Artificial Superintelligence: A Solution to All Problems, or Our Biggest Existential Threat? An Evolutionary Look

Student's Name

Institutional Affiliation

Course Code and Name

Professor's Name

Date

Introduction

This year, the term "Artificial Intelligence" (AI) turned 67, as it was first introduced during a workshop at Dartmouth College in 1956. Back then, the prominent scientists and thinkers of that time naively hoped to frame this concept within a month-long seminar and give it a finite goal of achieving human-level intelligence (Dartmouth Workshop, 1956). However, the discussion they've started is not over yet – some are pushing AI research forward, while others are voting to put it on halt.

We believe that the leap from weak to super AI appears intrinsically counterintuitive to human observers, despite the inevitability of this transition. Consequently, our present deliberations should not primarily concern whether to rigorously program AI or altogether halt its progression. Instead, we should focus on how humanity can adapt to the world in which it is no longer a dominant species.

The Debate

Since 1956, several types of AI have been introduced into the debate, including narrow AI (such as in today's smartphones), general AI (human-level intelligence), and super AI.

It is this last type of AI, the artificial superintelligence, that is now causing a divide among the best thinkers, splitting them into believers of the benevolent nature of the future superintelligent AI, and strong proponents of its intrinsically evil intentions promising catastrophic consequences for humanity.

The Biggest Believers in Peaceful Coexistence

- **Ray Kurzweil.** Ray is a famous futurologist, inventor, and one of Google's leading AI developers. He is known for his optimistic predictions of our future with artificial intelligence. In his work "The Singularity is Near" he gives plenty of arguments for why we should not be afraid to push forward with the development of super AI (Kurzweil, 2005).

- **Elon Musk.** While Musk is more cautious of the peaceful coexistence between the human race and superintelligent machines, and he advocates for the development of strict regulations and preventive policies, he is also the one who is actively investing in AI research and is propelling this whole area forward.
- **Max Tegmark**. A widely cited AI researcher, Max, in his bestseller "Life 3.0: Being Human in the Age of AI", suggests to take on the human values approach in creating the ever-powerful AI (Tegmark, 2017).

The Loudest Heralds of the Upcoming Disaster

- **Stephen Hawking.** Who else can speak more convincingly of the prospects of AI than an Einstein of the present day with a robotic (powered by AI) voice? Hawking explicitly warned us of the dangers of uncontrolled AI development in his latest works (Hawking, 2015).
- **Nick Bostrom.** A philosopher, and widely published writer Nick Bostrom, in his latest book "Superintelligence: Paths, Dangers, Strategies", thoroughly discussed the potential of super AI to break through any digital and physical cages we can make to tame its power (Bostrom, 2014).
- **Eliezer Yudkowsky.** Yudkowsky is a lead researcher at the Machine Intelligence Research Institute. In his recent publication in Times Magazine, he claims that any sufficiently advanced AI, shortly upon its creation, could instantly kill all life on Earth. Not because it hates life and us humans, but simply because we are made of atoms it can use for other purposes (Yudkowsky, 2023).

This is a highly simplified and condensed model of the ongoing debate on this topic. In reality, plenty of scientists are afraid to speak with skepticism about AI publicly, while they certainly express their concerns in private discussions. Furthermore, some of the above-mentioned opponents, in the past, had different views, and they’ve changed their opinions based on the breakthroughs in the field of AI. This is a natural behavior for someone dealing with something they cannot fully comprehend.

With the advent of Chat GPT-4 and similar advanced software, on the one hand, and noticeable progress in the development of quantum computers, on the other, humanity doesn't seem to have much time left for theoretical discussions.

Parallels with Biological Life on Earth

Just like the current narrow AI has its programmers, the biological life on Earth as we know it also initially had a programmer – the process of evolution. Scientists claim that the first living cell on Earth appeared 4 billion years ago (Dawkins, 1976). Back then, and up until the present day, cells had no other intentions but to replicate, i.e., produce copies of themselves.

The Aim Does NOT Always Justify the Means

Now think about this – How did this elementary and harmless replication program lead to the appearance of such complex and malevolent life forms as predators? Was it conceivable that in the process of achieving higher efficiency at replication, cells would evolve to form such complex organs as eyes and a brain, to form consciousness, and ultimately, to become self-aware?

The blind and slow process of biological evolution has led to the advent of the industrial revolution, advanced machinery, and, finally, the human level of intelligence, that is now giving birth to super AI and is scared of its potential to destroy life on Earth as we know it.

The Morale

The moral is simple – there is little to no sense in trying to program the to-be artificial superintelligence away from causing harm to its creators. As even the most humanistic initial goals, the super AI can quickly supplement with unpredictable by-goals and supporting means that we would call evil and immoral.

Who can be certain of what will become of this planet, its resources, and all the living things in the process of super AI achieving its programmed goals? Just like the harmless program of cell replication has been enhanced by the evolution to produce claws and teeth in a lion, with which it kills herbivorous creatures, or the F-16 jets and nuclear weapons used by one group of people to kill and dominate over the other.

On Exponential Growth and Knockout to Our Intuition

The progress that has led us to the advent of AI has been going on exponentially. According to the famous Moore's Law, the number of transistors on an integrated circuit doubles every 18 months (Moore, 1965).

Before the transistors, however, the exponential growth has been “exploiting” vacuum tubes, and relays, and even before them there was (and always is) the biological exponential growth. It took evolution billions of years to come from a single cell to a living organism, then hundreds of thousands of years for complex creatures like mammals to emerge, and only a few hundred years from man’s invention of simple work tools to sophisticated technology like smartphones.

However, nothing was ever able to stop or at least pause that growth – not a natural catastrophe, deadly disease, epidemic, economic recession, or even a war. Its exponential pace is stable and inevitable.

A Thought Experiment

To imagine how counterintuitive the technological exponential growth is, we suggest the following thought experiment. Imagine you are trying to cover a chessboard with rice grains. A chessboard has only 64 squares and your goal is to simulate an exponential growth by placing a single grain on the first square, two on the second, four on the third, and so on. How many rice grains do you think you need to cover the whole board in the same manner? A cup of rice? Or perhaps a bucket? No, the volume of rice you’ll need will far surpass the world’s annual rice production and can be compared to the size of Mount Everest.

Even if the current AI seems too weak (narrow) to be afraid of, your human, linear intuition is a bad advisor on the AI’s power in merely 10 years from now.

Conclusion

The debate over artificial superintelligence’s nature and impact on humanity rages on. While believers like Ray Kurzweil, Elon Musk, and Max Tegmark emphasize potential coexistence benefits and cooperation gains between humans and AI, voices of caution like Stephen Hawking, Nick Bostrom, and Eliezer Yudkowsky warn of dire consequences.

As we navigate this debate, it's essential to consider the parallels with biological evolution and the counterintuitive nature of exponential growth. The trajectory of AI's development calls for not just cautious regulation, but also thoughtful adaptation to a world where artificial superintelligence will surpass human supremacy.

Bibliography

1. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.
2. Dartmouth Workshop. (1956). Dartmouth College. Retrieved from https://www.dartmouth.edu/artificial-intelligence-conference/
3. Dawkins, R. (1976). The Selfish Gene. Oxford University Press.
4. Hawking, S. (2015). Brief Answers to the Big Questions. Bantam.
5. Kurzweil, R. (2005). The Singularity is Near: When Humans Transcend Biology. Penguin.
6. Moore, G. E. (1965). Cramming more components onto integrated circuits. Electronics, 38(8), 114-117.
7. Tegmark, M. (2017). Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf.
8. Yudkowsky, E. (2023). Pausing AI Developments Isn't Enough. We Need to Shut it All Down. Times Magazine. Retrieved from https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/