Proceedings of the 5th International Conference on Literature, Language, and Culture Development
DOI: 10.54254/2753-7064/2026.32918

From Categorical Imperative to Algorithms — A Dialogue Between Kantian Ethics and Artificial Intelligence Ethics

Yujia Wu

The University of Sydney, Sydney, Australia
3453967186@qq.com

Abstract. As the rapid development of AI (Artificial Intelligence), a series of profound ethical issues are illustrated to individuals. The most significant issues are the spread of algorithmic discrimination, and the encroachment of human dignity caused by instrumental rationality. In order to solve these issues, it is urgent for ethics to provide principled guidance. In this paper, I will display that Kantian Ethics, especially its core concept of categorical imperative, can provide irreplaceable formal foundation and critical lens for AI ethics. Through systematically explaining the key concepts of Kantian ethics, which are autonomy, self-legislation, and the principle that humanity is an end in itself. The current AI cannot be considered as a Kantian moral agent, because of its internal heteronomous nature and lack of free will. Based on this argument, this paper further discusses the programs of reconceptualized responsibility in the age of algorithms. Specifically, on the one hand, although AI is not a moral agent, its distributed feature may correspond to a distributed model of responsibility attribution. On the other hand, the formal features of algorithms provide a possibility for transforming the procedural cores of categorical imperative, such as the test of universalizability, into principles of ethical algorithm design. In conclusion, the dialogue between Kantian Ethics and AI is mutually enriching. Kantian ethics not only offers an essential ethical standard for evaluating and regulating AI systems, but also illustrates enduring vitality while facing contemporary technological challenges.

Keywords: Kant, Ethics, Categorical Imperative, Algorithm

1. Introduction

With the rapid development of AI, the ethics of AI need to be emphasized along with the technological development of AI, including its practitioners and its users. Users of AI may be negatively affected by algorithms, while practitioners need to safeguard the continued development of AI. This means that both AI practitioners and users need to actively address the dilemmas of AI ethics, improve AI ethics, and thus resolve the dilemmas they are caught in. In addition, seeking guidance from appropriate ethics is crucial to solving the dilemma of AI ethics. Pan Shilian mentions, "The integration of AI and Kantian moral philosophy is a mutual empowerment." [1] This paper similarly supports that Kantian moral philosophy can largely inspire AI ethics, and aims to discuss the significance of Kantian moral philosophy for AI ethics and how it can address the practical dilemmas of AI ethics: one is the algorithmic discrimination of the majority against the

DOI: 10.54254/2753-7064/2026.32918

minority; and the other is the dissolution of human dignity and purpose by the proliferation of instrumental rationality.

This paper will show how the core concepts of Kant's moral philosophy can be applied specifically to AI ethics, and how the problem of algorithmic discrimination can be solved by combining self-legislation and algorithms from the categorical imperative to the algorithmic imperative. Also how to stop the dissolution of the Kantian concept of dignity by referring to it in order to explore the balance between instrumental rationality and human dignity and purpose. In addition, this paper also affirms that Kantian ethics has indeed energized Kantian ethics in contemporary ethics through its intersection with contemporary AI ethics.

2. Key concepts of Kantian ethics

2.1. Free will and autonomy

In Kant's ethics, autonomy and self-legislation occupy an indisputably central position. To understand the authority of these two concepts, one must first grasp the intellectual development of Kant's conception of freedom, grounded in humanity's pursuit of freedom and the resulting notions of autonomy and self-legislation. Huang Suzhen demonstrates Kant drawing on ontology and his theory of the two worlds, which are the "world of the understanding" and the "world of the senses", and discusses the justification of free will from the perspectives of both speculative reason and practical reason [2]. In Kant's theory, when we talk about speculative reasoning, the object of "causality" is natural phenomena. People cannot change these objects in a big way. So freedom is only something we can imagine. We cannot prove it works in real life, and we cannot make sure it is real objectively. But Huang Suzhen has a different view. She talks about practical reason that focuses on causality. The person who takes action makes up the full and basic reason for the action. So free will is proved to be totally decided by itself. Freedom is completely real and objective. What's more, based on practical reason, Huang points out Kant's focus on morality. Freedom must follow basic rules [3]. Kant put forward the idea of positive freedom. It stands for making rules for oneself. It is different from negative freedom and rules that depend on outside things. Freedom must follow causal rules from one's own reason, and this is self-legislation. So practical reason can make rules for itself. Positive freedom is the basic rule of freedom. It sets limits to freedom to meet moral needs. This makes morality possible in the first place.

At the same time, when talking about the moral needs of freedom, Huang Suzhen says we must pay attention to the link among moral law, good will and free will [2]. To be specific, free will acts as a bridge to connect moral law and good will. Kant says, "In the world, we can only think of good will as something good without any conditions. No other thing can be like this, even if it seems to go beyond it." Besides, the idea of good will is part of the idea of duty. The idea of duty also helps explain good will clearly. An action can have moral value only when it is driven by duty. It is not enough just to follow duty passively. This means good will and duty are very important for the moral meaning of an action. So we need to talk about how to use good will and act for duty to make actions have moral value. So Huang Ge says, in Kant's theory, actions driven by duty need to be limited and guided by a rule to get moral value. This rule is moral law [3]. To be specific, people do not have perfect good will or unconditional goodness. So they must follow moral laws to make sure their actions are driven by duty. At the same time, people can follow moral laws only if free will exists. Without free will, people cannot choose to follow moral laws and act for duty on their own.

When we talk about linking Kantian ethics with AI, we need to think about whether AI can be a moral agent in Kant's theory. From the Kantian ethical ideas in this article and other scholars'

DOI: 10.54254/2753-7064/2026.32918

studies, rational beings with freedom and self-rule are usually human beings. AI copies human intelligence. In Kantian ethics, a moral subject must have self-awareness, practical reason and the ability to set goals. This article will discuss this in later parts. It will check if AI meets the conditions of Kantian ethics. If not, it will also see if AI can still be seen as a moral agent.

2.2. The categorical imperative

When we talk about the link between moral laws and the categorical imperative, Jining puts forward a clear view. Human beings do not have a complete good will, and they have limited ability, so they cannot always follow moral laws in a necessary way. For this reason, the categorical imperative that belongs to the human will is just the moral law itself [4]. Also, moral laws are objective principles, and they have binding power on the human will. So these moral laws are called "commandments", and the form of such a "commandment" is what we call an imperative. To explain the moral law in a clearer way, Kant makes a distinction between hypothetical imperatives and categorical imperatives. To be more specific, Kant divides all imperatives into three different types. They are "rules of skill", "advice of prudence" and "moral laws" respectively. The first two kinds of imperatives depend on practical goals, so they are hypothetical imperatives. But the "moral law" is objective, necessary and useful for everyone in all cases, so it forms the categorical imperative. So Kant holds the idea that for the human will, the moral law is an absolute imperative. This fact also shows that the moral law has real and proven authority.

Besides, Kant wants to show the standard form of the absolute imperative. After he compares hypothetical imperatives with categorical imperatives, he puts forward three formulas of the categorical imperative [4]. Among these three formulas, the formula of the universal law is the most important one. It builds the basic ground for the whole idea of autonomy and also for the other two formulas. Kant says, "The statement that 'in every action, the will makes a law for itself' only shows this principle: 'You should act only by the maxim that you can wish to become a universal law for all people at the same time.'" Also, humans are rational beings, so they must make sure that the maxims that guide their self-directed actions follow universal laws. These maxims should work for all people in the same way, instead of fitting the changeable and uncertain features of human nature [3]. So the maxim of the will has to match the universal law, and this maxim must come from a person's own reason. When it comes to the formula of the teleology of human nature, Huang proves that Kant builds a tight connection between two ideas. One is the concept of the end-in-itself or the necessary end, and the other is the formula of the universal law. This means that rational beings should be seen as ends for their own sake, not as tools to achieve the goals of the will. In this way, universality is shown in the desires of rational beings [3]. With this formula, Kant provides a real motive for the rational will to follow universal laws. The purpose of rational beings is the key and decisive basis for this motive. At last, Kant puts forward the "formula of autonomy". He improves the first two formulas by getting rid of all kinds of interests, like force and temptation. This is to make a clear difference between autonomy and heteronomy. To be specific, Kant thinks that if humans only follow a law, the reason must be that the law brings certain interests to them. This also means that the maxim of the will does not come from the universal law-making of one's own reason.

The categorical imperative is the procedural core of ethics, and this paper will also talk about the link between the categorical imperative and the ethics of artificial intelligence. To be specific, the categorical imperative is not a specific and fixed moral rule. Instead, it is a way for rational people to examine their own thoughts and actions. Can the formal feature of the categorical imperative be turned into the procedural logic of algorithms? According to the article written by Zhang Chuanyou, Kant's moral philosophy has two key paths, namely the "ascending path" and the "descending path".

DOI: 10.54254/2753-7064/2026.32918

The ascending path starts from general moral theory, then moves to the theoretical discussions in moral metaphysics. After that, it goes further to the level of the Critique of Practical Reason, and finally reaches the formal nature of pure reason [5]. The descending path is the process of going back from metaphysics to the real and empirical world. This descending path gives clear guidance to the ethics of artificial intelligence. It tells how to return from pure formalism to the real empirical world. In this way, it creates a possibility that is closer to people's daily life and easier for people to accept and understand.

3. A Dialogue between Kantian ethics and artificial intelligence ethics

3.1. Algorithmic cognition and moral agency in artificial intelligence

In terms of the relationship between Kantian ethics and AI ethics, the question of whether AI can become a Kantian moral agency is essential, and various philosophers hold different perspectives. Fan Xinyue argues that because in traditional ethics, consciousness is a significant prerequisite for free will and being a moral agent, the question of whether AI can have consciousness is the key to the debate. Meanwhile, Fan Xinyue expresses those algorithms as the core of AI, the question of whether it can imitate and generate consciousness also should be considered and be examined from ontological and ethical perspectives [6]. Specifically, from the ontological perspective, having consciousness and free will is a reflection of human agency, and biological mechanisms cannot completely revert it. As a result, even if AI algorithms can imitate biological mechanisms, it still cannot have human intelligence and human agency. From the ethical perspective, having consciousness and free will requires that individuals morally be able to take responsibility for their action, without being influenced by external constraints and enforcement. First, AI cannot understand moral laws. Furthermore, its algorithms and logic are established by its designers, and it could be considered constrained and forced by external factors, and generally cannot be responsible for its actions. At the same time, Liu Zuo claims that from a cognitive perspective, AI is predetermined because it is difficult for it to have spontaneity, and it is not free. From a practical perspective, AI barely has the ability to take actions, without the ability to act freely as a rational agent [7]. Therefore, according to aforementioned aspects, AI does not have consciousness and free will , and cannot be a Kantian moral agent.

On the other hand, there is an argument that although AI does not have human consciousness, it possesses machine consciousness, and may even have the opportunity to exceed human intelligence. Specifically, algorithmic expression of AI has formal characteristics, and theoretically human cognition and behavior which are formal representations, can be simulated by AI. As a result, AI may be able to imitate the activity mode of consciousness to some extent, and should not be simply determined as lacking consciousness and free will based on aforementioned reasons. Moreover, AI algorithms are formally following general programming principles, and they can be considered as following universal laws, without being impacted by weak will [7]. This implies that, to some extent, AI has the possibility of possessing ethical algorithms, and individuals should not hold complete negative attitude to the connection between Artificial Intelligence ethics and Kantian ethics.

In terms of the cognition aspects of AI, Fan Xinyue illustrates that AI is traditionally considered as incapable of cognition. Because although it can formally simulate human non-voluntary behaviours to archive intelligence, it is still difficult to explain the production of "qualitative experiences." [6]. This implies that individuals can ignore mistakes of rules in daily life, and accurately understand the meaning while dealing with information. Compared with individuals, it is

DOI: 10.54254/2753-7064/2026.32918

difficult for AI to understand new situations which are outside the established rules. Furthermore, as the development of AI and the improvement of programming compilation techniques, AI continues to gain cognitive insights of questions through a process of trial and error, and provides the evidence that AI has the potential to imitate human cognition. Therefore, AI can be considered as possessing cognitive abilities, although it cannot have human consciousness and be a Kantian moral agent, it is still deeply connected to human features, and be guided by Kantian ethics.

In summary, even though AI has cognitive abilities and its algorithmic characteristics are connected with concepts in Kantian ethics, it cannot be considered a Kantian moral agent because it lacks consciousness and free will. Furthermore, as the rapid development of AI, determining what is a better model of human machine interaction and establishing a new model of responsibility have become critical issues. On the one hand, the issue of assigning responsibility to current AI should be explained from various perspectives. On the other hand, in terms of which AI can be considered a Kantian moral agent, there is still room for further research.

3.2. Redefining responsibility in the age of algorithms

Regarding the issue of assigning responsibility for AI in the age of algorithms, accountability, responsibility, and the rule of law are generally considered as fundamental requirements to face this new technology. Fan Xinyue suggests that, in terms of taking moral and legal responsibility, according to Kant's deontological theory, moral responsibility can only be taken by moral agents with free will, while legal responsibility must be determined based on the allocation of moral responsibility [6]. This means that, from a deontological perspective, AI cannot currently take legal responsibility because it lacks free will. For the continued development of AI, it requires a mechanism for assigning responsibility in order to identify vulnerabilities and correct errors, and transform moral experience into algorithmic data. In terms of causal responsibility, it is difficult for individuals to infer moral responsibility from its attribution. Specifically, causal responsibility demands judgment based on causal relationships. In practice, the causal relationships involving AI are extremely complex, because they not only extend beyond the AI itself, but also the rights and capacities required for moral responsibility are not equal to those of the actor. As a result, in practice the actor who actually causes the outcome may cannot meet the requirements of moral responsibility [6]. So, causal liability is not a proper standard of responsibility for artificial intelligence. When it comes to duty-based liability, the subjects and actors of this liability are human beings, and its dividing lines are not clear enough. Even so, we can still prove that AI has a certain level of independent agency, so it is possible for AI to have the potential to take on this kind of liability. To be specific, we carry out moral sanctions because we require the acting subject to know its mistakes and fix them. This rule also fits the situation of artificial intelligence, for AI can both find out its mistakes and change its programs to correct these mistakes. This means that if AI has moral autonomy through its own algorithms, it can be given the ability to take responsibility under the duty-based liability standard. Besides, Fan Xinyue points out that AI's decisions and actions usually come from the repeated interactions of many different participants. For this reason, it is not right for only one single party to take all the responsibility, and as a result, AI's liability under the duty-based standard is shared by different parties [6]. So, when we talk about how to assign responsibility for artificial intelligence, we need to consider its distributed feature and the related questions about whether AI can bear moral responsibility. People still need to further improve and perfect the moral rules and the ethical governance system for artificial intelligence, so as to explore better and more effective solutions.

DOI: 10.54254/2753-7064/2026.32918

What's more, Li Yaming holds the view that we can build ethical principles for artificial intelligence on the basis of Kantian theory [8]. According to Kant's theory, the principles that can guide an acting subject's actions can only come from the subject's own self-reflection. When we act and reflect on our actions, we can draw a conclusion that reasonable actions need to be justified, and actions also need basic freedom and well-being, which are necessary valuable things. This means that as an acting subject, moral justification is very important for artificial intelligence. This kind of justification is formed through a process of careful thinking, and the standard for this thinking is "no reason to refuse". Based on this Kantian theory, Li Yaming proves the basic premises of AI ethical principles and gets three key principles from these premises. These three principles all focus on and protect the necessary valuable things of the acting subject. They are the non-aggression principle, the equality principle, and the assistance principle [8]. At the same time, these three basic moral principles also make up the requirements for "human moral status" in many different moral and philosophical theories. This reflects that although AI lacks free will and finds it difficult to bear moral responsibility, as a moral agent, it must adapt to the requirements and relevant standards of "human moral status." Furthermore, although the three principles based on Kantian theory seem to generally support the idea that agents, including AI, should not take proactive action, this set of principles can still provide excellent moral guidance for AI ethics and possesses advantages over other principles. The most fundamental moral obligations constitute the core principles of AI, thereby helping to prevent moral errors and even threats to human life.

4. Conclusion

The core concepts of Kantian ethics including free will, autonomy, and the categorical imperative, they offer outstanding theoretical contributions and solid practical guidance to the field of AI ethics, demonstrating its enduring relevance in the contemporary world. Kantian ethics not only provides a deep foundation for resolving ethical dilemmas in artificial intelligence but also offers excellent guidance for AI practice. Regarding ethical dilemmas in artificial intelligence, Gordon and Gunkel state, "The emergence of artificial intelligent robots compels us to reconsider many of our moral terms and to revise our existing ethical and moral conceptions." [9] "Whether artificial intelligence can become a Kantian moral agent" has become the core of the issue and the key to resolving the dilemma. This paper holds that at present, artificial intelligence cannot become a Kantian moral agent because it lacks consciousness and free will, and many philosophers share this view. Yet AI can have machine consciousness, and its algorithms can copy human cognition and behaviors through formal expressions, so it has certain potential for ethical algorithms. Besides, with the development of AI, it has owned cognitive abilities and mimics human thinking. So even though AI is not a Kantian moral agent, it imitates humans in many ways and can be guided by Kantian ethics. For the practical use of AI, we face governance challenges in the future human-machine coexistence society, so rebuilding responsibility in the algorithmic age is the most important thing. It is hard for AI to take moral and causal responsibility, but it can be given duties if it has moral autonomy via algorithms. Also, because of its distributed feature and based on the responsibility traceability principle, rules must make all human responsibility nodes in the distributed system clear and responsible. The making of these rules needs guidance from Kantian ethics, and we can build AI ethical principles on Kantian theory as well. So Kantian ethics plays a key role in both theoretical and practical parts of AI ethics. This paper argues that the development of AI ethics still needs the guidance of Kantian ethics, and Kantian ethics will keep up with the new times and gain new vitality in the modern world through such guidance.

DOI: 10.54254/2753-7064/2026.32918

References

[1] Pan, S. L. (2025). An Exploration of the Kantian Moral Philosophy Solution to Artificial Intelligence Ethics. Journal of Guiyang University Social Sciences, 20(5),48-56.
[2] Huang, S. Z. (2023). Kant's justification of free will and its significance in practice. Journal of Yunnan Social Science, 2, 26-33.
[3] Huang, G. (2020). A Study on the Idea of Autonomy in Kant's Practical Philosophy. Dissertation of Shandong University.
[4] Ji, N. (2017), Kant's Quest for the Highest Moral Principles. Dissertation of Central Party School of the Communist Party of China.
[5] Zhang, C. Y. (2007), The Rising Road and the Falling Road of Kant's Moral Philosophy. Journal of Morality and Civilization, 6, 63-66.
[6] Fan, X. Y. (2022), The Problem of Responsible Subjects in AI from the Perspective of Algorithm Cognition. Dissertation of Lanzhou University.
[7] Liu, Z. (2025), Can Artificial Intelligence be a Kantian Rational Being? Journal of Academic Forum, 5, 12-22.
[8] Li, Y. M. (2023), Contemporary Research on Moral Normativity and the Ethical Design of Artificial Intelligence. Journal of Philosophical Trends,5,105-114.
[9] Gordon, J. S., & Gunkel, D. J. (2022). Moral status and intelligent robots. The Southern journal of philosophy, 60(1), 88-117.