial general intelligence or HLMI - human-level machine intelligence. Most experts in the field believe its possible that we will be able to build AGI, and more importantly, by next two decades.

[Visuals - definition of AGI animated on screen, a chart showing the data from recent surveys of most cited AI researchers by the Future of Humanity Institute and so on]

The narrator now describes how human intelligence is limited by our biology. We think with our brains, which, even being one of the most complicated thinking systems we know of, is not that impressive. We have limited number of neurons to process information and a limited amount of memory to store information. But, an AGI system will have access to vast amount of computational power and the whole of human knowledge. Moreover, it processes data with electronic devices, which is much faster than our biochemical information processing. It will be able to learn all we know in a jiffy and then self-improve and create newer knowledge that we don't know yet. At this point, have a system that is smarter than all of humanity combined - an artificial superintelligence (ASI). This rapid development in intelligence is called the intelligence explosion.

The narrator continues, now what happens after the explosion? The Narrator asserts that it is improbable that there will be robot uprising kind of scenario that the Hollywood sci-fi films portray. But, most researchers in the field believe that we won't have any way to know the motives and inner workings of the ASI system, and there will be substantial chance that it will destroy humanity because of several reasons - maybe to maximize computational powers, perhaps to remove the threat to existence.

[Visuals - terminator and other sci-fi film clips]

But on the other hand, if we can build what is known as 'friendly AI' or 'benevolent AI' or 'human compatible AI' it may help us achieve unprecedented human flourishing, solving problems like immortality or intergalactic travel and so on.

[Visuals - utopian sci-fi book covers like Brave New World, an animated definition of friendly AI]

The narrator argues that this is the most important problem humanity has ever seen, and if we are to believe the researcher we don't have much time to solve this issue. Serious people around the globe are already thinking about this problem, and there are already a few government level efforts about this. But, we need better public consensus and collaboration at the global level to tackle this challenge. It's time for a worldwide, second Manhattan Project. This might be humanity's last chance. [visuals - screenshots of those institutes - MIRI, FHI, FLI, OpenAI, open philanthropy, stills of global conferences about the issue, first pages of drafts of White House and WEF regarding the issue]