Survey: Artificial Intelligence Ethics

Shuxi Wang [α] & Zengyan Xia [σ]

Abstract- **Artificial Intelligence Ethics is playing an important role with the development of Artificial Intelligence (AI). It is popular recognized that obeying to Artificial Intelligence Ethics guidelines and principles may resolve so many problems caused by Artificial Intelligence. This paper reviewed the development history of Artificial Intelligence Ethics, listed the main guidelines and principles of Artificial Intelligence Ethics, proposed the methods of Artificial Intelligence Ethics governance, discussed related algorithms to solve Artificial Intelligence Ethics problems.**

Keywords: artificial intelligence; artificial intelligence ethics; artificial intelligence ethics guidelines and principles; artificial intelligence ethics algorithms; artificial intelligence ethics governance.

I. Introduction

With the development of Artificial Intelligence, more and more ethical problems are caused by Artificial Intelligence, and Artificial Intelligence Ethics are drawing more and more attention.

To solve Artificial Intelligence Ethics problems, this paper reviewed the development history of Artificial Intelligence Ethics, listed the main guidelines and principles of Artificial Intelligence Ethics, proposed the methods of Artificial Intelligence Ethics governance, discussed related algorithms to solve Artificial Intelligence Ethics problems.

Artificial intelligence ethics is not only a social problem, but also a philosophical problem. This paper has the viewpoint that Artificial intelligence ethics should be computed. This paper attempt to use mathematics and algorithms to solve Artificial intelligence ethics problems. In this paper, one Artificial Intelligence Ethics model will be proposed to solve Artificial intelligence ethics problems.

II. What is Artificial Intelligence Ethics

Artificial intelligence ethics is an academic hotspot. Artificial intelligence ethics mainly include the following aspects: (1) Whether or not Artificial Intelligence should own moral awareness. (2) Whether or not Artificial Intelligence should own the sense of responsibility. (3) Should Artificial Intelligence make moral and ethical judgments regarding decisions related to human life and safety. (4) If Artificial Intelligence can learn and create independently, should Artificial Intelligence own intellectual property rights. (5) Whether or not the application of Artificial Intelligence meets ethical and moral standards. For example, whether the weaponization of artificial intelligence is acceptable.

The debate in the academic community regarding the moral judgment of artificial intelligence requires distinguishing between two issues: first, the moral evaluation of artificial intelligence itself; Secondly, the evaluation of the good and evil consequences of the development and application of artificial intelligence. The key issue of moral judgment on artificial intelligence has not been resolved, that is, the issue of evaluating the good and evil of artificial intelligence itself has not been distinguished from the evaluation of the good and evil of the consequences of the development and application of artificial intelligence. The key to solving the latter problem still lies in humanity itself. However, in order to solve the previous problem, we cannot judge it based on the existing ethical and moral framework, but should critically reflect on traditional technological ethics.

Overall, there are three main positions and viewpoints in the academic community.

a) *The First Optimistic Stance*

Experts and scholars who hold this position believe that artificial intelligence is just a means and tool, and it does not matter whether it is good or bad. The key lies in the human beings who use it. They hold an optimistic attitude towards the future development prospects of artificial intelligence. Overall, the research and widespread application of artificial intelligence have more advantages than disadvantages for human development, and can generate huge economic and social benefits.

Generally speaking, the optimistic stance is mostly upheld by some artificial intelligence professionals who are related to the research and application of artificial intelligence, or consider their own interests, or by scientists who blindly worship science and technology. Its flaw lies in the one-sided and isolated view of the positive aspects of artificial intelligence, such as its ability to generate huge economic and social benefits, reconstruct almost all industries including finance, healthcare, education, transportation, etc., and promote overall changes in human lifestyles. They intentionally or unintentionally ignore or conceal the negative effects of artificial intelligence, such as the emergence of killer robots that will pose a security threat to humanity and the potential

Author α: Faculty of School of Information Technology & Management, University of International Business and Economics Beijing, China.
e-mail: wangshuxi@uibe.edu.cn
Author σ: Faculty of School of Humanities, Beijing University of Posts and Telecommunications Beijing, China.
e-mail: minmin.xia@163.com

degradation of human civilization caused by excessive reliance on artificial intelligence.

b) The Second Type of Neutral Stance

Experts and scholars who hold this position acknowledge that artificial intelligence itself has the potential to "do evil", and its research and application pose a potential threat to humanity and may bring serious consequences. However, for some reasons, they still strongly support the development of artificial intelligence technology.

Artificial intelligence is currently in its early stages of development, and its harm is far from strong enough, so there is no need to worry too much;

For example, Tom Austin, a global leading analyst in the field of artificial intelligence, stated that Hawking's warning that "the complete development of artificial intelligence will lead to the complete destruction of humanity" is "very foolish", citing the reason that "artificial intelligence is still very low-level".

"Artificial things cannot surpass humans";

This viewpoint stems from a certain religious sentiment, which states that "the Creator is always superior to what he has created," so there is no need to worry excessively. It is ironic that the scientists who should have had the most atheistic spirit seek intellectual resources from religious beliefs.

Human beings can set moral standards for artificial intelligence, but there has never been an effective argument that artificial intelligence will inevitably comply with the moral standards set by humans;

Some people believe that as long as moral education is provided to artificial intelligence, it can ensure that they are dedicated to goodness and serve humanity wholeheartedly. However, the question is how moral education can prevent artificial intelligence from developing in an unethical direction.

c) The Third Pessimistic Stance

Experts and scholars who hold this position believe that artificial intelligence is no longer a tool. It has a sense of life and learning ability, and has two moral possibilities of "doing evil": one is that the powerful power of artificial intelligence may trigger "human evil", and the other is that artificial intelligence itself has the ability to "do evil", and humans cannot cope with the "evil" of artificial intelligence, ultimately leading to nothingness and destruction. Therefore, they expressed concerns that artificial intelligence may lose control or harm humanity in the future. In early 2015, Stephen Hawking, Bill Gates, Elon Marks, and others signed an open letter calling for control over artificial intelligence development. Max believes that artificial intelligence can "summon demons", posing a greater threat to humanity than nuclear weapons. Hawking explicitly asserts that the complete development of artificial intelligence may lead to the extinction of humanity. People who hold a pessimistic stance believe that technology is not the path to human liberation. It is "not liberated from nature by controlling it, but rather a destruction of nature and humans themselves. The process of constantly killing living beings will ultimately lead to overall destruction."

In the application of Artificial Intelligence, balancing the development of technology with ethical and moral standards is an important issue. We need to consider technological innovation and development, while also paying attention to the interests and rights of society and humanity. For example, in the field of autonomous driving, it is necessary to consider vehicle safety and pedestrian safety, establish sound safety standards and regulations, and strengthen the supervision and management of artificial intelligence to avoid safety and ethical issues.

Dealing with ethical issues related to Artificial Intelligence requires multifaceted work and measures. Firstly, we need to strengthen research and discussion on ethical issues related to artificial intelligence, and establish ethical and moral standards for artificial intelligence. Secondly, it is necessary to strengthen the supervision and management of artificial intelligence, standardize its application and development, and prevent safety and ethical issues from arising. In addition, it is necessary to strengthen public participation and supervision, establish a sound feedback mechanism, and ensure the safety and ethics of Artificial Intelligence.

Artificial intelligence ethics must conform to human morality: not infringing on privacy, not harming humanity, not influencing the political situation, especially in the field of AI weapons, more caution should be exercised.

In short, the development and application of Artificial Intelligence will inevitably face ethical and moral challenges. We need to strengthen research and discussion on artificial intelligence, establish ethical and moral standards for artificial intelligence, and strengthen supervision and management of artificial intelligence to avoid safety and ethical issues. In the long run, only by achieving a balance between technological development and ethical standards can the sustainable development and application of artificial intelligence technology be achieved.

III. The Principles of Artificial Intelligence Ethics

Artificial intelligence ethics must ensure that AI does not make decisions that contain bias or discrimination, and establish mechanisms to interrogate AI to ensure that they comply with human ethical standards. Serving the interests of humanity and never harming humanity. The principles of Artificial Intelligence Ethics includes:

1. Developing artificial intelligence is for the common benefit of humanity.

2. Artificial intelligence should ensure fairness and be easy to understand.
3. Artificial intelligence should not be used to infringe on people's privacy.
4. All citizens have the right to receive education, enabling them to prosper and develop spiritually, emotionally, and economically alongside artificial intelligence.
5. Artificial intelligence should not be endowed with the autonomy to harm, destroy, or deceive humans.

Artificial intelligence is not without risks, and the formulation of these principles helps to mitigate risks. If there are methods to prevent the misuse of artificial intelligence technology, the public will trust AI more and better apply this technology.

AI will eliminate old jobs and create new ones. During the transition from old to new, the government must do a good job in vocational re-education to ensure that those who have been robbed of their jobs find new jobs.

IV. Specific Artificial Intelligence Ethical Implications

Artificial Intelligence ethics include so many aspects. The following are some specific Artificial Intelligence ethical implications.

Medical Field

With the rapid development of technology, artificial intelligence technology has made breakthrough applications in the medical field. The development of artificial intelligence technologies such as artificial intelligence assisted diagnosis, intelligent drug design, and artificial intelligence assisted surgery has brought unprecedented development opportunities to the medical industry. However, the development of artificial intelligence in the medical field has also brought ethics and challenges. How to ensure that the use of artificial intelligence in the medical field does not cause potential harm to human health has become an urgent issue that needs to be addressed.

The basic principles and applications of artificial intelligence in the medical field.

a) The basic principles of artificial intelligence in the medical field.

Artificial intelligence (AI) technology is the ability to simulate human intelligent activities and achieve certain tasks. Its core is based on algorithms and big data, using techniques such as deep learning and machine learning to enable computers to recognize, analyze, and process large amounts of data, thereby achieving the acquisition, understanding, and application of information.

Harmful effects: exploring the ethics and challenges of artificial intelligence in the medical field.

The application of artificial intelligence in the medical field mainly includes the following aspects:

1. *Auxiliary Diagnosis*

By analyzing a large amount of medical data, artificial intelligence can assist doctors in disease diagnosis and improve the accuracy of diagnosis. For example, artificial intelligence can recognize and analyze medical images through technologies such as depth learning, assist doctors in detection and localization, and reduce surgical risks.

2. *Intelligent Drug Design*

Artificial intelligence can assist scientists in drug development, improving drug efficacy and reducing drug side effects. By deeply mining a large amount of bioinformatics data, artificial intelligence can predict the molecular structure and properties of drugs, providing scientists with targeted directions for drug development.

3. *Artificial Intelligence Assisted Surgery*

Artificial intelligence can assist doctors in surgical simulation and planning, improving the safety and efficiency of surgery. For example, artificial intelligence can provide real-time feedback to doctors by simulating surgical scenarios, assisting them in performing precise operations during the surgical process and reducing surgical risks.

The Ethics and Challenges of Artificial Intelligence in the Medical Field.

b) Privacy Protection

The data in the medical field is highly sensitive and involves patient privacy information. The application of artificial intelligence in the medical field requires strict protection of patient privacy to avoid the leakage of patient personal information.

1. *Data Security*

In the medical field, artificial intelligence needs to process a large amount of data, including sensitive information such as patient medical records, images, genes, etc. The confidentiality and security of these data are crucial, and artificial intelligence enterprises need to take strict security measures to ensure that these data will not be leaked during transmission, storage, and use.

2. *Anonymity*

In the medical field, artificial intelligence needs to process a large amount of anonymous data, such as patient medical records, medication usage records, etc. Although these data do not contain personal identification information, they are still sensitive. Therefore, artificial intelligence enterprises need to adopt strict policies and measures to ensure the security and confidentiality of these anonymous data.

c) Moral Responsibility

The application of artificial intelligence in the medical field has a strong moral responsibility. The

development of artificial intelligence technology requires adherence to ethical standards and respect for human dignity and rights.

1. *Respect Individual Rights*

The application of artificial intelligence in the medical field may have an impact on the personal rights of patients, such as infringement of patient privacy and leakage of genetic information. Therefore, artificial intelligence enterprises need to respect the individual rights of patients, protect their dignity and privacy.

2. *Adhere to ethical standards*

The application of artificial intelligence in the medical field requires adherence to medical ethical standards, respect for medical ethics and professional ethics. For example, artificial intelligence technology needs to ensure that there is no misdiagnosis or missed diagnosis in the detection and positioning process, to avoid unnecessary harm to patients.

d) Publicity

The application of artificial intelligence in the medical field needs to ensure fairness and avoid unfair distribution of medical resources due to factors such as race and gender.

1. *Public Allocation of Medical Resources*

The application of artificial intelligence in the medical field needs to follow the principle of public allocation of medical resources, ensuring that everyone can access public and reasonable medical resources.

2. *Oppose Discrimination*

The application of artificial intelligence in the medical field needs to oppose discrimination and ensure that everyone can receive equal medical treatment.
Harmful effects: exploring the ethics and challenges of artificial intelligence in the medical field.

The use of artificial intelligence in the medical field has great potential and development space. However, in order to ensure that the use of artificial intelligence in the medical field does not cause potential harm to human health, it is necessary to comply with relevant ethical standards and legal regulations. Artificial intelligence enterprises need to shoulder moral responsibilities, protect the privacy and dignity of patients, ensure the allocation of public medical resources, and make positive contributions to human health and the development of the medical industry.

Field of Human Rights

The risks of basic rights include personal data and privacy protection, as well as non-discrimination. The use of artificial intelligence may affect the fundamental values of the European Union and lead to the infringement of fundamental rights, including freedom of speech, freedom of assembly, human dignity, non-discrimination based on gender, race or ethnicity, freedom of religious belief, or non-discrimination based on disability age, sexual orientation (applicable in certain fields), protection of personal data and private life, and the right to effective judicial remedies and fair trials, and consumer protection rights, etc. These risks may be due to flaws in the overall design of artificial intelligence systems (including human supervision), or may be due to possible biases not being corrected when using data (for example, the system only uses or primarily uses data from men for training, resulting in poor results related to women).

Prejudice and discrimination are inherent risks in any social or economic activity. Human decision-making cannot avoid errors and biases. However, when the same bias appears in artificial intelligence, it may have a greater impact, and without social governance mechanisms that control human behavior, it can affect and discriminate against many people. This situation also occurs when artificial intelligence systems are "learning" during operation.

Field of War

For a long time, discussions on the application of artificial intelligence in military have been limited to autonomous weapons and the ethical issues they bring. With the development of technology, attention should now be paid to the impact of artificial intelligence on security and other aspects of the military field.

Artificial intelligence is greatly changing the civilian sector, such as improving efficiency, reducing costs, and automating processes, and the military will also usher in an AI revolution.

At present, all countries must obtain human permission before using weapons, which is also in line with human values. However, what problems will occur when opponents deploy autonomous weapons without human permission needs to be urgently discussed.

There is reason to believe that even countries that have imposed some restrictions on artificial intelligence capabilities will encounter such opponents, which puts the countries that have imposed restrictions at a disadvantage. Therefore, countries must have a comprehensive understanding of what artificial intelligence can do.

Although autonomous weapons have attracted a lot of attention, most conversations about this technology are negative, leading people to overlook the positive applications of artificial intelligence in areas such as military protection and reducing civilian casualties.

The Advantages of Artificial Intelligence

Artificial intelligence has broad application potential in optimizing human-machine collaboration in fields such as command chain communication and logistics, as well as predicting opponent maneuvers. Numerous countries, including the Israeli Defense Force, are conducting corresponding research.

Military commanders will use artificial intelligence to solve the dilemmas of war. Artificial intelligence will enhance the decision-making ability of commanders by providing more accurate battlefield situational awareness and higher responsiveness, thanks to constantly updated sensor data.

Artificial intelligence technology will also help decision-makers and analysts cope with the impact of information overload, better organize and process ever-growing opponent data, and enable troops to make predictions about future events and outcomes, enabling them to better prepare for combat.

Better understanding of opponents is becoming one of the most promising application directions for artificial intelligence. Artificial intelligence will achieve faster and more real-time information collection, detection patterns, communication network drawing, and even better sensing of opponent morale by analyzing their language on social media and other platforms. These new AI features are equivalent to Intelligence Gathering 2.0.

This type of analysis can be extended to the military communication and social media activities of civilians in hostile countries, in order to better understand a country's willingness to war at any time, which is the most critical factor in human warfare and will have a huge impact on decision-makers in both civilian and military fields.

In the field of military logistics and maintenance, artificial intelligence can create revolutionary cost saving efficiency, which is why most military forces prioritize conducting research in this area. Logistics support may lead to the most fundamental changes in the military.

Artificial intelligence systems can also optimize the procurement process and achieve supply chain automation, predict the demand for maintenance equipment and order supplies, while minimizing costs. Artificial intelligence can also be used for personnel allocation, helping the military identify which soldiers are most suitable for which unit. Unlike other aspects of artificial intelligence, these applications are unlikely to raise any significant legal or ethical issues.

Artificial intelligence based technology can also enhance the capabilities of individual soldiers, which should not be seen as an unethical or dangerous way.

What is the Application of Artificial Intelligence in National Defense?

At the strategic level, artificial intelligence can enhance the capabilities of air defense systems. Emerging weapons, such as hypersonic missiles, are difficult to detect by existing defense systems due to their speed, while air defense systems that use artificial intelligence processing capabilities can detect and intercept such missiles.

In addition, in the field of information warfare, artificial intelligence has great potential in quickly verifying information or identifying opponents.

V. Artificial Intelligence Ethics Model and Algorithm

As an architecture system of autonomous intelligent agents, artificial intelligence's subjectivity or subject structure is somewhat similar to how we move certain functions of the human brain (or functions similar to the human brain) into machines. If the biological basis of human subjectivity is "neural", then the subjectivity of artificial intelligence can only be an imitation of human subjectivity. The scientific basis and manifestation of this imitation is the "algorithm". Setting aside the extent to which artificial intelligence agents are similar to human agents, researchers point out that the key element in making artificial intelligence products intelligent agents is the "moral algorithm" - an algorithm that teaches autonomous artificial intelligence devices to act responsibly, embedded in the algorithm system of artificial intelligence.

The subject mode of artificial intelligence faces significant ethical challenges on this issue. Taking autonomous robots in healthcare and the battlefield as an example, how should robots make decisions when facing the dilemma of life and death for human life? When inappropriate decisions lead to avoidable harm, whose responsibility is that? On this issue, although the subject mode of artificial intelligence highlights the importance of moral algorithms, its deeper and more important dependence is undoubtedly the "good law" established by humans for themselves.

Currently, there are roughly three algorithms that endow artificial intelligence with moral abilities.

One is to expand the moral logic through semantic networks, forming the concepts of obligation and permission;

The second is to establish association rules through knowledge graphs to detect moral judgment situations;

The third is to explore relevant relationships through cloud computing, evaluate or predict the consequences of actions.

Moral algorithms are algorithmic programs embedded in the algorithmic system that need to be improved. It itself is constantly changing and developing, rather than a specific existing thing, nor is it an ultimate assumption that can be achieved overnight or once and for all. It, as an artificial construction, is a "manual goodness" that leads to the "purpose goodness" and therefore depends on the human subject pattern. At this scale, algorithms can only promote the evolution of moral algorithms and their embedding in machines in a responsible manner by reflecting or following the "good law" of human subjectivity. This is the basic principle that

moral construction in the era of artificial intelligence should follow, that is, algorithms follow the "good law". In this principle, although the term "good law" is abstract and ambiguous, the scale of human subjectivity it represents may also cause controversy in specific content, but it clearly points to two moral forms on the human scale in form.

The first form of morality is dominated by common human subject patterns and involves all ethical issues that humans may bring when expanding artificial intelligence. Specifically, when people view artificial intelligence as a tool, its moral specificity and importance always call for the return of the moral responsibility of human subjects. This is a simple normative orientation, which means that humans should plan and embrace the advent of the artificial intelligence era in a responsible way. A prudent ethics suggests that the greatest threat that artificial intelligence may face is not from machines, but from humans or their intentions and actions. Considering that the algorithm that endows robots with moral abilities is essentially an algorithm that mimics human morality, how is it possible to present human morality in the form of algorithms in machines if humans cannot obtain clarity on moral issues? The problem paradoxically illustrates the moral construction of artificial intelligence implosion. It responds in some way to James Moore's demand for ethical intelligence subjects to have moral clarity, that is, as autonomy increases, artificial intelligence with autonomous moral abilities must be able to make clear rational decisions when facing moral dilemmas or conflicts of different moral principles. This demand for moral clarity, in turn, constructs or depicts the characteristics of "good law" at the human scale, forcing the human subject model to do everything possible to break out of various moral ambiguity zones that may lead to dark consequences (or even disasters).

The second form of morality is dominated by the "inter subject" mode of interaction between human subjects and artificial intelligence subjects, involving the moral construction of the dependent relationship between human subjects and intelligent subjects. This is a new field. Moral algorithms can only continuously correct biases or errors, further upgrade and improve in the repeated game between human machine interaction subjects. Autonomous robots may make decisions that we believe are morally wrong - such as being authorized not to provide pain relievers to patients, or biased artificial intelligence may self reinforce and harm society. However, this should not be a reason for humans to reject robots, but rather an opportunity for robots or artificial intelligence to improve and enhance their moral form. With the establishment of interdependence between human subjects and artificial intelligence subjects, autonomous robots with self decision-making ability, once they learn to develop decision-making algorithms from a moral perspective in their interaction with human subjects, can become a "good law" of interdependence between humans and machines to avoid harm.

VI. Conclusion and Future Work

This paper has the viewpoint that Artificial intelligence ethics should be computed. This paper attempt to use mathematics and algorithms to solve Artificial intelligence ethics problems. In this paper, one Artificial Intelligence Ethics model will be proposed to solve Artificial intelligence ethics problems. The future work will focus on related algorithms about Artificial intelligence ethics.

References Références Referencias

1. D. L. Medin, “Structural principles in categorization”, in T. J. Tighe & B. E. Shepp (eds.) Perception, Cognition and Development: Interaction Analyses, Hillsdale, N. J.: Lawrence Erlbaum, 1983, p.1469.
2. S. Shanker, Wittgenstein's Remarks on the Foundations of AI, London & New York: Routlege,1998, p.187.
3. Beauchamp, T. L. and Childress, J. F. 2012. Principles of Biomedical Ethics 8th. Oxford University Press, Oxford.
4. Biller-Andorno, N.; Aebi-Mueller, R.; (...); Sedlakova, J.2021. Ineffectiveness and unlikelihood of benefit: Dealing with the concept of futility in medicine. Swiss Academy of Medical SciencesBern.
5. Biller-Andorno, N; Ferrario, A; (...); Krauthammer, M. Mar 2022. Mar 2021 (Early Access). JOURNAL OF MEDICAL ETHICS 48 (3), pp.175-183.
6. Biller-Andorno, N and Biller, A. Oct 10 2019. NEW ENGLAND JOURNAL OF MEDICINE 381 (15), pp.1480-1485.
7. Giubilini, Alberto and Savulescu, Julian. 2018. Philosophy & technology 31 (2) , pp.169-188.
8. Hermann, H; Feuz, M; (...); Biller-Andorno, N. Jun 2020. Apr 2020 (Early Access). MEDICINE HEALTH CARE AND PHILOSOPHY 23 (2) , pp.253-259.
9. Loi, M; Ferrario, A and Viganò, E. Sep 2021. Oct 2020 (Early Access). ETHICS AND INFORMATION TECHNOLOGY 23 (3), pp.253-263.
10. Meier, LJ; Hein, A; (...); Buyx, A. Jul 3 2022. Mar 2022 (Early Access). AMERICAN JOURNAL OF BIOETHICS 22 (7) , pp.4-20.
11. van de Poel, I. Sep 2020 | Sep 2020 (Early Access). MINDS AND MACHINES 30 (3), pp.385-409.
12. Shaw, D; Trachsel, M and Elger, B. Jul 2018. BRITISH JOURNAL OF PSYCHIATRY. 213 (1), pp.393-395.
13. den Hartogh, G. Mar 2016. MEDICINE HEALTH CARE AND PHILOSOPHY. 19 (1) , pp.71-83.
14. Hermann, H; Trachsel, M and Biller-Andorno, N. Sep 2015. JOURNAL OF MEDICAL ETHICS. 41 (9) , pp.739-744.