

Research on the Ethical Dilemma and Outlet of Strong Artificial Intelligence [†]

Lu Wang

Department of Philosophy, School of Humanities and Social Sciences, College of Marxism, Xi'an International Studies University, Xi'an 710128, China; wl00213680@163.com

[†] Presented at the 5th International Conference of Philosophy of Information, IS4SI Summit 2021, Online, 12–19 September 2021.

Abstract: Today, the rapid development of intelligent technology, especially Strong AI, which has become a staple in fiction, from various films and television shows to literary works, has become a contentious topic within popular culture, which approaches the topic with generalized and prejudicial panic. In view of this, this paper analyzes and compares Strong AI and human intelligence through the lens of multi-view fusion, pointing out that Strong AI may appear presently, and be present for a long time to come. Therefore, people should not be excessively anxious and afraid, as this could lead to the blind rejection of high technology and even social instability. What we should do now is make use of this anxiety, be vigilant and sensitive in maintaining ethical practices, consider the future development and trends of Strong AI, and think about how to eliminate or avoid the anxiety it may prompt in human beings in the process.

Keywords: strong AI; human; anxiety; three laws



Citation: Wang, L. Research on the Ethical Dilemma and Outlet of Strong Artificial Intelligence. *Proceedings* **2022**, *81*, 100. <https://doi.org/10.3390/proceedings2022081100>

Academic Editors: Yixin Zhong and Kun Wu

Published: 12 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. How Can Strong AI Induce Human Anxiety?

1.1. Anxiety about the Tendency of "Anti Anthropocentrism"

Humanity's fear of Strong AI is rooted in the fact that human beings in traditional political societies have a misguided fear that their own subjective status could disappear as a result of its occurrence. This anxiety, of humanity's "central digestion", is ultimately rooted in their fear of the "other". It reflects the inequalities found, both within and without, traditional societies. However, what human beings must accept is that this phenomenon is gradually approaching reality, just as we have accepted other paradigm shifts regarding the way in which things are done, for example, that we can no longer enforce excessive demands upon nature. With this in mind, human beings must take the initiative to "decentralize".

1.2. Human Anxiety about Self-Identity

Regarding the concept of the "name reality" relationship, because Strong AI simulates human intelligence and may eventually develop independent consciousness, human beings could themselves be categorized as Strong AI. Taking a robot with independent consciousness as an example, the robot itself does not make any demands for a "name" of its own, which is given to and by human beings, subject to their own perspectives, life experiences, and intrinsic preferences. The body is simply named "robot". This kind of machine, one that has a "human" title, will undoubtedly impact humanity's subjective identity. These lines, between organic and binary, have already been blurred; people living in the era of big data are even called "information people", amalgamations of acquired personal private data based upon huge data caches. The traditional "single" definition of a person is beginning to collapse. Hence, the anxiety over human identity. As such, from this point of view, the essence of human anxiety about Strong AI relates to external performance and of the uncertainty of self-identity and subjectivity.

1.3. The One Sidedness of Human Cognition and Prediction and the Immaturity of Current Artificial Intelligence Technology

In an experiment, a research team working for “Facebook” asked two AI robots to have a conversation. As the dialogue continued, the researchers were surprised to find that the two AI robots had developed a specialized language that was difficult for human beings to understand. In order to avoid the dialogue between the two robots proceeding beyond the point of human comprehension, they terminated the experiment [1]. The general population also recognizes the one-sided cognition and predictive limitations of artificial intelligence. Frequent accidents still occur, such as driverless cars (such as the 2016 Tesla that failed to identify a white tractor when viewed against a bright sky) causing accidents, or medical casualties, caused by the wrong command output of a medical machine still. This shows that artificial intelligence technology is still immature at its present stage, which is another cause for mistrust and anxiety in humans.

1.4. The Continued Fear of the Prophecy That Had Been Realized

Ray Kurzweil, a man hailed as “the future of artificial intelligence”, stated that: when the development of strong AI reaches a certain stage, in other words, when it breaks through the singularity, it will have abilities that exceed human beings, and thus, a super artificial intelligence will appear. It is predicted that such a singularity will occur in 2045. Since the prediction that robots would eventually defeat human players at Go in 1998, AlphaGO became the first computer program to defeat a professional human player in 2016. In response to this, a number of people feared that this heralded the realization of the “singularity”, that it was only a matter of time. From a post-humanist perspective, this is another highlight of “Cartesian anxiety” on the issue of subjectivity, which demonstrates the difficulties faced by physicalism in explaining human cognitive problems [2]. For instance, one should consider the impact that the emergence of Strong AI will have upon jobs that originally belonged to humans.

Additionally, the excessive dependence of human beings on Strong AI will rock the economic world to its foundations and lead to a decline of the humanistic spirit. It is reasonable to have fears and worries. The essential point demonstrated here is that human beings continuously worry about “human nature”, as they have done throughout the ages. As such, why should they be any less uncertain about another being’s “own nature”.

2. Is the Fear and Rejection of Strong AI Necessary?

To tackle this subject, we must explore the differences between artificial intelligence and human intelligence as follows:

Firstly, we should understand how we define “intelligence”. Only with such a foundation, can we clearly understand the delineation between “artificial” intelligence and “human” intelligence. The word “intelligence” first appeared in Xunzi (*Zhengming chapter*): “The ability of human beings to know things is called cognitive ability, the conformity of cognition with external things is called wisdom, some abilities of human beings are called instinct, and the conformity of instinct with external things is called talent.” To summarize, intelligence is not just cognition, wisdom, and consciousness, but also the product of emotion, ability, and action. This complex and fascinating comprehensive phenomenon has always been regarded as exclusive to human beings.

To create artificial intelligence is to create machines, or other artificial systems, that simulate human intelligence [3]. It requires comprehensive knowledge of computer science and technology, neurophysiology, psychology, philosophy, and other interdisciplinary fields. It mainly uses a computer, or computers, to simulate a human’s cognition, but as an agent without a body. The term “Strong AI” refers to a machine that is not only a tool but also has the consciousness of thinking being, comparable to that of a human being in its cognitive abilities. Examples of what this may look like can be found in many films, as previously mentioned. Films, such as “Jiwa” in the sci-fi movie “Ji Ji”, and “Anyi” and “Li Yao” in “Hello, Anyi”. However, the words “simulation” and “tool” posit that a Strong AI

is dependent on human-manufactured hardware, non-autonomous, and belongs to human beings. This is the basic position and starting point of understanding Strong AI.

Looking at the differences in cognitive strategy, “human” and “agent” take completely different paths. John Searle proposed that: “there is no need to imagine an information processing level or representation level between the neurobiological level and the psychological level of the brain. At the same time, there is no functional equivalence between the human brain and the computer” [4] (p. 39). Similarly, Colin Allen, among others, also believes that there is an essential difference between top-down learning and bottom-up learning in the cognitive learning ability of robots and humans [1].

As previously mentioned, the construction of “knowledge” and “wisdom” in the service of human intelligence requires one to understand the meaning of “knowledge”. “Knowledge” is the result of “practice” for the sake of “seeking”. It not only has the function of consciousness in the mind, but also requires practical experience with regards to the relationship between the “heavenly officials”, that is, sensory organs, and objective objects. “Knowledge of behavior” is finally implemented through “knowledge of virtue”, the unity of knowledge and practice. Following this path of logic, from “can know” to “already know”, the key transformation point, “action”, lies in the subject’s inner experience and emotional resonance (not excluding being forced to act). It is worth noting that the experience of each subject’s intuitive contact is not the same and that the emotion will therefore be different. The body is not just physical. Humans are incomparably and discriminatorily superior to Strong AI, an issue cannot be resolved technologically, legally, or ethically.

As far as the current technological means are concerned, existing research is still in its early days, and there are still many technical problems that cannot be solved when simulating a human brain. So, is still a long way to go before we meet the necessary conditions for AI output. Zhai Zhenming pointed out that, “any artificial system that does not take the material with consciousness function as the matrix, unless there is sufficient reason to conclude that the mechanism or content of consciousness is introduced and retained in the process of its artificial generation, we must think that the system has no consciousness like the original matrix material, no matter how close its behavior seems to the behavior of the conscious subject.”, And put forward a “and this possibility is only possible in quantum mechanics” [5] point of view.

3. Dispelling and Avoiding Anxiety towards Strong AI via the “Three Laws”

3.1. Positioning

Although Strong AI is “strong”, it is still the product of human intelligence. In order to understand and make use of its mysterious and charming consciousness, human beings turn it into an object-oriented machine carrier with the help of high-tech means, so as to achieve the purpose of serving themselves and society. It is clearly defined as an “intelligent carrier” to serve human society and create a harmonious living environment with human beings. It is not to be seen as a “replica”, raising humanitarian ethical issues, nor should it be considered a rash attempt driven by curiosity and interest. At the same time, human beings have prioritized humanity when designing robots; that a robot should “avoid actions or situations that could cause it to come to harm itself” is Asimov’s third law, not the first. Therefore, the construction of robots should be considered altruistic. Of course, human beings are also amiable towards altruistic behavior, but this kind of complete altruism is unlikely [6]. Therefore, the design and application of Strong AI must enforce Asimov’s “three laws of robots” for the lifetime of the construct. On the premise that “a robot shall not harm a human”, we must cultivate positive interactions between artificial intelligence and human beings. We must abide by the core of these applied ethics, so as to realize the more economic and efficient operation of human society.

3.2. Establishing Correct Values and Moral Norms Is the Internal Standard of Technological Ethical Behavior

We should ensure the construct's virtue through the guidance of responsibility ethics, strengthen the moral consciousness of developers, sellers, users, strong artificial agents, and establish people-oriented, fair, harmonious, and sustainable values which embody both individual and public responsibility. Friedrich Rapp, a German philosopher of technology, proposed that "all technological actions reflect certain values. If we want to change our existing actions, we must rethink and change our attitude" [7] (p. 148). This, essentially, is to change people's "conscious intention and subconscious motivation", which requires a transformation from moral heteronomy to self-discipline. This requires that developers adhere to the "cautious", "simple", and "restrained" attitude to programming and control. The group should maintain a sense of "worry" and "anxiety" to build a "human being unified in every way", and it must build a "human nationality", the purpose of which is to ensure that they "listen to the supranational cause of human beings on the premise of respecting nationality" [8] (p. 25). Here, a strong artificial intelligence can adopt the moral norm model of "top-down" and "bottom-up" [9].

3.3. Through the Improvement of Spiritual Civilization Construction and Cultural Confidence, We Can Enhance Self Identity

At present, peoples' identities have changed quietly when compared with the other historical transitions. People need to face this and change their way of thinking from "keeping the achievement". They must reconstruct their own identities through reflection and become conscious of their traditional historical culture. They must strengthen their cultural identity, reinforce their internal spiritual structure, cultivate the nation's cultural psychology, and uphold national belief and cohesion. We should adhere to the socialist core-value system: people-oriented, scientifically rich, and promoting the cultivation of an ideal personality, which combines the best of traditional and modern social developments.

3.4. Formulate Relevant Laws and Regulations, and Constantly Improve upon Them with the Development of Artificial Intelligence Technology

We should implement "three-step legislation" before, during, and after the event. The so-called "prior legislation" supplies the legal means meant to regulate the development of Strong AI, ensuring that its development occurs within the scope of human control, and to predict and avoid possible ethical issues. Secondly, "legislation in the matter" defines the basis and standards of the responsibility division, which legally enforces the responsibilities and obligations of R&D developers, users, and Strong AI agents, so as to ensure the effect of "there are laws to follow, law enforcement must follow". Responsibility cannot be shirked onto others if and when a problem arises. Finally, "legislation after the event" compensates for the errors found in aspects of the first two steps. For this process, we should establish a third-party testing system to evaluate, certify, and otherwise restrict control over technical and moral forbidden zones. For the sake of fairness, these laws and regulations should not conflict with each other.

3.5. Human Beings Need to Have "Self Restriction" While Improving Their Cognitive Ability and Technical Means

Firstly, AI is the crystallization of human intelligence. The development of Strong AI is closely related to human cognition. Therefore, we should not reject or accept Strong AI. We should treat it with a rational, dialectical, developmental, tolerant, but cautious attitude, which requires the confirmation of a human level of cognition and the unity of regularity and purposiveness, that is, all technologies that human beings create, and will create in the future, should be consistent with the laws of nature [7]. This requires greater human cognitive literacy. Secondly, current Strong AI technology choices should reserve space for future humans, including their opportunities for repentance. It should combine the traditional value of maintaining social connections with that of maintaining the personal unity of the subject, ensuring the future space, which is the way of "the purpose

remains unchanged, the means can be changed" [9]. We should establish a detailed and perfect "generation development" data file, that continuously tracks the technical services for each strong artificial agent. Human beings have the tendency and ability to adapt to cultural changes. Just as we adapt to technology, we make and use tools for a better collective future.

3.6. Creating a Transparent, Cooperative, Shared, and Harmonious World Environment

In order to design reliable Strong AI and give it more power, we need to implement the aforementioned preventative measures, and also reasonably guide it with the help of social cooperation. To strengthen individual, national, and international exchanges, and cooperation, and to establish common ethics and legal norms, we require the free and transparent dissemination of science and technology, rather than its being owned by individual people or groups alone. In order to prevent the hegemonic countries from using these huge data resources and advanced technologies, such as Strong AI to serve their ambition of dominating the world, we should establish a mutual restriction mechanism between the "data oligarch" companies and the government. At the same time, we should restrict some commercial companies that only serve their own interests while ignoring inconvenient information on potential dangers. In particular, we should regulate the ethics of "data oligarchs".

4. Conclusions

This analysis posits that we should uphold a sense of ethics and morality that is invisibly sliding between "yes" and "should be able to", rather than stubbornly adhering to the conservative notion that any deviation from the traditional sciences should be regarded as "irrational", "fictional", and "fantastical", that is, anything that has not yet been achieved is regarded as "irrational". It's not ethical to worry about. However, the fact is that invention itself is usually regarded as "good", and its relevant ethical supervision is to ensure that the expected goodwill not be derailed due to reckless greed for private profit. Therefore, expert thinking focuses more on predicting "impossible" technologies, preventing "unrealistic" ethical speculation via irrational public anxiety, advocating for a human society with unlimited creativity. Only when the threat to human dignity becomes reality, can we reflect on the ethical consequences of this far-reaching science and technology.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Guo, J.; Wang, W. How to avoid the ethical problems caused by artificial intelligence. *People's Forum* **2018**, *11*, 56–57.
2. Ye, N.; Wang, H. Utopia and practicality—Understanding the relationship between the object and the ego in the era of artificial intelligence. *Philos. Sci. Technol.* **2017**, *12*, 113–119.
3. Chen, J. Integration of technology and Ethics: On the human culture of artificial intelligence technology. *Academia* **2017**, *9*, 102–114.
4. Searle, J. *Heart, Brain and Science*; Yang, Y., Translator; Shanghai Translation Publishing House: Shanghai, China, 1991; p. 39.
5. Zhai, Z.; Peng, X. How strong artificial intelligence will change the world—Technological leap of artificial intelligence and technological prospect of Applied Ethics. *Acad. Front.* **2016**, *4*, 22–33.
6. Cui, Z.; Wang, H. The philosophical basis of human-computer interaction in artificial intelligence research: A discussion from Mello Ponty's direct perception of social otherness. *J. Xi'an Jiaotong Univ. Soc. Sci. Ed.* **2019**, *1*, 130–137.
7. Friedrich, R. *Introduction to Philosophy of Technology*; Liu, W., Translator; Liaoning Science Press: Shenyang, China, 1986; p. 148.
8. Hans, J. *Technology, Medicine and Ethics: Practice of the Principle of Responsibility*; Zhang, R., Translator; Shanghai Translation Press: Shanghai, China, 2008; p. 25.
9. Mo, H. Thinking on ethical issues of strong artificial intelligence and weak artificial intelligence. *Soc. Sci.* **2018**, *1*, 14–24.