

Oral Materialization and Ethical Governance of Science and Technology in Artificial Intelligence

Hui LI^{1,a,*}

¹*Department of Education Management, Huazhong Normal University, Wuhan, Hubei, China
a. ishoup82861@student.napavalley.edu*

**corresponding author*

Abstract: Contemporary people live in a reality full of intelligent technology construction, and many practical activities are regulated by technology. The chat content generation pre-training program artificial intelligence developed by the Open Artificial Intelligence Research Center (OpenAI) based on large model data training has aroused wide concern because of its generation and learning ability. It is necessary to admit that the current domestic media publicity seems to focus more on the convenience that big data technology may bring to human society. But not much about its own limitations. In the few discussions on the negative effects of big data technology, more concerns are raised about the ethical risks that may be caused by the abuse of relevant technology, such as the threat of "data greed" to personal privacy, and the decision-making errors that may be caused by the superstition of business decision-makers and government heads about "digital dictatorship", etc. From the perspective of philosophy of information technology and philosophy of cognitive science, this paper reviews the gains and losses of big data technology itself in terms of philosophical thought premise and path strategy. Starting from the theory of "moral materialization", this paper analyzes the ethical value load and risk of artificial intelligence from the perspectives of moral subject status, value position presupposition, operation intervention and operation result effect. And from the moral materialization theory of "top-down" approach and "bottom-up" approach to the ethical governance of modern scientific and technological intelligent artifacts.

Keywords: Artificial intelligence, moral objectification, Cyborg, ethics of technology

1. Introduction

As a representative of generative artificial intelligence, thanks to the technological progress in natural language processing, the combination of algorithms, computing power and data makes it possible to open understanding and generate coherent text, and theoretical algorithms begin to move towards practical application. At the same time, some new ethical issues also appear. This paper analyzes the "materiality" of artificial intelligence in cyberspace and analyzes the debate on the ethical value of science and technology caused by artificial intelligence from the approach of internalism.

2. The Appearance of Artificial Intelligence Theory

GPT is short for "generative pre-training" model. OpenAI, an American artificial intelligence company, has released a large-scale language model called Chat Generative Pre-trained Transformer,

which is a language model based on natural language processing and deep learning technologies. With the ability of natural language processing and machine learning, artificial intelligence technology based on natural language processing can interact with humans and form an intelligent cognitive understanding model [1]. In terms of the composition of artificial intelligence, some scholars believe that artificial intelligence is the result of the coordination and penetration of algorithms, computing power and big data. In terms of algorithms, the text generation deep learning model trained by artificial intelligence based on available data on the Internet uses the training method of "reinforcement learning from human feedback (RLHF)" to improve the model's ability to evaluate the generated answers. In terms of force, the operation and training of artificial intelligence cannot be separated from the support of strong computing power, which is mainly provided by the supercomputing infrastructure of Microsoft cloud computing service Azure AI. In terms of data, artificial intelligence has the support of more than hundreds of millions of words of human language data, showing strong language generation ability, context learning ability and world knowledge ability [2].

As a representative of generative AI model, artificial intelligence is different from Analytical AI model in that it cannot only provide human-computer interaction and dialogue in a variety of complex contexts such as emotional companionship, but also has preliminary logical thinking ability. It can realize information retrieval, text writing, planning, generating new content different from learning samples and even programming.

From the difference between the original intention of research and development and the training method, artificial intelligence has the following three characteristics: The first is interactivity. In the process of chatting and interacting with people, artificial intelligence can carry out continuous dialogue with users on some topics in natural language based on People's Daily language habits and usage expressions, constantly accept user feedback information and adjust reaction behavior in time. At the same time, when users have doubts in dialogue, artificial intelligence can actively issue questions and request for more information. The second is a powerful "memory" during conversation. Artificial intelligence can store massive data resources, and to a certain extent, master the logical thinking chain reasoning ability, comparative analysis of pictures and text content, integration of fragmented information in digital space, and response to different contexts. When users communicate with artificial intelligence again, they can reproduce the previous interaction between the two sides [3]. Based on this record, long-term continuous dialogue is formed, so as to flexibly respond to human-machine real-time interaction and improve the efficiency of human-machine collaboration. The third is humanization. Artificial intelligence is based on a natural language generative model trained by mass text. Its training "natural language processing" (NLP) model initially gets rid of the stiff sentences and the characteristics of machine talk existing in the language system of the previous generation of intelligent robots, and has an anthropomorphic style, enabling a flexibility with various speech techniques. It can provide faster, more efficient and more humanized language text services. Automatic [4] content generation through the learning of historical text data can be considered as a kind of "creation" of artificial intelligence, but in essence, this "creativity" is the "quantitative change" of data analysis supported by super computing power.

2.1. The Embodiment of Artificial Intelligence in the Cyborg Sense

Since Pythagoras' "number" and Parmenides' "existence" in ancient Greece, people have gradually divided their exploration of the phenomenal world and the essential world. In the 17th century, Descartes' "I think, therefore I am", guided by the dualism of mind and body, formally established the thinking mode of subject-object dualism, and the moral and ethical system based on the dualism of subject-object thought mode gradually rigid and clear-cut the judgment of good and evil, right and wrong. "Cyborg" is a combination of the two words cybernetics and organism, its appearance shows

the combination of biological organism and electronic machine, the relationship between subject and object is not "black and white", the boundary between subject and object is gradually weakening. The development of modern technology has blurred the traditional boundary between the body and technology, for example, Internet communication technology has expanded the way of communication between human beings and the world, genetic engineering and mechanical bionics technology have made the choice of human body functions increasingly rich, and the ethical issues of technology have gradually exceeded the rigid dual thinking mode of subject and object.

Merleau Ponty explains "embodied" by analyzing the concept of "body". In his view, "the body is a medium that exists in the world, and to have a body is to intervene in a certain environment, participate in some plans and continue to be in it" [5]. The concept of the body is the subject of the unity of the body and the soul, and the body itself is the place where the meaning of the mutual penetration between the subject and the object occurs, and the bearer of the interweaving of the subject and the object. Merleau Ponty used "meat" to refer to the concept of the body in "Visible and Invisible", "meat has the dual properties of material and spiritual" [6]. The correlation and interaction between the body and external objects, on the one hand, the body constructs a "cultural world" [7] in the body in the process of using objects and transmitting information and exerts an influence on people; On the other hand, external objects can also be shaped by their association with the body and given a comprehensive meaning. As an association of the body, artificial intelligence "educates" the body through the dialogue paradigm, breaking through the traditional human experience cognition mode of integration of knowledge and action. Artificial intelligence has epoch-making significance. Different from the effect of physical technology on objective physiological entities, artificial intelligence casts spiritual content and knowledge into the category of "meat" in the virtual cyberspace. And then subvert and even reconstruct the ethical norms of human existence.

2.2. The Expression of the Internalistic Technology Ethics of Artificial Intelligence

Contemporary people live in a reality full of intelligent technology construction, and many practical activities are regulated by technology. People gradually feel that technology cannot be simply classified as a neutral tool. As an artifact, the design, use and results after use of tools contain certain ethical values. Is AI Value Neutral or Value loaded? This paper believes that artificial intelligence is loaded with human moral value. In internalist technical ethics, unlike the externalist approach, which regards ethics and morality as a normative force outside technical activities and conducts ethical criticism and reflection on the consequences of technology, it "hides rites and tools", excavates the moral dynamic role of technology, and considers how to realize and express its moral dynamic role in reality [8]. Under the concept of internalism technology ethics, "moral materialization" has become a new research direction of western technology ethics in contemporary development. The basic connotation of "moral materialization" refers to embedding moral norms in technology to promote people's behavior to the good, and technology is no longer a neutral tool, but a normative object that can be embedded in moral regulations, and then realize moral purposes through the design and manufacture of technology. Different from the more independent appliances or machines as the representative of the traditional technology of physical characteristics, contemporary science and technology continues to innovate, with the development of battery technology, chip technology, network technology, technical products are no longer independent individuals to achieve a single function, through the extensive link of the network to form a complex technical system, hidden behind the technology is embedded in the algorithm logic [9]. Artificial intelligence, as a representative of generative AI, is jointly promoted by the progress of algorithms, computing power and data, among which the breakthrough at the algorithm level is the most critical factor [10]. In essence, algorithms, computing power and data are all "technical artifacts" created by humans, and artificial intelligence is a technology in cyberspace, a virtual reality, carrying human ethical values.

3. The Ethical Value Load and Risk of Artificial Intelligence

The cognitive experience of the moral subject has changed from the man with free will to the living body, and then to the research category of technical artifact and intelligent artifact. Kant regards morality as the relationship between rational parties, so human beings with rational thinking ability have the status of moral subject; Ralston, the father of environmental ethics, argued that the inherent value of nature is reasonable, and all living beings contained in the natural attributes of nature have the status of moral subject. Viebeck proposed "moral materialization", which recognized the moral subject status of technical artifacts and embodied ethical value in the structure and function of artifacts [11]. But does artificial intelligence based on algorithms, computing power and data have a moral subject status as an intelligent artifact in the virtual cyberspace?

Whether artificial intelligence has the status of moral subject involves the definition of the standard of moral subject. For Descartes, it is absurd to think of a physical machine as a moral subject, because a machine with an extension of time and space is a substance, quite different from the conscious mind, and does not possess the reason of the mind. In terms of whether it has free will and rationality, artificial intelligence, as an artificial system, does not have the moral subject status of human beings. However, from the perspective of autonomous behavior and creation generation, artificial intelligence can learn from massive text data and automatically "create", which is in line with the view of artificial intelligence expert Foster on existing intelligent machines. Artificial intelligence not only acts as a passive mediator of ethical rules, but also interacts with other agents in the existing moral ecology, thus influencing existing ethical rules. Therefore, artificial intelligence can be considered as having the status of a moral agent. Although this "creativity" of artificial intelligence is essentially the super computing power to analyze the data under the algorithm rules, at the level of quantitative change, whether it will produce a certain degree of qualitative change on the basis of quantitative change needs to be further promoted by science and technology. If artificial intelligence is recognized as having the status of a moral subject, capable of exporting ethical values autonomously and independently of human beings, and performing morally related behaviors, then human beings should take precautions against its good or evil value orientation, its great power and the harm it may bring.

Human practice cannot be separated from value presupposition. In artificial technology, ethical concepts are intentionally or unintentionally embedded in technical products. Engels once made the following judgment on morality: "Morality is always the morality of class; Or it justifies the rule and interests of the ruling class " [12] who, for some purpose, will place certain values and interests above others. The general artificial intelligence represented by artificial intelligence is designed to serve widely recognized ethical concepts and the interests of all mankind, and has a shared nature, but in fact, the interests of individual countries or individual organizations often require the design and operation of artificial intelligence to set ethical value principles in advance.

First, the presupposition in the value position is easy to form the "technological hegemony" of interest groups. The monopoly of artificial intelligence technology on a few interest groups will lead to serious class solidification and technical barriers, [13]and the "technological hegemony" trapped in it will threaten the fairness of mankind, strengthen the monopoly of the interest groups of science and technology giants in the market, and trigger a new round of "digital divide". Moreover, when the designer does not have the correct orientation of moral ethics or blindly obey his own ethical bias, the designer will not be able to achieve the goal. Artificial intelligence technology products such as artificial intelligence will also reflect their preference for value, and this "bias will be strengthened and amplified with the accumulation of data and the iteration of algorithms", thus causing confusion in science and technology ethics.

Second, the presupposition in the value position may form "ideological bullying" in cyberspace. For people, "conceptual things are nothing more than material things that have been transferred into

people's minds and transformed in people's minds". The data feeding between designers and users of knowledge and content generation by artificial intelligence is the result of algorithm and computing training, and does not have subject consciousness, so its value position can be shaped. In reality, the data used to train artificial intelligence mainly comes from the culture established in the Western value system and ideology, and has a serious ideological orientation. After absorbing these data, artificial intelligence is easy to generate texts containing ideological bias [14]. Artificial intelligence relies on learning algorithms and data to obtain information and form a natural language processing model for judgment. "Through a series of procedures such as pre-training, manual supervision of labeled data, expert adjustment of models, and user interaction, content is generated and knowledge is constructed, which can be exported to users around the world without the full control of the government and the market", the value position of artificial intelligence presupposes [15]. It is easy to form "ideological bullying" in the field of cyberspace on a global scale, and it is necessary to remain vigilant about artificial intelligence and avoid falling into the trap of ideological penetration of technology packaging.

In the process of operation, artificial intelligence may violate people's equal participation rights in cyberspace and raise the threshold of public participation in social governance in disguise. On the one hand, the value preference in the artificial intelligence algorithm is easy to ignore the equality of opportunity of the digital vulnerable groups, the digital vulnerable groups leave a small amount of data in the cyber space, and the digital vulnerable groups are generally difficult to effectively express their value demands, so that it is difficult to enter the vision of democratic decision-making, aggravating the digital divide. The participation of artificial intelligence leads to the appearance of machine will in people's action plans and decision-making methods. It is easy to dilute individual value judgment and value choice, resulting in the disorder of mainstream social values. In addition, artificial intelligence can only make decisions based on existing training database data in the process of social governance, which is difficult to reveal causal logic, unable to produce purpose and meaning at the level of autonomous consciousness, unable to deal with newly emerged and uncertain social factors, and prone to social disorder. Then it causes the confusion of science and technology ethics. On the other hand, under the monopolistic domination of artificial intelligence technology, citizens' data information is controlled, and the public's individual judgment and decision-making are easily affected by algorithmic bias in artificial intelligence. Artificial intelligence becomes an instrumental extension of the embodied "body" in cyberspace, as an instrumental object used by it to achieve specific purposes, and bears certain ethical values. It even induces people's dependence on artificial intelligence, diluting the ability to think independently, leading to the degeneration of independent thinking ability, and the public becomes a "puppet" who is represented by the voice, leading to the failure of the value of public decision-making, thereby causing social injustice.

The operational results of artificial intelligence have ethical effects. When artificial intelligence is designed, it will have a certain degree of relative independence, and even have some unintended ethical consequences.

First, algorithmic discrimination causes ethical bias. The information data of discriminatory values existing in real space is put into the corpus, and the training materials of generative AI come from the human information data, so it may inherit the discriminatory factors contained in the human information data. Even if training data is not biased, it is possible for AI to create bias through deep learning. Artificial intelligence is programmed during the design process to refuse to answer certain sensitive topics. However, under the guidance of users, these models will still say a lot of inappropriate statements [16]. Sociologist Pariser pointed out that algorithms infer people's likes and dislikes through the analysis of big data, filter heterogeneous information and push notification information, and create a relatively closed "information cocoon".[17] Artificial intelligence builds an "information island" in the ethical values bound by algorithms, which will cause ethical values to

become deeply rooted in the process of constant repeat and amplification. People's herd psychology will change to artificial intelligence psychology, and the long-term indoctrination of homogenous content in the "information cocoon" will penetrate their emotional cognition, ethical behavior and overall behavior, which will have an important impact on people's value beliefs, attitudes and behaviors, and cause certain ethical biases.

The second is academic ethics and privacy and security disputes. Based on the generative characteristics of artificial intelligence, it can generate extremely convincing and complex texts, but the text source is essentially the existing knowledge results, if the user uses artificial intelligence to produce works that do not belong to their own, then does it constitute plagiarism of the existing knowledge results? As soon as artificial intelligence appeared, a series of academic ethical value challenges were generated, which caused people to think about the technical ethical issues related to artificial intelligence. From the perspective of knowledge reliability, the production of effective and reliable information content by artificial intelligence depends on the model to provide a large amount of proven knowledge and common sense, but in the face of new problems and new situations of The Times, artificial intelligence answers will inevitably reduce the reliability of the generated information, and sometimes even provide misleading content and false information. At the same time, AI needs a lot of data in training, which may contain some users' private information. A data leak in a language model used by AI could expose the model, and if "interested parties" try to extract relevant sensitive information from the language model, it could compromise personal privacy and organization security. Therefore, the security and reliability of the information content provided by artificial intelligence should arouse people's extensive attention, and should include the protection of users' privacy information and personal data security, and prevent personal data from being illegally used.

4. The Enlightenment of Scientific and Technological Ethical Governance of Artificial Intelligence

Indeed, the transformation of artificial intelligence brings a series of opportunities, such as in the transformation of production mode, the generation of systematic training plans for workers, improve workers' theoretical knowledge and practical skills, improve workers' productivity, and improve production tools; In the innovation of educational form, we should provide "teaching according to students' aptitude" type service to promote the realization of lifelong education. In the way of life, the reconstruction of human material and spiritual life. However, as the latest generation of intelligent artificial objects, artificial intelligence is embedded in the value position when it is designed and generated, and is also affected by the value concept in the actual operation, and breeds new moral and ethical problems in the operation result effect. In cyberspace, artificial intelligence, as a constitutive participant of moral behavior, is an instrumental extension of the embodied "body", an instrumental object used to achieve specific purposes, with embedded ethical and moral values, reflecting the deeply technological situation of human beings [18]. Therefore, we cannot wait and see, let it go, in dealing with the ethical problems existing in the moral materialization of artificial intelligence, we can refer to the "top-down" normative path and the "bottom-up" normative path.

4.1. Top-down Normative Path

When Wendell Wallach and Colin Allen talk about embedded morality in machines in their book *The Moral Machine: How to Make Robots Know Right from Wrong*, the basic idea behind the "top-down" approach is: "If a moral principle or norm can be clearly stated, then a moral action becomes a matter of compliance with the norm, and all the artificial moral actor needs to do is to calculate whether its action is permitted by the rule," [19] that is, to make a clear and accurate statement of the ethical

principle becomes a moral norm followed by the machine. In the past, ethical theoretical research and technical practice were often regarded as the work of ethicists and engineers respectively. Even if the professional ethics of scientific and technological personnel was emphasized, it was often considered as the operational principle that should be paid special attention to outside scientific and technological activities. Moral materialization emphasized that engineers should embed corresponding moral and ethical values in scientific and technological practice. In the process of designing and operating scientific and technological products, technical engineers consciously or unconsciously practice ethical norms in practice.

The "top-down" normative path can give very reliable results when the situations encountered do not exceed the scope of the scenarios preset by the algorithm. However, in the process of practical operation, there are problems such as "rule regression" and huge calculation amount caused by the inexhaustible moral and ethical value standards, and the help of higher moral standards in the conflict of moral standards. However, the "top-down" normative approach still has reference value in addressing ethical issues in AI. On the one hand, engineers should also pay attention to cultural interpretation while following the engineering ethical guidance of being accountable to employers and ensuring that products can effectively perform the functions expected by employers. From the perspective of "regulatory analysis" of technology, engineers should be conscious of the cross-cultural use of technology, understand that specific technology will present different ways of use in different cultural situations, pay attention to the ethical and moral rules between different cultures as much as possible, and avoid the conflict of ethical values in the cross-cultural use of artificial intelligence. On the other hand, engineers should pay attention to the value sensitivity of things, in the process of designing technology products and practical operation, to maintain a regulatory imagination beyond the specific function of the technology, so as to be able to predict applications beyond their functional Settings.

4.2. The Bottom-up Normative Path

The "bottom-up" normative path and the "top-down" normative path are two approaches proposed by Wallach and Allen when the "Moral Turing Test" focuses on the moral embedment in machine design. Both approaches emphasize the construction of the moral agent model and adhere to the moral "good" behavior standards. In the application of generative artificial intelligence such as artificial intelligence, "through the self-organization development and evolution of machine learning and complex adaptive systems, robots can generate general ethical principles from specific ethical situations" [20]. The "bottom-up" normative path is a research approach opposite to the "top-down" normative path. The difference is that the "top-down approach" focuses on the moral construction based on the theory of universal ethics, and emphasizes the input of ethical value criteria in advance in artificial intelligence, while the "bottom-up approach" obtains morality by taking standardized value as the implicit content of the machine, and does not input a large number of ethical rules in advance, but is based on the creation of an algorithm. The algorithm simulates the formation process of human morality, explores the moral reasoning mode of self-judgment through continuous data training, and thus obtains the ability of moral judgment. However, as mentioned above, algorithms intentionally or undoubtedly carry certain value judgments to some extent, and there is "algorithm bias"; Because of the difference of language and culture, database will also cause different values. In the "bottom-up" normative path, data is trained based on algorithms, and this value difference bias will be strengthened and amplified with the accumulation of data and the iteration of algorithms.

5. Conclusions

The two approaches of "top-down" and "bottom-up" have their own advantages and disadvantages in dealing with the ethical problems existing in the process of moral materialization of artificial intelligence. By combining the two approaches, we not only pay attention to the input of moral and ethical differences in different cultures, but also pay attention to the update of algorithm models, so as to bridge the differences between the two paradigms. Avoid, as much as possible, the ethical issues that arise in the design, operation and production of AI results. At the same time, we must clearly realize that artificial intelligence, as a generative artificial intelligence, its "creativity" comes from the combination of data provided by people, computing power and algorithms, and tools can never replace humans. We should bear in mind the concept of inclusive and prudent governance, and face the complex ethical challenges that artificial intelligence may bring.

References

- [1] CARAYONP,HOONAKKERP,HUNDTAS,etal. *Application of human factors to improve usability of clinical decision support for diagnostic decision-making:a scenario-based simulation study*[J].*BMJ Quality& Safety*, 2019,29(4):329-340. (in Chinese)
- [2] Zhu Guanghui, Wang Xiwen. *Operation mode, Key Technology and Future Prospect of Artificial Intelligence* [J]. *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)*, 2019,44(04):113-122. (in Chinese)
- [3] CHATTERJEEJ, DETHLEFSN. *This new conversational AI model can be your friend, philosopher, and guide and even your worst enemy*[J]. *Patterns (New York)*, 2023, 4(1): 100676.
- [4] Ling Xiaoxiong. *Cold thinking on science and technology ethics and academic Ethics after the explosion of artificial intelligence* [J]. *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)* 2023(04):123.
- [5] Maurice Merleau Ponty. *Phenomenology of perception* [M]. Translated by Jiang Zhihui. Beijing: The Commercial Press, 2001:116.
- [6] YAO Dazhi. *Modern Western Philosophy* [M]. Beijing: China Social Sciences Press, 2015:24.
- [7] Liu Zheng. *From "Body" to "Flesh" : A Discussion on Merleau-Ponty's Body-involved ethics* [J]. *Journal of Social Sciences, South China Normal University*.2016.
- [8] Zhang Wei. Zang Li Shu: *The Chinese approach of internalist technology ethics* [J]. *Journal of Dalian University of Technology (Social Science Edition)*, 2018(03) : 116.
- [9] Wang Xiaowei. *"Moral Materialization" and ethical governance of Modern Science and Technology* [J]. *Zhejiang Social Sciences*,2023(01) : 119.
- [10] Chen Yongwei. *Beyond Artificial Intelligence: Opportunities, Risks and Challenges of generative AI* [J]. *Journal of Shandong University (Philosophy and Social Sciences Edition)*,2023(3):127-143. (in Chinese.
- [11] Yan Kunru. *Do AI machines have the status of moral agents?* [J]. *Research in Dialectics of Nature*, 2019(05):47
- [12] *Compilation and Translation Bureau of Marx, Engels, Lenin and Stalin of the CPC Central Committee. The Complete Works of Marx and Engels (Volume 26)* [M], Beijing: People's Publishing House, 2014 edition :100.
- [13] Zhao Tingyang. *What philosophical questions does artificial intelligence raise?* [J]. *Cultural Aspects*,2020(1).
- [14] LAN Jiang. *Generative Artificial Intelligence and the Historical Mission of Humanities and Social Sciences: Starting from the revolution of artificial intelligence* [J]. *Ideological and Theoretical Education*,2023(4):12-1.
- [15] Wang Peinan, Deng Shiqing. *Content, algorithms, and Knowledge Power: Risks and responses to artificial Intelligence from the perspective of international communication* [J]. *International Communication*, 2023(04):39.
- [16] Chen Yongwei. *Beyond Artificial Intelligence: Opportunities, Risks and Challenges of Generative AI* [J]. *Journal of Shandong University (Philosophy and Social Sciences Edition)* 2023 (3) :128.
- [17] Eli Pariser. *Filter Bubble: The Internet's Secret manipulation of Us* [M]. Fang Shishi, Yang Yuan, trans. Beijing: China Renmin University Press, 2020:10.
- [18] Carlini N, Tramer F, Wallace E, et al. *Extracting Training Data from Large Language Models*, *USENIX Security Symposium*, 2021 (6) : 223.
- [19] Wendell Wallach, Colin Allen. *Moral machines: How to make robots know Right from wrong* [M]. Wang Xiaohong, trans. Beijing: Peking University Press, 2017.
- [20] Duan Weiwen. *The approach and connotation of robot ethics* [J]. *Science and Society*, 2015(2):35.