Does AI Share Same Ethic with Human Being?

Zilong Feng

▶ **To cite this version:**

Zilong Feng. Does AI Share Same Ethic with Human Being?. 2nd International Conference on Intelligence Science (ICIS), Nov 2018, Beijing, China. pp.465-472, 10.1007/978-3-030-01313-4_49 . hal-02118822

HAL Id: hal-02118822
https://inria.hal.science/hal-02118822

Submitted on 3 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Does AI share same ethic with human being?

——from the perspective of virtue ethics

Zilong Feng

Peking university
Department of Philosophy, and of Religious Studies
No.5 Yiheyuan Road Haidian District, Beijing, P.R.China
100871 Beijing

13210160038@fudan.edu.cn

Abstract. From the perspective of virtue ethics, this paper points out that Artificial Intelligence becomes more and more like an ethic subject which can take responsibility with its improvement of autonomy and sensitivity. This paper intends to point out that it will produce many problems to tackle the questions of ethics of Artificial Intelligence through programming the codes of abstract moral principle. It is at first a social integration question rather than a technical question when we talk about the question of AI's ethics. From the perspective of historical and social premises of ethics, in what kind of degree Artificial Intelligence can share the same ethics system with human equals to the degree of its integration into the narrative of human's society. And this is also a process of establishing a common social cooperation system between human and Artificial Intelligence. Furthermore, self-consciousness and responsibility are also social conceptions that established by recognition, and the Artificial Intelligence's identity for its individual social role is also established in the process of integration.

Keywords: Artificial Intelligence virtue ethics responsibility

1 Introduction

Imagine that the human world was to suffer from a war with AI, and they compromise after a long time. After the war, artificial intelligence (AI) establishes its own country in the Antarctic. Can we imagine the feature of its social system? In the economics of human society, we should assume some premises based on humanity to construct economics system, for example, market hypothesis or rational person. But in the field of ethics, the question becomes more complex. It is obvious in the framework of virtue ethics to admit that all kinds of virtues have its own premises which are decided by social conditions and history conditions. Of course, what kind of factors are the most important may be vary from different scholars. Karl Max and Macintyre may have different opinions on the status of the contradictions between economic and superstructure, but they all convey the same opinion that we cannot think of a kind of system of ethics without its special social and history conditions. On the other hand, we

also cannot think of the social and history conditions of ethics without considering humanity. However, it is quite difficult to definite the so-called humanity of AI. How can we give a set of human moral laws to AI by programming without permission from AI? Is it fair to give the set of human moral laws—— just as the Isaac Asimov's three laws of robotics ——to AI? The viewpoint of this paper is that answers to these questions are depended on the degree of the AI robot's involvement into human society. I want to prove the opinion with the basic theory structure from the virtue ethics of Macintyre.

2 The problem of abstract moral law and AI

There is a granted idea that we could program abstract moral laws into the codes of robots. Maybe the most famous moral laws given in the history are Isaac Asimov's three laws of robotics. And the three laws are listed as follow:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[1]

However, the kind of idea may face the counter-argument which holds the idea that moral laws cannot perform without considering concrete situations —— a kind structure of argument which can always be seen when you want to find some counter-argument for Kant's moral theory. The kind structure of counter-argument for Kant can be characterized more clearly when related with a problem of AI, the so-called framework problem.[2] Framework problem claims that AI is sensitive to variables around them.

Another argument for the idea to apply existing moral laws today into the area of AI or in the future society and this kind of argument can be called 'technological transparency', which means that fundamental moral laws can stay unchangeable during the developing process of technology.[3] The basic argument of technological transparency is that technology is only a kind of means which should not take the responsibility and the subject of responsibility can only be human. And the counter-argument for "technological transparency" can be more easy today, because with the development of AI, the subjective status of AI becomes not so difficult to answer. AI now becomes more and more intelligent, and it is so difficult to test whether it is really not autonomy. At least in the process of deep learning, programmer now more and more put their attention to adjust parameter rather than instruct AI directly. The process of

[1] Wikipedia, https://en.wikipedia.org/wiki/Three_Laws_of_Robotics, last accessed 2018/08/07

[2] B. Meltzer & Donald Michie (eds.): Machine Intelligence 4. Edinburgh University Press, Edinburgh (1969), pp. 463--502.

[3] Blay Whitby: Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. J. Interacting with Computers, Volume 20, Issue 3 (2008), pp. 326–333.

deep learning is a black box process. According to the embodied cognition, the conceptual structure is influenced by the body of organism.[4] The consciousness criteria of human cannot be generalized. Today, whether the AI could possess ownership of property is not so absurd to talk about. If it is possible, maybe we also need to consider whether it is necessary to give them social insurance.

In fact, attitudes towards the responsibility of AI have altered and more and more scholars are now beginning to consider it as a serious question. Part of the reason is that the combination of biology and computer science, but the most important is that freedom are always related to responsibility. Now the freedom of robots become more than ever, especially when the freedom relates to its specialist areas, and maybe the best example is automatic drive. Colin Allen and Wendell Wallach argues that whether the subject of action can become the subject of moral depends on two factors: autonomy and sensitivity. If the two factors are very high, then we can say that the AI is more like a subject of moral rather than a tool. In such condition, the responsibility belongs to the AI rather than the operator of the AI except for in some special areas, according to Gert-Jan Lokhorst and Jeroen van den Hoven, such as military robots in the war.[5] Colin Allen and Wendell Wallach believes that we can make the kind of robot which can be the subject of moral in the future although it is very hard today.

With the development of AI, scientists now are developing a kind of being which can take its own responsibility in some sense. This kind of AI has high sensitivity and autonomy, so they are subject of moral in some degree. What human should face is not only a tool but a rational being. The foundation of moral for the two situations are different. If robots are only tools that be controlled by human, then the same argument structure can be applied to the problem of ethics of robot. However, now AI is in some sense a kind of being which has the ability to make its own choice, is it still adaptable and justice to coerce them to accept human ethics?

It is important to reconsider the norm between human and AI with the thoughts of Macintyre. As Macintyre had mentioned when he talk about virtue ethics, every kind of moral system has its own historical and social premises. The norm between human and AI has to be established on the base of the thought.

3 Three stages of virtue ethics of Macintyre and AI

At first, I want to introduce the opinions of Macintyre briefly which is rarely mentioned in the analysis of robot's ethics. It is very complicated to category all kind of virtue ethics, for example, according to The Stanford Encyclopedia of Philosophy, virtue ethics can be defined in four ways. [6] In this paper, I will discuss ethics of AI with the thoughts of Macintyre who is one of the most famous scholars in this field. I will

[4] Varela, Francisco J., Thompson, Evan T., and Rosch, Eleanor: The Embodied Mind: Cognitive Science and Human Experience. The MIT Press, Cambridge, Massachusetts(1991), p. 4.

[5] Lin, Patrick: Robot Ethics, The Ethical and Social Implications of Robotics. The MIT Press, Cambridge, Massachusetts(2012), p. 154.

[6] Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/entries/ethics-virtue/, last accessed 2018/08/07.

introduce the thoughts of Macintyre in short and explain the foundation of the relationship between its thoughts and AI' ethics. There are three stages of Macintyre's thoughts in this book —— his theory of practice, theory of narrative and what makes up a tradition of moral.

We have to define what is the so-called virtue according to the text of *After Virtue: A Study of Moral Theory*. In this book, Macintyre have analyzed the conception of virtue in history to get a universal definition of virtue, and his virtue ethics established on the universal definition. The most important feature of the conception of virtue is as follow:

One of the features of the concept of a virtue which has emerges with some clarity from the argument so far is that it always requires for its application the acceptance for some prior account of certain features of social and moral life in terms of which it has to be defined and explained.[7]

For example, he lists three kinds of virtues in history, and the representations are Homer, Aristotle and Franklin. The virtues of them can be summarized as social role、the achievement of *telos*、the utility in achieving earthly and heavenly success.[8] All of three conceptions above are rooted in their social background. Of course, the influence of the Karl Marx is very obvious in this kind of viewpoint of Macintyre. The virtue of Macintyre's definition is historical and it cannot be separated from its social background. The question of ethics of AI is at first a question of social problem rather than a technology problem. Only if we can analyze its social background then we can find the basement of ethics of AI. Of course, this is not to say that it is not important at all to pay attention to the Kant's style of moral theory, just as many scholars have done so far. They designed the moral system of AI with adding to it a moral module. But this kind of module only reveals abstract conditions of moral. The thoughts of virtue ethics should be mentioned when we want to talk about particular questions.

The aim of virtue ethics is constrained by history and society. In order to argue this influence, Macintyre introduces the conception of practice and this is the first stage of his virtue ethics. His conception of practice is as follow:

By a "practice" I am going to mean any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended.[9]

We can see from his words that this conception is at first a social conception, and it is used to describe the cooperation relationship in human society. With the formation of relationship, the notion of goods internal to social activities can be defined and

---

[7] Alasdair Macintyre：After Virtue: A Study of Moral Theory. University of Notre Dame Press, Notre Dame, Indiana (2007), p. 186.

[8] Ibid., p. 185.

[9] Ibid., p.187.

admitted. Goods internal to social activities is beneficial to all members who participated in the practice. Moreover, he also mentions as follow:

A practice involves standards of excellence and obedience to rules as well as the achievement of goods. To enter into a practice is to accept the authority of those standards and the inadequacy of my own performance as judges by them.[10]

That is to say, internal goods and practice establish the regulations and criteria of ethical community. If AI wants to live in human society as an ethical subject, then at first it should be integrated into the practice of human society and have the common internal goods with human. The fear that AI may threaten to human's existence can be resolved in this social cooperation system.

The second stage succeeds with the problem of first stage. Although the practice in the first stage provides the social background of virtue, the conflicts among individuals still exist. In order to solve the problem, Macintyre puts forward the theory of the unity of human life. This theory emphasized that everyone's life is a unity, and this unity provide a full *telos* to ethics. The unity of life comes from "the unity of a narrative embodied in a single life".[11] To realize the unity of personal life has to answer what is good to human. Macintyre dose not describe the mothed of judging good, but according to his text, a good life has the characteristic of internal goods to practice. We cannot understand it without describing the process of seeking a good life, and virtue must be defined in the same process. Just as Macintyre has said as follow:

The good life for a man is the life spent in seeking for the good life for man, and the virtues necessary for the seeking are those which will enable us to understand what more and what else the good life for man is.[12]

If AI wants to get the status of moral subject, it has to have a united life which has a teleology structure. For a robot who has the ability to choose with at least limited reason, the kind of choice has to be restricted by the third stage.

The third stage of Macintyre's theory wants to unite the social person with historical practice. For Macintyre, the subject cannot be separated from its history and society. The subject is something inherited from its past, as he has said：

What I am, therefore, is in key part what I inherit, a specific past that is present to some degree in my present.[13]

For the historical practice, Macintyre says as follow:

It was important when I characterized the concept of a practice to notice that practice always have histories and that at any given moment what a practice is depends on a mode of understanding it which has been transmitted often through many generations.[14]

The unity of a narrative embodied in a single life is in the historical social practice, and social practice constitutes a relative open tradition which faces the future. The person in the history tradition is not only constricted by the tradition, but also creates

[10] Ibid., p. 190.
[11] Ibid., p. 218.
[12] Ibid., p. 219.
[13] Ibid., p. 221.
[14] Ibid., p. 221.

the tradition. For AI robots, if they are not tools but something who can take responsibility, then they are rolled in a kind of tradition and they have to take its responsibility for the tradition. The aim of AI robots' lives also relies on this tradition. Human as members of the same ethical community have right to require AI robots to take the choice which is good for the community. That is to say, AI robots have the responsibility to get the education in an ethical community.

According to Macintyre's thoughts above, every kind of ethical community has to establish itself by a social cooperation system. This kind of social relationship has its own internal goods, and virtues also depend on it. The historical subject who has its unity of narrative is in the social cooperation system. The members of the ethical community share a common ethics system in this sense. If AI really shares a common ethics system with human, they have to take part in the same ethical community in the same way. This kind of theory model can solve some problems of abstract moral module especially in the moral dilemma (as the Trolley dilemma showed), because it not only pays attention to action compared with abstract moral module. The importance of action is not so high because virtue can be potential, and it can be a kind of character even though it is not actualized. The kind of characteristics of virtue ethics are related with social role, and the following part of the paper will resolve the problem of how to identify individual and its social character in a society.

4 AI and the recognition of society

The question of self-consciousness is not only an important issue when we talk about AI, but also the core question in the modern history of philosophy. In the field of philosophy, the conception of I is not only discussed in the cognition field, but also is a conception related to society and ethics. When we talk about self-consciousness, we should not ignore the dimension of society. The phenomenon of self-consciousness cannot be explained without considering social dimension. The interaction of human and AI robots is the social premise of self-consciousness of AI robot, and it is also the premise of ethics of AI robots. To talk about ethics of AI robot usually relates to autonomy, teleology or responsibility. However, all of these conceptions are related to self-consciousness. Self-consciousness is not only a physiological phenomena, and it should be explained with the dimension of society. Even if the AI technology can make a real brain, it does not mean that the brain can really have self-consciousness. Because recognition is an important process in the establishment of self-consciousness.

We cannot have self-consciousness without being in human society. In the field of AI, conflicts such as individual and community also exist, and the difference of ethics and moral also should not be forgotten. Axel Honneth points out that ethics should be seen as follow:

The concept of "ethical life" is now meant to include the entirety of intersubjective conditions that can be shown to serve as necessary preconditions for individual self-realization.[15]

[15] Axel Honneth: The Struggle for Recognition: The Moral Grammar of Social Conflicts. translated by Joel Anderson, the MIT Press, Cambridge, Massachusetts (1995), p. 173.

Individuals must know that they are recognized for their particular abilities and traits in order to be capable of self-realization, they need a form of social esteem that they can only acquire on the basis of collectively shared goals.[16]

Society is constructed by all kinds of recognition, such as love and law. And individuals can deepen their recognition to itself in this process of recognition. Authors of science fiction and screenplays of science film are maybe more clear about this point. In many science fictions and movies, human and robots establish recognition to each other in the process of love and war, and all of these processes can be seen as the construction process of recognition. In the process of recognition, self-consciousness of AI is established too.

Of course, in the theory framework of technology of transparency, the opinions above are too weird. Because to the viewpoint of technology of transparency, AI robot is still something rather than a relative rational subject, so they belong to human and the responsibility of their behavior should also be ascribed to human too. The ethics of robot are still the extension of the ethics of human in the framework of technology of transparency. But this viewpoint has too many problems with the developing autonomy of AI robot now. Maybe it sounds so weird that a computer should be responsible with you. In fact, according to the outcome of questionnaire designed by Jamy Li in 2016, many people opposed to see autonomous car as the subject of responsibility.[17] But in the field of law, it is not so weird for a nonhuman subject to take responsibility. In fact, legal person such as company has the independent legal personality to take responsibility on many legal actions just as human. In the present framework of law, the possibility of designing a regulation about taking part responsibilities by AI robot still exists. It is too cruel for companies that produce autonomous car and buyer to take all responsibilities of autonomous car, and it could reduce the passion to develop such technologies. This is also the reason why this paper does not hold the point that it is too early to talk about the responsible subject of AI robots. The key point is the question of recognition, and law system is also the appearance of recognition.

5 Conclusion

The viewpoint that we can solve the ethical problem of AI by programming an abstract moral module has many theory problems. With the development of AI's autonomy and sensitivity, AI robots should be at least seen as part responsible subjects. From the framework of Macintyre's virtue ethics, the relationship of AI and human should be considered with premises of society and history. Of course, the dimension of recognition is also an important element which cannot be ignored when we consider the question of self-consciousness of AI which is related to the legal and moral status of AI. It is at first a social integration question when we talk about AI's ethics, and this

[16] Ibid., p. 178.

[17] Li, Jamy, Xuan Zhao, Mu-Jung Cho, Wendy Ju, and Bertram F. Malle: From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Acci-dents with Self-Driving Cars. In: SAE 2016 World Congress and Exhibition, SAE Tech-nical Paper 2016-01-0164(2016), https://doi.org/10.4271/2016-01-0164

is the premises of designing the moral module to AI robots. If we want to answer the question of AI robots, especially the ethical relationship between AI robots and human, we have to know that in what kind of degree AI can share the same ethics system with human equals to the degree of its integration into the narrative of human's history, and the degree of its integration into human's social cooperation system with goods internal to practice.

References

1. Alasdair Macintyre：After Virtue: A Study of Moral Theory. University of Notre Dame Press, Notre Dame, Indiana (2007)
2. Axel Honneth: The Struggle for Recognition: The Moral Grammar of Social Conflicts. translated by Joel Anderson, the MIT Press, Cambridge, Massachusetts (1995)
3. Varela, Francisco J., Thompson, Evan T., and Rosch, Eleanor: The Embodied Mind: Cognitive Science and Human Experience. The MIT Press, Cambridge, Massachusetts(1991)
4. B. Meltzer & Donald Michie (eds.): Machine Intelligence 4. Edinburgh University Press, Edinburgh (1969)
5. Blay Whitby: Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. J. Interacting with Computers, Volume 20, Issue 3 (2008)
6. Wikipedia, https://en.wikipedia.org/wiki/Three_Laws_of_Robotics
7. Lin, Patrick: Robot Ethics, The Ethical and Social Implications of Robotics. The MIT Press, Cambridge, Massachusetts(2012)
8. Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/entries/ethics-virtue/
9. Li, Jamy, Xuan Zhao, Mu-Jung Cho, Wendy Ju, and Bertram F. Malle: From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars. In: SAE 2016 World Congress and Exhibition, SAE Technical Paper 2016-01-0164(2016), https://doi.org/10.4271/2016-01-0164