Can Artificial Intelligence Defeat Humanity? — A Postprint of the XWorld Conference Roundtable

Authors:

Date: 2023-10-08T00:00:00+00:00

Abstract

When humans make decisions, they must collect diverse data, perform calculations, and identify patterns within that data. Artificial intelligence proves far superior to humans in making decisions through such processes. However, this is not a matter of opposition or determining a victor in some confrontation. What is more probable is a fusion that transforms humans into hybrid entities—partially organic and partially inorganic. Thus, I propose reframing the question: rather than asking whether A…

Full Text

Preamble

When humans make decisions, they must collect diverse data, perform calculations, and identify patterns within that data. Artificial intelligence proves far superior to humans in making decisions through such processes. However, this is not a matter of opposition or determining a victor in some confrontation. What is more probable is a fusion that transforms humans into hybrid entities—partially organic and partially inorganic. Thus, I propose reframing the question: rather than asking whether AI will wage war against humanity to determine a winner, we should view it as a "wedding"—a marriage between humans and artificial intelligence. And I believe this marriage will inevitably occur.

Introduction

On July 6, 2017, the "XWorld Conference"—co-hosted by CITIC Publishing Group and Baifendian Group—was successfully convened. The roundtable featured Yuval Harari, the "young genius" and author of *Sapiens* and *Homo Deus*;

Su Meng, Chairman and CEO of Baifendian Group; Wei Kunlin, Professor at Peking University's School of Psychological and Cognitive Sciences; Yin Qi, founder and CEO of Megvii (Face++); and Ai Cheng, founder of iAsk Media. Together, they engaged in a stimulating intellectual exchange with both live and online audiences on three pivotal topics: "Can artificial intelligence defeat humanity?", "Will artificial intelligence cause severe class differentiation?", and "How should we view the future?" Below are the participants' principal viewpoints.

Roundtable Discussion

Q1: Will Artificial Intelligence Defeat Humanity?

Wei Kunlin: As a researcher of the human brain and cognition, I increasingly concur with Mr. Harari's perspective. We can decompose human capabilities into distinct modules—perception, decision-making, and creativity—and observe AI surpassing humans across all of them. In terms of memory, AI has already exceeded human capacity. Perception—how we interpret the world—is another domain where AI is catching up, as evidenced by autonomous driving systems that must rapidly perceive pedestrian movements and make correct judgments. Finally, creativity largely involves the random recombination of existing information followed by trial-and-error testing for utility and feasibility. AI is advancing here as well. In Go, for instance, AlphaGo frequently makes moves that humans could never conceive—this represents genuine creativity.

Yuval Harari: This question fundamentally concerns power and decision-making. While these currently remain in human hands, they will increasingly be transferred to computers and artificial intelligence.

Su Meng: I believe AI has already defeated humans in numerous domains beyond just Go. First, AI is stackable. Although human storage and computational capacities are considerable, our brains cannot connect as efficiently as computers. A million Go masters combined do not achieve greater Go proficiency because their intelligence cannot integrate effectively, whereas machines can. Second, AI's learning capability far exceeds that of humans. The amount we learn from primary school through university is minuscule compared to what machines can rapidly absorb, read, and comprehend. Consequently, I am convinced that humans will ultimately be unable to defeat artificial intelligence.

Yin Qi: Surprisingly, I am the only participant who believes AI will not ultimately defeat humanity. First, we must consider the timeline—discussing ultimate outcomes while ignoring temporal context often leads to inaccurate conclusions. Although AI is developing rapidly, it has not yet truly touched upon its most essential core, leaving us a window of 50 to 100 years for technological evolution. Within this extended timeframe, the fusion and co-evolution of AI and humans may occur before AI surpasses human intelligence—AI+HI might reach that critical threshold first. If such a prolonged cycle exists, fusion will precede defeat.

Second, technology essentially provides only possibilities—it may point to options A, B, or C for the future, but factors beyond technology may ultimately determine humanity's fate. Moreover, if AI could reach human-level intelligence in fifty years, Mr. Harari's current global warnings about AI defeating humanity might, like a chaotic system, alter AI's trajectory due to human prediction. Based on these considerations, I believe AI and humanity will more likely follow a path of co-evolution.

Q2: Will AI and Human-Machine Fusion Exacerbate Class Division?

Yin Qi: Throughout history, all technological and productivity advancements have resulted in job displacement—for example, carriage drivers became obsolete. Productivity improvements themselves never cause class division; rather, class division stems from production relations or the influences behind productivity. AI itself is fundamentally neutral. I object to the question of whether AI will divide humanity, though I agree that AI has significant social effects.

Su Meng: Mr. Harari, you mentioned the intensification of social stratification. Have you observed data monopolies by giants in certain fields?

Yuval Harari: Indeed. Timeframes differ dramatically across perspectives. For entrepreneurs, five years represents a very distant future—they typically focus on the coming months. For a historian, however, five years is exceedingly brief; even fifty or five hundred years seems short. Different people operate on different temporal scales. I am not discussing the emergence of a useless class in five years, but rather the possibility in fifty years. One might argue this remains distant and that we have time, but that is precisely what makes our situation dangerous—this seemingly alarming scenario demands our serious attention and action today.

Wei Kunlin: Today's discussion touches on timescale issues. We consistently underestimate long-term developments while overestimating short-term progress—this represents a cognitive bias. If we adopt a longer perspective, many actions become necessary now, such as educational reforms to equip future generations with greater adaptability and flexibility in both professional and personal life.

Ai Cheng: Mr. Harari, you firmly believe division will occur. How might we achieve greater fairness? Your books do not elaborate on this—could you share your thoughts? If you were to become one of the 1% "superhumans," what treatment would you hope for the 99% "useless" people?

Yuval Harari: We possess the power to create a favorable environment for the 99% because we have technology capable of providing them with food, water, education, healthcare, and more. I am not so pessimistic as to believe that a future "useless class" will lack basic sustenance. The resources exist; the problem lies in motivation—whether the 1% will have the incentive to create a good life for the 99%. This requires global coordination. No single country can

solve this problem alone. Differentiation will occur between nations—some will benefit enormously from AI and become more prosperous, while others will be left far behind.

The challenge, therefore, is how to manage and coordinate this globally. This is particularly critical for developing countries whose economies are built largely on cheap labor. How will their children prepare within the education system for AI's arrival? What happens if we no longer need their labor in thirty years? Can their children adapt? Addressing this requires international cooperation and coordination. Our concern is that the current state of the world is moving further away from such coordination.

Q3: Optimistic or Pessimistic About Humanity's Future?

Yuval Harari: Human history offers both optimistic and pessimistic reasons. The last technological revolution, for instance, introduced nuclear weapons that threatened human survival. In the 1950s and 1960s, many feared nuclear weapons would lead to humanity's destruction. Yet this did not occur—humanity bravely met the challenge, avoided nuclear war, and global violence levels declined significantly. Today, we witness fewer violent events than at any period in human history; more people die from overeating than from human violence. This history provides many optimistic reasons—we can also resolve the AI challenge.

Conversely, we have negative cases of human failure, such as climate change and global warming, also resulting from technological progress. Our response to this problem has thus far been woefully inadequate, and our climate and ecosystem are deteriorating rapidly. At least to date, humanity has not done enough to prevent this dire situation—this is our pessimism.

I believe humans are capable of making both wise and profoundly unwise decisions, and we cannot predict which in advance. This is why I am both optimistic and pessimistic.

Note: Figure translations are in progress. See original paper for figures.