ISSN 2706-6827 Vol. 5, Issue 10: 55-61, DOI: 10.25236/IJFS.2023.051011

The Moral Status of Artificial Agents

He Huanhuan[1,a,*]

[1]School of Politics and Public Administration, Qufu Normal University, Rizhao, Shandong, 276827, China
[a]huohuo707@163.com
[*]Corresponding author

Abstract: *The fact that decisions made by artificial agents in the real world are often ethically tinged has led many authorities to advocate the development of "artificial moral agents." From the essence of moral agents, as long as the AI can deliberate, has autonomy, and can be responsible for its actions, then the AI can have moral status. It is theoretically and technically reasonable to "embed" morality into AI, and it is necessary, on the one hand, because of the inevitability of its development, on the other hand, because the functions presented by AI development become more and more complex, making it more and more difficult to predict its behavior, and at the same time, to prevent the infringement of human rights. AI agents inevitably need to make autonomous moral decisions, so AI agents may need to be given the same moral status as humans.*

Keywords: *Artificial agent; Moral framework; Moral status*

1. Introduction

The concept of artificial intelligence emerged from research into the nature of human intelligence, with Norbert Wiener, through his famous work on feedback loops, arguing that intelligent human behavior stems from feedback loop mechanisms and that machines have the potential to simulate these mechanisms [1]. Wiener's research was influential in the early development of artificial intelligence, leading engineers to try to push the boundaries of current technology and blur the lines between human intelligence and artificial intelligence. Ray Kurzweil made a very bold prediction about the development of computer intelligence, "I set the date of the singularity to the time of a very deep and divisive transition - 2045." Abiotic intelligence this year will be a billion times greater than all human intelligence today" [2]. In Ray Kurzweil's view, regarding AI agent technology, "the real issue involved here is strong artificial intelligence" [2], which is an inevitable trend in the development of artificial intelligence. As AI becomes more complex and autonomous, yet exists outside our moral sphere due to its limited capabilities, we have to address questions about the moral status of AI agents. Assuming that there will be artificial agents that can pass the Turing test in the future, what will it take for these artificial agents to become members of the human moral community?

2. Analysis of the moral nature of artificial agents

According to many philosophers, subjectivity requires entities to be able to perform certain actions, that is, subjectivity is a dual causal relationship between an entity and an action, and the entity is considered to be the source of the action. So, what is behavior? In the usual sense, behavior refers to something that happens to us or that we do. We can call these two events and actions respectively. The main difference between the two is that events are purposeless, while actions are purposeful. For example, we can argue that breathing and writing are both actions, both dependent on entities, but that breathing is an unconscious act, while writing is purposeful. So we can think of breathing as an event and writing as an action [3]. The subject is necessarily related to action, and an entity can only be called a subject if it acts. Therefore, intentionality is a necessary condition for an entity to become a subject, and highly evolved AI agents can also "exhibit human-like subjectivity characteristics that are different from ordinary artificial technical entities", and "AI agents can not only become a new category of agents, Moreover, the emergence of artificial intelligence agents also expands our understanding of the inner state of the agents to a certain extent" [4]. Because the mechanism inside the AI works very similar to the consciousness generated by the human brain. However, intentional entities are not the same as rational agents. For example, although animals are capable of certain intentional actions, animals clearly cannot be considered rational agents.

So what is a rational agent? This requires an explanation of the reason. Since the Renaissance, there has been controversy about the explanation of rationality, but there are generally two kinds of explanations for rationality: One is the non-naturalistic explanation, which regards rationality as a kind of ability. The most representative explanation is Kant's explanation of rationality. Kant believes that rationality is an intrinsic human ability and moral law comes from reason and is constructed by human reason. The second explanation is a naturalistic one, which argues that reason depends on our physiology or can be reduced to some physiological phenomenon, such as the causal or functional effects of the neural structure or abstract neural patterns of the brain. According to the first interpretation, it is assumed that rationality is an intrinsic ability, so AI can only be considered rational if it has autonomy because the first interpretation suggests that rationality involves a kind of autonomy. According to the second interpretation, rationality can be reduced to a certain function or structure, and if the AI has the same function and structure as humans, then the AI can be considered rational [5]. Naturalistic explanations of reason are more reasonable than non-naturalistic explanations of reason. Reason cannot be completely reduced to a certain structure or function, because the reductionist explanation of reason will face the problem of infinite regression. For example, reason is interpreted as a structural function, and the explanation of this structural function is necessarily dependent on other structures and functions. This way of explaining reason is more cumbersome than the way of explaining reason as an internal ability, so the non-naturalistic explanation is more reasonable.

In addition to the characteristic of autonomy, another remarkable characteristic of rationality is its thoughtfulness. Deliberation refers to the subject thinking about the reasons behind its actions, whereas the animal's intentional behavior does not think about the reasons behind things, so the animal cannot act as a rational subject. In contrast, part of human behavior is characterized by deliberation, and therefore man is regarded as a rational subject.

What is the connection between rational subjects and moral subjects? A moral subject refers to "a moral actor who is self-conscious, capable of moral cognition, reasoning and self-judgment, moral choice, moral behavior, and moral responsibility" [6]. It can be seen from this that only humans with the characteristics of rational agents can become moral agents, that is, as long as artificial agents become rational agents, artificial agents may become moral agents. There are only two technical problems that need to be solved here: one is whether artificial intelligence can have the ability to deliberate. The other is whether the AI can act autonomously. In other words, from the perspective of research and development, we need to make artificial agents have the ability to think and act rationally like people, so "we can regard the research of artificial agents as the design and development process of rational agents". "Rational design includes the input of learning function, moral function, behavioral function, and language function, and the output is anthropomorphic intelligence" [7]. However, in addition to the characteristics of rational agents, namely the capacity for moral reasoning, moral agents seem to have other characteristics, namely that moral agents are associated with responsibility. This means that when a moral agent commits a moral act, he is always praised or blamed under certain moral standards. In particular, it should be noted that moral agents are related to autonomy, that is, moral agents must make moral judgments and conduct moral actions by themselves. Because "value judgment mainly discusses the behavioral standards of the value subject, it indicates that the subject that can be included in the scope of discussion must be a free subject with independent consciousness" [8]. To develop a theory of artificial moral agents, three technical questions need to be considered: First, can artificial agents be capable of deliberation? Second, can AI be autonomous? Finally, can AI agents be held accountable for their actions? If the above problems can be solved, then it is clear that AI agents can become a kind of moral agent [9]. Therefore, for AI agents to become a member of the human moral community, the above three conditions must be met at the same time to give them moral status.

3. Why do artificial agents have moral status

There are two main objections to giving artificial agents moral status: one is that it is technically and theoretically impossible to give artificial agents moral status; Another view is that giving AI agents moral status is unnecessary. By summarizing and refuting some important objections to the construction of artificial agents, the moral status of artificial agents is explored more deeply.

3.1. The refutation of giving artificial agents a moral status

Robert Sparrow argues that giving artificial agents moral status is not feasible. On the one hand, morality is so closely related to our own complex emotions and neural networks that technologists may underestimate the technical difficulty of solving this difficult problem, so it seems technically impossible to give artificial agents moral status [10]. Patrick Chisan Hew similarly argues that it is not technically feasible to give artificial agents moral status. He believes that if an ethical AI is to be built, its rules of behavior and the mechanisms to provide those rules cannot be entirely provided by external humans, and such technology has little prospect at present [11].

On the other hand, in Robert's opinion, we lack an accurate definition of "ethics", while there are many contradictory ethical theories, it is difficult for us to determine which ethical theory should be given to the artificial agent, and it is difficult to determine whether the artificial agent is "ethics". Therefore, in theory, we can not give "ethics" to artificial agents. Similarly, Ariela Tubert argues that moral rules are difficult to program due to their theoretical complexity because moral rules need to be applied in the right context and to make the right trade-offs in a given context, but there is a long-standing disagreement about which rules to apply and when. That is, we have not yet developed a proper set of rules to explain our ethical views [12]. Therefore, due to the complexity of the moral rules themselves, we cannot design and program an executable moral program for an AI agent. In Ariella's view, the learning of artificial agents is to learn information input by humans or the information searched on the Internet, and the information input by humans is not always ethical, and artificial agents cannot accurately distinguish unethical information, which makes artificial agents not always able to learn correct moral rules. In other words, AI may also learn unethical behaviors and make unethical judgments. For example, Microsoft's chatbot Tay posts racist information on Twitter, and Google Translate translates in a sexist way, all of which are caused by incorrect information input or collected by AI during moral learning.

Even if they could solve the above problems, Amitai Etzioni and Oren Etzioni argue that AI agents do not have to be moral agents, or that it is unnecessary to give AI agents moral status. Because they believe that AI agents do not need to make moral decisions. On the one hand, in most cases, AI agents only need to submit to legal arrangements. For example, in the case of self-driving cars, we do not need to let self-driving cars themselves morally consider whether they should be regulated, but simply obey the arrangements of traffic laws and drive according to traffic laws. So what the designers of self-driving cars need to do is not set some kind of ethical framework for self-driving cars, but make self-driving cars obey the law like all other cars. On the other hand, they believe that even in the face of some moral dilemmas, the artificial agent does not need to make moral decisions at all, that is, the designer does not need to input the moral framework for the artificial agent, but only needs to input the user's moral preferences into the program of the artificial agent, and the artificial agent can analyze such preferences. The preferences of users are summarized to assist users in making ethical decisions. Here, the AI only serves as a tool to help the user make ethical decisions in emergencies or moral dilemmas when the user cannot make quick ethical decisions, and the AI does not have to make any ethical decisions made by itself. Therefore, they argue that in most cases, AI agents can keep order through legal means, and in other cases, let them comply with users' moral preferences [13]. Ariella also believes that it is unnecessary to give artificial agents moral status because giving artificial agents moral status means giving artificial agents autonomy, and we do not need artificial agents to have autonomy, because artificial agents have autonomy means that artificial agents have strong unpredictable behavior. This unpredictability can be harmful to humans, and we should minimize the risk of harm caused by AI agents. Therefore, Ariella believes that we only need to control the AI and make its behavior follow a certain pattern so that the AI can avoid making bad decisions. Compared with autonomous artificial agents, non-autonomous artificial agents are less harmful, and will not cause liability problems, because the responsibility of non-autonomous artificial agents should be attributed to its designer or controller, while the responsibility of unpredictable artificial agents is vague, so we should not give artificial agents moral status.

3.2. The defense of granting artificial agents a moral status

There are two reasons to give AI agents a moral status. On the one hand, it is technically and theoretically feasible to create artificial moral agents. In terms of technology, the research and development of artificial intelligence agents have made some progress, and shortly, like the complete simulation of the human brain on the computer, that is, Ray Kurzweil believes that "an important application is to connect the human brain and the computer" [2], and then combine biotechnology and nanotechnology, to achieve strong artificial intelligence. Such technology is possible in the future, that is, AI agents can have autonomy and moral status.

In theory, not having an exact definition of the nature of "ethics" is not a valid reason not to give artificial agents moral status, because the idea implies that if a concept is unclear, then we can't use it, which is counterintuitive to our daily lives. We will always encounter inaccurate concepts in daily life, but this does not prevent us from using and understanding them. Taking "morality" as an example, we have disputes about what is "morality" in our daily lives, but this does not prevent us from giving "morality" to ourselves. So even if we don't have an exact definition of what is "moral" or "ethical, " we can still give an AI a moral status. And we think that at least two ethical frameworks are possible, one deontological or Kantian, and the other consequentialist. Wan Kim, John Hooker, and Thomas Donaldson have argued that one can program artificial agents using Kantian ethics in deontic modal logic. In their view, AI can be aligned with human values if it follows the principles of universality, autonomy, and morality.

The consequentialist ethical theory can simplify the behavior of the AI agent into a proposition, that is, the behavior is correct if and only if, from the perspective of justice, the behavior has the best result among all the alternatives within the scope of the agent's power, and the behavior with the best result is the mandatory behavior that the agent must take. And the agents should view the consequences of their actions fairly. Josiah Della Foresta argues that the calculation of utility by AI agents based on the correct actions of the best results in a consistent, complete, and practical manner is more reliable than human moral calculation, and that consequentialist monism avoids the thorny problem of moral conflict [14]. In other words, consequence-oriented ethics asserts that "the criterion to judge whether an action is moral is not the goodwill of the actor, but the actual effect of the action. As long as the result of the action is conducive to the interests and happiness of the majority of people, the action is moral" [15]. This is a path of "moral objectification", which is to combine traditional moral principles with artificial agents, to successfully make artificial agents into the role of norms of behavior.

On the other hand, it is also necessary to give artificial intelligence a moral status. First, because of the inevitable need for its development. With the development of artificial intelligence, strong artificial intelligence is bound to replace weak artificial intelligence and occupy the mainstream of artificial intelligence, because strong artificial intelligence is a conscious artificial intelligence with a brain [16]. In his book The Singularity is Near, Kurzweil emphasizes the so-called "singularity moment", arguing that "singularity is a transcendence", "transcendence refers to the various levels of reality" [2], including us and our technology, culture, art, as well as the creation of emotions, spiritual emotions, that is, the creation of the natural world. In other words, "when artificial intelligence crosses this tipping point, it will surpass human intelligence by hundreds of thousands of speed and efficiency" [17]. At this time, strong artificial intelligence is super thinking that has the characteristics of human thinking and far exceeds the human brain. If the moral framework is not designed for it, it may be used incorrectly. Therefore, the development of artificial moral subjects is also an inevitable trend.

Second, as the functions presented by the development of artificial agents become more and more complex, it is increasingly difficult to predict the behavior of artificial agents. Due to the complexity of the development of artificial agents and the difficulty of predicting their behavior, people must set a certain moral framework for artificial agents to manage these artificial agents [18]. In the case of autonomous vehicles, autonomous vehicles may face real-time ethical decisions, such as the question of how to choose in a situation similar to the trolley dilemma, that is, the question of whose safety to prioritize. In short, "driverless cars and other AI-equipped machines - which make their own decisions - seem to need moral guidance" [13]. Therefore, it is inevitable that some AI agents will need to make autonomous moral decisions.

Third, in the process of using artificial intelligence agents, if the artificial intelligence agents do not set certain moral norms, there may be infringements on the rights of people. For example, a nursing robot whose sole concern is to ensure the patient's health as much as possible may force the patient to take medication to ensure the patient's health, but the nursing robot's coercion of the patient ignores the patient's autonomous right not to take medication. Therefore, to prevent the infringement of human rights, it is also necessary to set a reasonable moral framework for AI agents [18].

4. The exploration of the moral framework of artificial agents

From the above, it can be seen that it is feasible and necessary to give artificial intelligence agents moral status. So how do you give artificial agents moral status? Three main approaches have been proposed: the top-down approach, the bottom-up approach, and the hybrid approach. Among them, the hybrid approach is the most advantageous. In addition, as the development of artificial agents becomes more complex, giving artificial agents moral status must solve two key problems: the problem of control and the problem of moral consistency.

4.1. A possible model for conferring moral status on artificial agents

Moral status is "the position of an actor in a moral activity". This position determines not only the role of the actor (such as a moral subject or object) but also the responsibilities and rights of the actor [19]. So, when will we be able to give artificial agents moral status? When an AI can deliberate, autonomy, and accountability for its actions, it can be given moral status. The AI is made up of two parts: one, which we call the hardware, is made up of "a complex set of physical rather than biological beings that are undeniably devoid of self-awareness and intention" [15]; The other part is composed of software, its core is a set of complex algorithm program, and according to the view of the computer school, whether human or computer, its cognitive behavior can be reduced to calculation, mentioning this point we naturally can not avoid the very famous Turing Test in the history of artificial intelligence development. Turing "devised the classic Turing Test in the history of artificial intelligence, concluding that a computer that passes the test is intelligent" [20], that is, an AI that passes the test can think. This further means, then, that the A can be considered a rational agent and thus a moral agent. Although Searle proposed the "Chinese House" test to refute the Turing test, arguing that "the computer program that constitutes a strong AI is grammatical in nature, but the human mind is semantic", therefore "the general AI lacks the human mind's ability to understand semantics, and it is impossible to obtain the same moral status as human beings" [19]. But Arielle L. Zuckerberg doesn't see it that way, because it seems unfair that we would require an AI to provide such proof to grant it moral status when we don't require other people to prove that they are conscious, but simply act as if they are [1]. Therefore, our social and moral evaluation should be entirely dependent on a person's appearance, rather than the requirement to be practical, and according to a social norm of human interaction, an AI that can pass the Turing test can also be a member of our moral community. Therefore, AI agents that can pass the Turing test may somehow join our moral community, and we have no reason to treat them differently.

So, how to make artificial agents able to make moral choices autonomously, that is, to give artificial agents moral status? Three main approaches have been proposed: the top-down approach, the bottom-up approach, and the hybrid approach. The top-down approach refers to the input of specific moral principles into the AI agent, that is, the moral principles are programmed into the AI agent's system so that the AI agent can make correct moral judgments in the corresponding situation. For example, it is possible to input a consequentialist or deontological moral program for the AI agent, requiring the AI agent to make moral judgments according to its rules, such as the deontological DIARC/ADE cognitive robot developed by computer scientists Briggs and Sultz at Tufts University. Morteza Dehghani, a scholar from the University of Southern California in the United States, and others have modeled a cognitive model of Moral Decision-Making-MoraIDM(Moral decision-making), which uses first-principle reasoning and analogical reasoning to improve the moral decision-making of artificial agents [21].

The bottom-up approach does not assign any specific moral program to the AI, but instead allows the AI to construct its moral framework by observing human moral behavior. For example, an early self-driving car developed by researchers at Carnegie Mellon University was able to drive on a highway in just 2-3 minutes after being trained by a human driver, A team at NVIDIA Corporation has demonstrated a driverless car that uses "end-to-end" machine learning technology to drive itself after 72 hours of observing human driving data [13].

The bottom-up approach can continue to be subdivided into two types, one of which simulates an environment in which there are evolutionary pressures that might induce an AI to become a moral agent by iteratively interacting with other AI agents. Another bottom-up approach is similar to teaching children to learn moral knowledge, that is, through reinforcement learning and case training, AI agents become moral agents [14]. For example, the case-based reasoning (CBR) based BDI agent model, Casuist BDI (Belief-wish-Intention) agent model, designed by Iranian scholars A. R. Honarvar and N. Ghasen-Aghaee, adjusts moral behavior [21].

The hybrid approach refers to the combination of the top-down approach and bottom-up approach, that is, to set certain moral rules for the artificial agent, and then conduct moral learning under the framework of moral rules. The main function of such moral rules is to provide moral guidance for the artificial agent to learn moral knowledge. For example, CareBot, designed by Australian scholar Marder and American philosopher Franklin, is an agent based on LIDA, that is, its moral decision-making adopts a combination of top-down and bottom-up models [21].

The hybrid approach has advantages over the first two approaches because it combines the advantages of both top-down and bottom-up approaches. Moreover, Aristotle had already provided an important revelation for this, in his view, it is intellectual virtue that can be taught, and ethical virtue that needs to be acquired through practice. That is to say, "different virtues need to be placed in AI in different ways" [22]. T two approaches have their respective effects. On the one hand, the top-down approach of setting moral rules for the AI agent can avoid the AI agent from learning wrong moral knowledge as much as possible and reduce the probability of the AI agent's unpredictable behavior; On the other hand, the bottom-up approach gives the AI the ability to react to different situations, thus increasing its flexibility, and the AI can continuously develop itself through learning so that it can cope with more complex moral situations.

4.2. Possible problems in granting moral status to artificial intelligence agents

There is no denying that the development of AI agents will have a huge impact on our lives in the future, so it is even more important to understand how we might interact with and have ethical obligations to future AI agents. Therefore, giving AI agents moral status inevitably requires addressing two key issues, namely the problem of control and the problem of moral consistency. First, if we give the AI a moral status, do we still have control over the AI? On the one hand, if we control the AI, can the AI still have autonomy? Once we have too much control over AI, it is obvious that the will lose its status as a moral agent and become a tool. On the other hand, if we do not control AI, then we may be exposed to many uncertain risks. In addition, AI agents may not be able to recognize the value of humans, which could pose a potential threat to the survival of the human race. That is to say, if we do not control it, then there may be a situation in which artificial agents harm humans, such as an out-of-control AI that in turn exterminates or enslaves humans [23]. Therefore, the issue of control is the primary issue to give artificial agents moral status, and we need to find a balance between preventing the infringement of artificial agents on human beings and respecting the autonomy of artificial agents. Secondly, it is also an important issue how to make the artificial agent's moral framework consistent with human's correct or reasonable moral framework when constructing it [24], because the artificial agent may not understand human moral knowledge due to the complexity of human morality during moral learning. Making it morally compatible with humans is another important issue that needs to be addressed to give artificial agents moral status. The requirements for such moral consistency are also multifunctional."People not only need the artificial moral subject to consistently abide by its established moral norms but also require the constructed machine ethical framework to maintain internal consistency, to ensure that similar moral value judgments can be made when similar moral situations occur in the future" [25].

In conclusion, morality has always been the hottest topic in the discussion of artificial agents, and whether artificial agents can be moral agents is a long and controversial issue. The more we expect of an AI, the more necessary it is for its application to be constrained by an appropriate moral framework, so it is necessary to build an "artificial moral subject", that is, to embed "morality" in an "AI agent". The debate over whether AI agents can have moral status, and how we can make them, shows that understanding the concept of morality is just as difficult as how to embed it in AI agents, which means that we will face many challenges on the road to giving AI agents moral status.

References

[1] Zuckerberg, Arielle L. Moral Agency and Advancements in Artificial Intelligence [J]. CMC Senior Theses, 2010, 36.

[2] By Kurzweil, Li Qingcheng, Dong Zhenhua, Translated by Tian Yuan. The Singularity is near [M]. Beijing: China Machine Press, 2011. 9.

[3] Boyles, Robert James M. Philosophical Signposts for Artificial Moral Agent Frameworks[J]. Suri, 2017, 6(02):92-109.

[4] Liu Yong'an. The Structural Perspective of the Ethical Discourse Power of Artificial Technology Entities [J]. Research in Dialectics of Nature, 2021, 37(5):29-35.

[5] Sahu M K. Kantian Notion of freedom and Autonomy of Artificial Agency[J]. Prometeica-Revista De Filosofia Y Ciencias, 2021, 23:136-149.

[6] Yan Kunru. Do Artificial Intelligent Machines are moral Agents? [J]. Research in Dialectics of Nature, 2019, 35(05):47-51.

[7] Jian Xiaoxuan. Study on Moral Status of Artificial lntelligence from the Perspective of Responsibility [J]. Journal of Xiangtan University (Philosophy and Social Sciences Edition), 2019,

44(05) : 133-138.
[8] Yan Tao. Why the Ethical Construction of Artificial Intelligence ls Possible [J]. Philosophical Analysis, 2023, 14 (01): 148-158+99.
[9] Sullins J P. When is a robot a moral agent[J]. International Review of Information Ethics, 2006, 6 (12).
[10] Robert Sparrow. Why machines cannot be moral[J]. AI and Society, 2021, 36(3):1-9.
[11] Patrick Chisan Hew. Artificial moral agents are infeasible with foreseeable technologies[J]. Ethics and Information Technology, 2014, 16 (03).
[12] Tubert, Ariela. Ethical Machines? [J]. Seattle University Law Review, 2018, 41 (04).
[13] Etzioni, Amitai & Etzioni, Oren. Incorporating Ethics into Artificial Intelligence[J]. The Journal of Ethics, 2017, 21(4).
[14] Della Foresta, Josiah. Consequentialism & Machine Ethics: Towards a Foundational Machine Ethic to Ensure the Right Action of Artificial Moral Agents[J]. Montreal AI Ethics Institute, 2020.
[15] Zhang Jinjie. Discussion on the Status of Moral Subject of Srtificial Intelligence [J]. Seeker, 2022(01): 58-65.
[16] Zhang Yijia. Ethical Dilemmas and Risks Faced by Strong Artificial Intelligence Devices [J]. Jiangsu Social Sciences, 2023, (02): 58-67.
[17] Pan Bin. Moral Embedding in Artificial Intelligence [J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2020, 34(2):1-6.
[18] Paul Formosa, Malcolm Ryan. Making moral machines: why we need artificial moral agents[J]. AI & Soc, 2021(36):839-851.
[19] Cheng Peng, Gao Siyang. Philosophical Reflection on the Moral Status of AGI [J]. Studies in Dialectics of Nature, 2021, 37(7):46-51.
[20] Sun Hui. Will Humans Be Replaced by Artificial lntelligence? lmitation, Understanding and Intelligence [J]. Journal of China University of Mining and Technology (Social Sciences Edition), 2021, 23(03): 140-150.
[21] Cheng Haidong, Chen Fan. The Moral Practice of Artificial lntelligence Agents from the Perspective of lnteraction Construction[J]. Journal of Beijing Normal University (Social Science Edition), 2022(06): 145-153.
[22] Wu Tongli. Is Al Qualified to Be a Moral Agent [J]. Philosophical Trends, 2021(06): 104-116+128.
[23] Arvan Marcus. Varieties of Artificial Moral Agency and the New Control Problem[J]. Humana. Mente - Journal of Philosophical Studies, 2002, 15(42).
[24] Muller Vincent C. Ethics of Artificial Intelligence and Robotics[J]. The Routledge social science handbook of AI, 2020:1-70.
[25] Su Lingyin. The Construction of Ethical Framework:The Main Assignments of Machine Ethics [J]. Journal of Shanghai Normal University (Philosophy and Social Sciences Edition), 2019, 48(1):76-86.