ISSN 2706-6827 Vol. 5, Issue 12: 130-134, DOI: 10.25236/IJFS.2023.051221

Reflections on the Moral Status of Robots—The Establishment of Moral Status Standards for Robots in the Era of Strong Artificial Intelligence

Yingdong Shen[1,*], Ziyi Gong[2]

[1]School of Politics and Public Administration, Qufu Normal University, Rizhao, Shandong, China
[2]School of Politics and Public Administration, Qufu Normal University, Rizhao, Shandong, China
[*]Corresponding author

Abstract: *The rapid development of artificial intelligence, such as robots, has not only improved people's lives and accelerated the development of society, but also triggered ethical debates about whether robots have moral status. This paper discusses the importance of the internal "spirit" (or "soul"), and discusses the transformation of the moral status of robots from the perspective of human-machine relationship. It is concluded that the current robots still do not have moral status, but at the same time, it also raises the question of whether the era of strong artificial intelligence should give robots moral status. Based on this, the theory of ethical behaviourism is introduced in order to establish a comprehensive new criterion for judging moral status in response to the debate about the moral status of robots with full autonomous consciousness in the future.*

Keywords: *Robot, Moral Status, Moral Consciousness, Ethical Behaviourism*

1. Introduction

With the advent of the information age, the rapid development of science, the gradual increase in the degree of autonomy of robots, and the increasing resemblance of robots to human beings in certain aspects of their behaviour, appearance and way of thinking have triggered controversies about robots in terms of technology and ethics, in particular, the discussion on whether robots should have moral status is becoming more and more heated. At present, the debate on the moral status of robots mainly focuses on the fact that robots do not have complete autonomous consciousness and are a kind of tool-like artificial manufacturing machinery, so it is difficult to assign moral status to robots. However, with the imagination of future robots, after entering the era of strong artificial intelligence, robots undergoing deep learning can acquire autonomous consciousness in the true sense of the word, and due to the special artificial nature of robots, it is not possible to rely solely on the awakening of autonomous consciousness alone to confer a moral status on them. John Danaher proposed and defended a theory of "ethical behaviorism", which no longer pays attention to the internal "spirit" and makes judgments only through external behavior, to identify and support the moral subject status of robots. This point of view for the debate on the moral status of robots to provide a new idea, but it also aroused a lot of opposition and questioning the voice of the existence of a certain degree of one-sidedness. Therefore, this paper attempts to explore a new and more comprehensive evaluation standard to discuss the moral status of robots in the era of strong artificial intelligence, in order to promote the orderly development of robots.

2. Manuscript Preparation

2.1. Prerequisites for the moral status of robots

2.1.1. The importance of the inner "spirit"

Moral status refers to the position, qualification, or role of a moral subject in the moral community.[1] For the debate on whether robots should be given moral status, we should first judge whether robots can become moral subjects. Kant's understanding of the moral subject points out that “the moral subject should have the rational freedom of unity of law and will. The moral subject should have: autonomy, that is, the ability of the moral subject to grasp the moral law and practice the moral law according to the will; self-determination, that is, the ability to choose their own behavior of good and evil; self-discipline, that is, moral subjects can legislate for themselves and follow their own conscience. From a moral philosophical point of view, the idea that robots can act as fully autonomous moral agents today is quite unconvincing."[2] The premise of the ability to choose to do good and evil is to have moral perception and moral consciousness, and to distinguish what is good and what is evil. "From a macro-historical perspective, consciousness evolved in response to environmental challenges to individual and group survival, which is based on physical sensibility; from a micro-individual perspective, the formation of consciousness, self-consciousness, and ethical awareness is closely related to a person's body-based sensibility, life history, and reflective capacity."[3] As Hume thought, "Morality is felt rather than judged." Johnson has proposed that a "moral subject" must meet five conditions, that is, it has an internal state composed of desires, beliefs or other intentional states that can cause action, an external event that can prompt the subject to act, an internal state that is the cause of an external event, an external action that has an external effect, and an external effect that has a moral bearer.[4] We can realize that the inner "spirit" such as autonomy and will is an indispensable prerequisite for judging whether an entity has moral status.

Humans and animals are living beings, which are composed of complex combination of organic matter, while machines are non-living beings made of complex combinations of inorganic matter, and robots are completely different from the biological nature of humans or animals. Robotics, as a machine consisting of a composite of inorganic materials, is defined internationally as "a programmable and multifunctional manipulator; or a specialized system with computer-changeable and programmable actions for performing different tasks."[5] Robots are seen as a system, a tool, an artifact, not biological, not in the category of human or animal. "Intelligent robots can be broadly divided into two parts, namely the brain and the body. The core of its brain part is the algorithm program, while the body part is composed of a series of physical existence."[6] From the beginning of the design to the application and operation stage, the robot is inseparable from the external intervention of programmers and operators. It relies on algorithm logic and program design to simulate and practice human behavior, thinking and emotion, and concretes moral and complex abstract ethical principles into data and programs built into the machine so as to obtain autonomy, make moral judgments, and make behaviors that conform to moral principles. While the application of human-like sensory systems in robots allows them to feel pain using some sort of algorithm within their bodies, it is essentially just a computational procedure written in advance by a programmer for a finite number of possibilities, and robots are not capable of sensing in more complex situations. It is difficult for robots to understand the moral principles behind moral behavior and moral choices, for example, the "three laws of robotics" proposed by Isaac Asimov in his book *Runaround* stipulates that "robots shall not harm humans or cause humans to be harmed by inaction." When robots comply with this law, they can make behaviors and choices to avoid harm to humans, but because obedience to the built-in program does not really understand the moral significance of respecting and protecting life behind not harming humans, most of the current robots cannot reach the standard of full autonomy.[7]

2.1.2. The Moral Status of the Relational Approach

Kukolberg proposed a relational method for the judgment of moral status, which is: "Moral status is accompanied by the relations between entities, and to define moral status independently of these relations is itself contrary to morality. The application of this method to robots will mean that in order to determine their moral status, it is necessary to know its relationship with other machines and humans, to know the scene, history and location of the machine, and also to know how it is embedded and constituted naturally, materially, socially, and culturally."[8] Starting from the perspective of the relational approach, robots exist and are rapidly developing in the fields of politics, military, business, medicine, education, public life, etc., but regardless of the field of life, the ultimate purpose of the current application of robots is to serve human beings and ensure the stable and sustainable development of society. The original intention of the robot is "slave, servant". At the beginning of its design, it was used as a tool with higher work efficiency, improving people's production efficiency and quality of life, and always serving people. "Bryson's new instrumentalism regards the robot as a slave, and believes that the responsibility of the robot belongs to its producer or operator, and the robot itself cannot be held responsible. From the perspective of dialectics, the essence of the human-machine relationship is the relationship between man and object, a master-slave relationship, while the essence of the human-machine relationship is the relationship between man and man, a social relationship of equality. Replacing human relations with human-machine relations is a kind of scientific and technological alienation, and even more so a kind of alienation of social relations."[9] "Whether in the present or in the future, the essence of intelligent machines is still 'genus humanity', a 'high-level' objectified product of human self-generation and self-realization, which should ultimately be used as a means to serve human beings as an end in itself."[10] Therefore, in the development of artificial intelligence such as robots, the purpose of robots is always designed to serve human beings, and the human-machine relationship is always a kind of master-slave relationship, which does not become a moral subject and cannot gain moral status. From the development of the weak artificial intelligence era to the present day, the moral relationship between robots and humans has been deformed from the beginning, and this unequal human-machine relationship cannot give robots a moral status from a teleological point of view. As robots develop towards the era of strong artificial intelligence, we are beginning to realize the problem of the moral status of robots. On the one hand, this understanding comes from the rapid development of robot intelligence, and powerful intelligent systems are likely to realize the self-awakening of moral subjects. On the other hand, it comes from the re-understanding of human-machine interaction. Robots have been very different from non-biological concepts in the past, and now we call them silicon-based organisms. In short, the human-machine relationship has begun to shift, and thinking about the question of the moral status of strong artificial intelligence is quietly taking place.

It is widely believed in academia that artificial intelligence can be classified into three main categories based on its degree of autonomy: weak artificial intelligence, strong artificial intelligence and super artificial intelligence. “Judging from the definition criteria of moral subject, only rational subjects with rational thinking and decision-making ability and independent responsibility can have moral subject status, and now the development of artificial intelligence is only in the stage of weak artificial intelligence, human beings cannot design a machine with human emotions, independent consciousness, freedom of choice and independent responsibility. Therefore, the machine does not have the status of moral subject.”[11] Although most of the current robots are in the era of weak artificial intelligence robots that do not have autonomous consciousness and do not have complete intelligence, robots have shown amazing wisdom in the actual application process. In 2017, at a Future Investment Initiative conference in Riyadh, the “sweet-looking” and “articulate” Sofia was granted Saudi Arabian citizenship.[12] With the imagination of future robots, the degree of autonomy of robots has increased to the era of strong artificial intelligence or even super artificial intelligence, where robots can learn by themselves and thus possess true intelligence and autonomy, and are able to independently think and make choices, and where inner “spirit” is no longer a limitation for granting robots a moral status, can robots be granted a moral status at that time? Obviously, due to the artificial nature of the robot, it is difficult to judge them solely on the basis of the classical theory of moral status, and it should be improved on its basis to establish a new standard for the establishment of moral status.

2.2. Ethical Behaviorism's Recognition of the Moral Status of Robots

In his article *Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism*, John Danaher expresses his recognition and support for the moral status of robots, proposing and defending the idea of “ethical behaviourism”. John Danaher argues that the argument for the moral status of robots lies in the fact that if a robot is roughly equivalent in performance to another entity, and the other entity is widely considered to have an important moral status, then it is right and appropriate to give that robot the same status.[13] He named it “ethical behaviorism”, and this idea comes from behaviorism, which used to be ontological behaviorism, that the inner “spirit” is something that can be ontologically reduced to the behavioral level, and the inner “spirit” is an abbreviation for a set of actions. The methodological behaviorism believes that it is difficult for us to observe the internal “spiritual” state that people think constitutes intelligence, and therefore can only observe recorded and measurable external data, including behavior and brain phenomena, so it is difficult to define robot morality from the inside. In comparison, John Danaher identified ethical behaviorism as the application of methodological behaviorism in the field of ethics, which is a Turing test in ethics. In the understanding and knowledge of external things, it respects the limits of human beings and agrees that perception provides a basis for the metaphysical attributes of human beings. However, in practical ethics, behavior is the only insight into the metaphysical basis of moral status.

To a certain extent, this kind of behaviorism can make the standard of moral status more observable and predictive, and make the moral nature of metaphysics become a more realistic value judgment. The traditional ontological view naturally rejects this idea, and the most crucial question is whether robots can be defined as moral creatures. Ontological scholars generally believe that moral status first needs to be established on biology, and all non-living things do not have moral attributes at present, so they hold a negative attitude towards the moral status of robots. However, after the rapid development of science and technology, robots are no longer non-living things in the traditional sense. Although they are still inorganic in structure, robots can reach the standard of living things from the perspective of perception and are approaching to living things in behavior and life. Ultimately, relying exclusively on species to confer moral status is not in itself moral enough. At the same time, the programmed inputs make the robot's behaviour controllable, and people can intervene in the robot's ethical behaviour by predicting and analyzing the robot's feedback. This can greatly increase the sense of trust in human-machine interaction, and it is also the moral red line of human-machine interaction. But in the era of strong artificial intelligence, this red line is not a simple set of ethical procedures, because such algorithmic morality is controversial. The red line should be the spontaneous moral bottom line of robot moral consciousness and the ethical principle consistent with human moral behavior. When robots are able to spontaneously generate this moral bottom line, then from the perspective of ethical and moral principles, they have the prerequisites for moral autonomy, which is an important manifestation of the awakening of the moral consciousness of robots.

Ethical behaviorism provides a new way of thinking on the issue of whether robots have moral status. Traditional ethics believes that the reason why the moral status of robots is controversial lies in the lack of internal spirit of robots, and that morality cannot be programmed. It is true that if ethics were simply compiled into a program code, the ethical system of human society would collapse in an instant. John Danaher shifted his vision from the internal to the external, focusing on behavior, making the moral status more realistic and practical, which is a breakthrough in the past metaphysical moral and ethical concepts. The ethical behaviorism proposed by John Danaher is a new cognition of the development of robots to a new stage, which gives robots a moral status from the external behavior equivalence. However, the premise of establishing such moral status is not to consider the internal, that is, the inner soul or consciousness of robots, as a judgement of the moral status, the complete abandonment of the individual's internal still remains singularly one-sided, and therefore, we propose a more comprehensive judgement criterion, which regards the robot's behaviors and its own awakening of its own moral consciousness as a mark of the possession of a moral status.

2.3. Establishment of A New Standard for the Moral Status of Robots

As an entity, the presentation of moral status should be internal and external together, requiring moral autonomy and self-consciousness internally and moral behaviour and performance externally. No one-way view is sufficient to support the establishment of a new moral status for robots. If the robot already has moral consciousness, but does not behave morally, we consider it to have no moral status as well. This is because ethics, as a social code of conduct for dealing with relationships, should be placed in the social environment. "All social life is essentially practical. All the mysterious things that lead theory to mysticism can be reasonably solved in human practice and in the understanding of this practice."[14] Without social relationship communication, it does not matter whether you have moral status, because moral status is established in practice. Having only moral consciousness without moral behaviour, the moral status of robots will remain at the metaphysical level, and only the combination of moral consciousness and moral behaviour can reflect the rationality of the moral status of robots. In other words, the moral behavior of robots is based on the performance of moral consciousness and can bear moral responsibility, and in the human-machine relationship robots have to use moral ethics as the principle of interaction, and it is in this kind of social environment that we can judge the moral status of robots more clearly.

At this point, we can clearly understand how the moral status of robots in the era of strong artificial intelligence is established. Internally, we focus on the prerequisites for moral status, which requires both the self-awakening of moral consciousness and the establishment of equal social relations. This is the reason why the moral status of robots has always been difficult to establish in the past. In terms of external behavior, robots should make ethical and moral performance behaviors and practice moral and ethical principles in social relations. This creates an ethical system in which moral consciousness and moral behaviour are mutually reinforcing.

It is worth mentioning that the moral status of robots established in this paper should not be ontological, we believe that moral status should be reflected in moral relations. Throughout the past moral theory, pay more attention to the absolute status of the inner "spirit". However, in the face of the future era of strong artificial intelligence, do we need to change our attitude and rethink the importance of the inner "spirit" for moral status? In this regard, we deeply recognize John Danaher's shift in perspective from the internal to the external, but at the same time we recognize that the internal "spirit" plays an important role in moral status. Therefore, we critically combine classical moral theory with ethical behaviorism, and try to respond to the moral status problems faced by robots in the era of strong artificial intelligence by establishing a comprehensive moral status judgment standard, while also creating a harmonious atmosphere for human-machine interaction.

3. Conclusions

With the continuous development of artificial intelligence, robots have gradually penetrated into every field of life, and whether we should give them moral status has become an urgent problem to be solved. The reflection on the ontology of robots at the present stage shows that robots are tools or machines designed by humans to simulate human beings in appearance and some functions, serve humans in purpose, have no moral perception and moral consciousness, and cannot make moral choices. In essence, it is an algorithm input into the body, which can execute moral instructions, but it cannot understand the moral significance behind it. It is not desirable to give the current robot moral status. If with the development of science and technology, robots enter the era of strong artificial intelligence or super artificial intelligence, and can achieve true self-awareness through deep learning, perhaps robots at that time can gain moral status. However, the existing criteria for judging moral status tend to focus on the internal “spirit” of living beings, such as humans or animals, or to judge them on the basis of measurable behaviors, which are theories based on species with living characteristics. Therefore, in the face of such special silicon-based creatures as robots, a comprehensive new type of judgement standard should be established for robots in the era of strong artificial intelligence, and discuss the internal consciousness awakening and moral behavior of robots, so as to solve the ethical dilemma of the moral status of robots and promote the stable development of human-machine relationship.

References

[1] Tongjin Yang. On the Moral Bearer Status of Robots and Its Normative Implications. Philosophical Analysis, (2019) 10 (06), 14-33+191.

[2] Hongyu Liu. Research on Robot Moral Decision-making Approach and Model. Studies in Dialectics of Nature, (2021) 37 (09), 28-34+82.

[3] Le Yu. Can Intelligent Robots have Rights? Journal of Huazhong University of Science and Technology (Social Science Edition), (2020) 34 (05), 17-24.

[4] Deborah G. Johnson. Computer Systems:Moral Entities but not Moral Agents. Ethics and Information Technology, (2006) 8 (4).

[5] Donghao Wang. Exploration of Robot Ethical Issues. Future and Development, (2013) 36 (05), 18-21.

[6] Yibin Dai. How is Artificial Intelligence Ethics Possible?—Based on the Perspective of Moral Subject and Moral Receiver. Studies in Ethics, (2020) 05, 96-102.

[7] Bekey G A. Autonomous Robots: From Biological Inspiration to Implementation and Control. MIT Press, (2005).

[8] Lingyin Su. A Review of Current Foreign Research on Robot Ethics. Journal of Xinjiang Normal University (Philosophy and Social Science), (2019) 40 (01), 105-122.

[9] Tieyan Yin, Jinping Dai. Philosophical Implications, Practical Issues and Governance Approaches of Artificial Intelligence Ethics. Journal of Kunming University of Science and Technology(Social Science), (2021) 21 (06),28-38.

[10] Qiongqiong Li, Zhen Li. The Dialectic of “Human-Machine Relationship” in the Age of Intelligence - Contemporary Echoes of Marx's Thought on Man and Machine. Studies on Mao Zedong and Deng Xiaoping Theories, (2021) 01, 71-79+108.

[11] Kunru Yan. Do Artificial Intelligence Machines have a Moral Subject Status? Studies in Dialectics of Nature, (2019) 35 (05), 47-51.

[12] Li Zhang, Peng Chen. A Critique of the Theory of Robot “Personality” and the Legal Regulation of Artificial Intelligence Objects. Academics in China, (2018) 12, 53-75.

[13] Danaher John. Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. Science and engineering ethics, (2019) 26 (4).

[14] Karl Marx, Frederick Engels. Karl Marx, Frederick Engels: collected works (Volume 1). People's Publishing House, (2009), 501.