

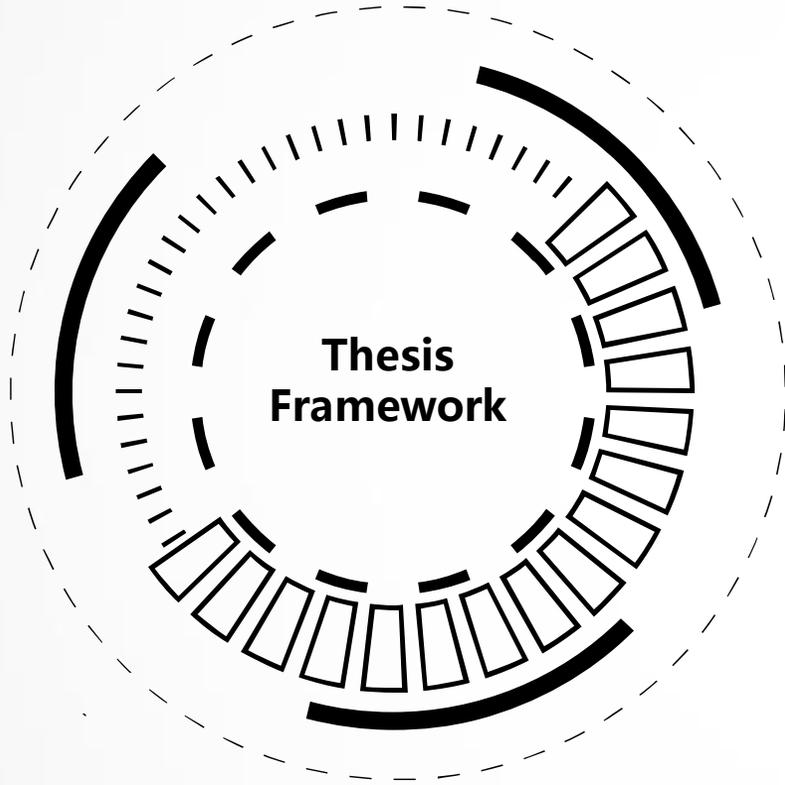
# Ethical Governance of Artificial Intelligence ( AI ) in China Based on Responsible Innovation

Chen Fan, Wang Jian and Chen Jia

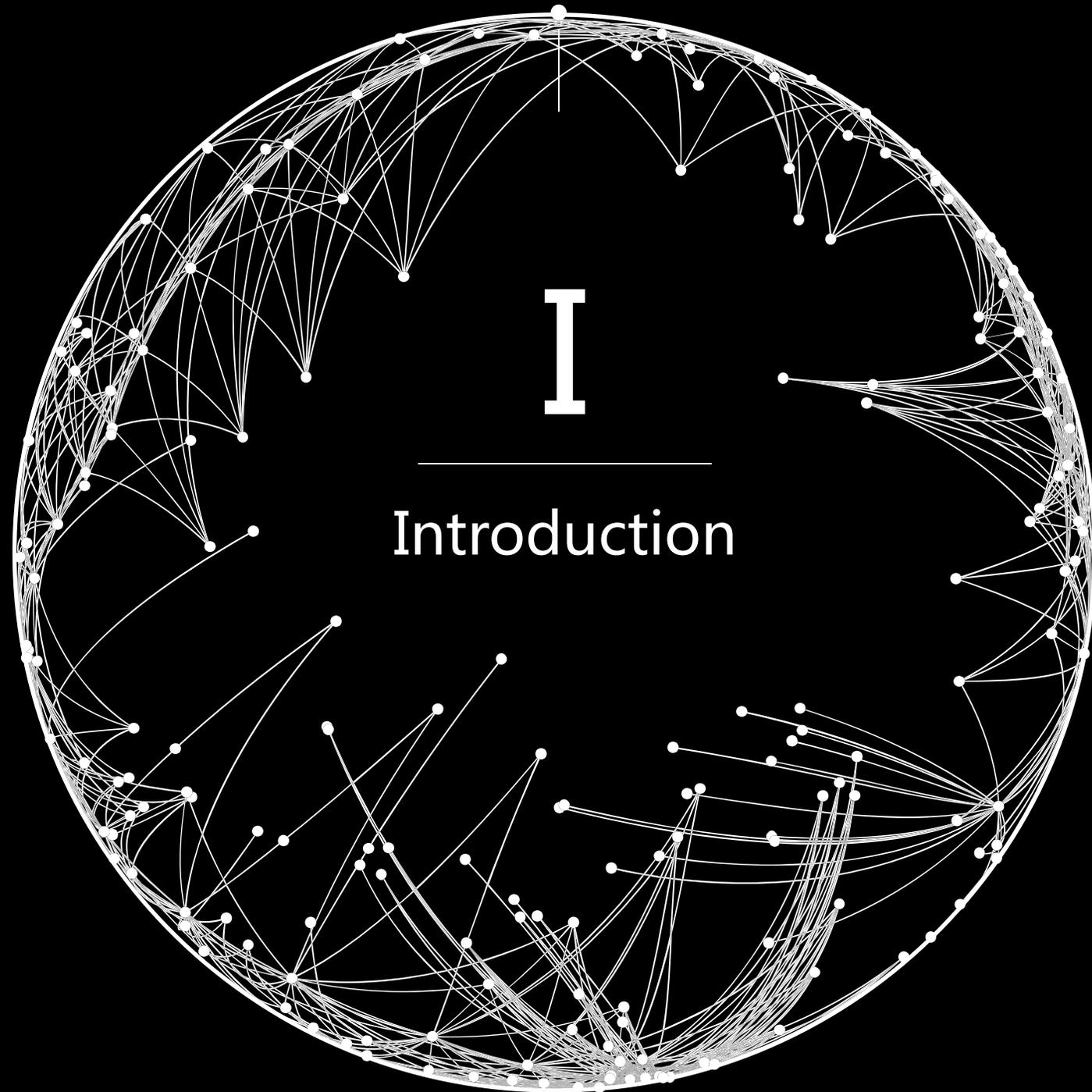
Research Center for Philosophy of Science and Technology

Northeastern University, China

5/12/2019



- I Introduction
- II Collingridge dilemma of Artificial Intelligence ( AI ) governance and necessity of ethical governance
- III Technology domination by morality ( 以道驭术 ) : Chinese tradition of Artificial Intelligence ( AI ) ethical governance
- IV Responsible innovation: action framework of Artificial Intelligence ( AI ) ethical governance



I

Introduction

# I

---



Alan Turing proposed 'thinking machine' in the paper 'can machines think?' in 1950. Artificial Intelligence ( AI ) has experienced a history of more than 60 years since the concept of 'Artificial Intelligence(AI)' was proposed during Dartmouth Conference in 1956.

Since Artificial Intelligence ( AI ) technology brings infinite possibility development to the future of human beings, it becomes a new competition hot-spot in various countries, own Artificial Intelligence ( AI ) development strategies are formulated one by one in the United States, Europe countries, Russia, Japan, South Korea, India and China since the 21st century. The United States is the first country to develop Artificial Intelligence ( AI ) to the national strategic level all over the world, and Artificial Intelligence ( AI ) strategic plan is regarded as new Apollo moon landing program in the United States.

## • Recent events on AI development for main counties :

—— The UK formulated '*Development Strategy 2020*', and the EU launched the largest civil robot R&D program all over the world in 2014.

——Japan formulated '*Robot Strategy of Japan: Vision, Strategy and Action Plan*' in 2015.

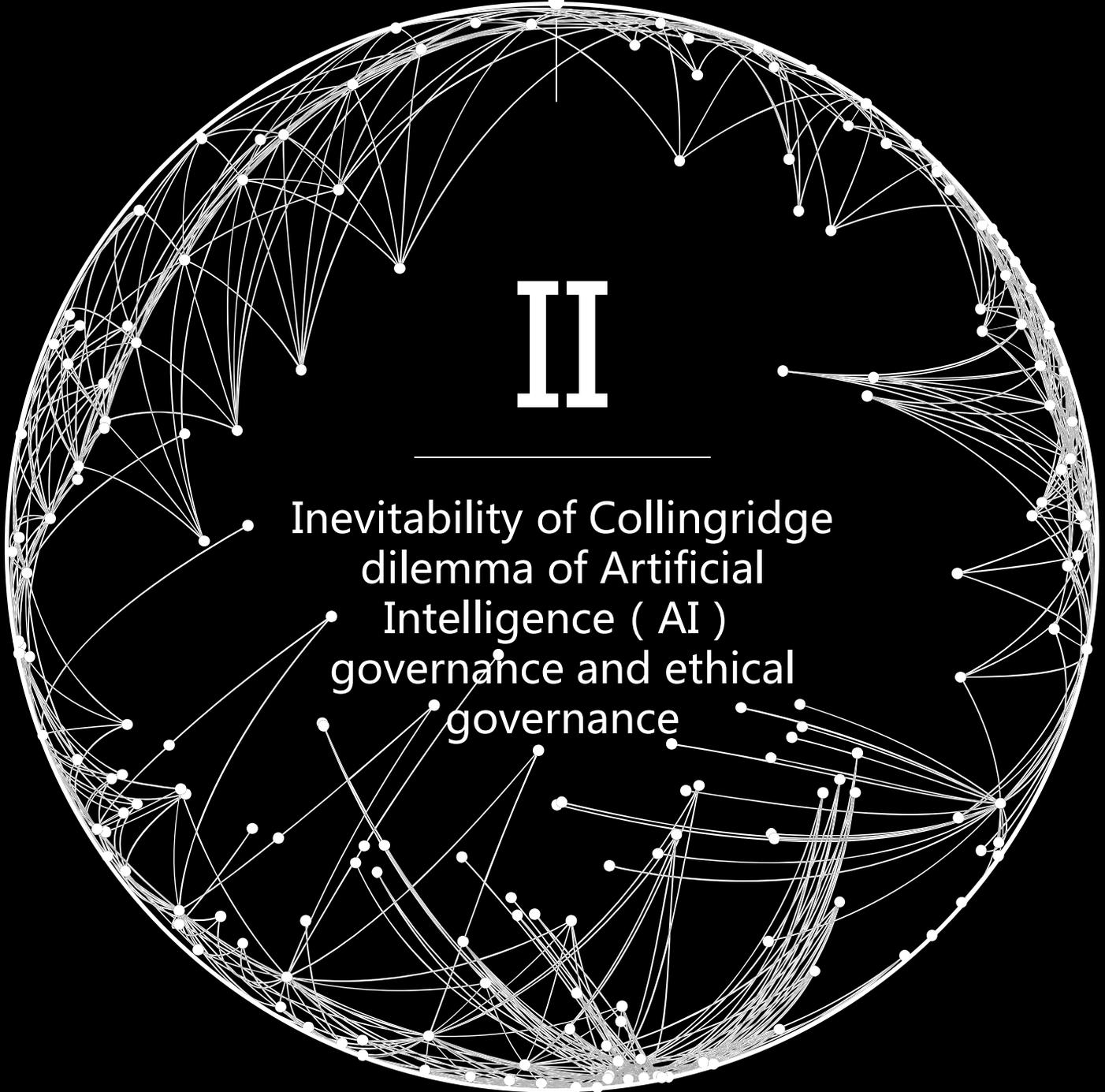
——The U.S. Government issued White House report 'preparation for future artificial intelligence' in 2016, including encouragement of collaborative innovation, protection of the public interest, development of regulatory measures for Artificial Intelligence ( AI ) technologies, thereby ensuring that technology applications are fair, secure, and controllable.

——'*Development Plan for The New Generation of Artificial Intelligence*' was released by China in 2017. Laws, regulations and ethical norms to promote Artificial Intelligence ( AI ) were formulated.

——25 EU member states signed 'Declaration on Cooperation in Artificial Intelligence' in 2018. Member states were encouraged to cooperate in the three aspects of jointly promoting innovation, improving skills and legal assistance, etc. 14 articles of consensus are reached, such as Artificial Intelligence ( AI ) R&D and application, security and accountability, etc.

---

It is undoubtedly that human governance actions are started in the face of Artificial Intelligence ( AI ) with fortune or misfortune. However, human beings also fall into governance Collingridge dilemma. How to ensure that governance belongs to 'good governance' rather than 'evil governance' (containment) and how to get rid of the Collingridge dilemma are common concerns of the government, industry and academia.



# II

---

Inevitability of Collingridge  
dilemma of Artificial  
Intelligence ( AI )  
governance and ethical  
governance

## II

---

The so-called Collingridge dilemma refers that it is almost impossible control a technology. When it can be controlled at the technology development stage, people are not able to fully understand its harmful social consequences and control it, however control cost becomes expensive and lagged when these social consequences appear.

The risks can not be completely determined for Artificial Intelligence ( AI ) at early research and development stage of artificial intelligence. An overly strong governance will make Artificial Intelligence ( AI ) in the development stage suffer from 'growth stagnation', perhaps human beings may lose the most effective ways to solve resources, poverty and diseases. It can not be governed due to too expensive governance cost until the mature development of Artificial Intelligence ( AI ) and even the appearance of 'singularity'.

Therefore, it is necessary to apply the value judgment of goodness to all stages and links of Artificial Intelligence ( AI ) from research and development, experiment, promotion and use in a responsible governance way in order to avoid the control Collingridge dilemma so as to guide Artificial Intelligence ( AI ) to develop in the direction of improving the well-being of human community instead of the reverse direction.

## II

---

Gong Ke (龚克, *Doctor of Engineering, professor, doctoral supervisor, foreign academician of Russian Academy of Aerospace Sciences*) proposed the follows in the World Federation of Engineering Organizations WFEO: 'Artificial Intelligence ( AI ) development faces serious challenges in terms of technical level and ethical governance, Artificial Intelligence ( AI ) technology should be more available, more credible, more reliable and more efficient at the same time. Therefore, international cooperation in technology innovation is required on the one hand, a consensus on ethical standards should be achieved with global and interdisciplinary dialogue in all countries, thereby transforming the consensus into policies, rules, laws and technology standards, and ensuring that the ethical requirements can be embedded into algorithm of Artificial Intelligence ( AI ) or intelligent systems, and the process is testable and verifiable.



In short, good Artificial Intelligence ( AI ) governance should be ethical governance of artificial intelligence.

## II

---

Artificial Intelligence ( AI ) ethical governance refers that various technical subjects and stakeholders of technology activities should negotiate and achieve 'moral consensus' in the activities of artificial intelligence ( AI ) technology, they should responsibly penetrate moral view, ethics and moral principles to all stages and links of the Artificial Intelligence ( AI ) technology activities, Artificial Intelligence ( AI ) technology can be guided and constrained to enhance human joint well-being.

## II

---

The ethical governance of Artificial Intelligence ( AI ) is the most suitable governance mode of Artificial Intelligence ( AI ) and the rational choice to avoid Collingridge dilemma.

White House released a report on Artificial Intelligence ( AI ) in 2016, and it was required that both Artificial Intelligence ( AI ) practitioners and students should receive ethics training.

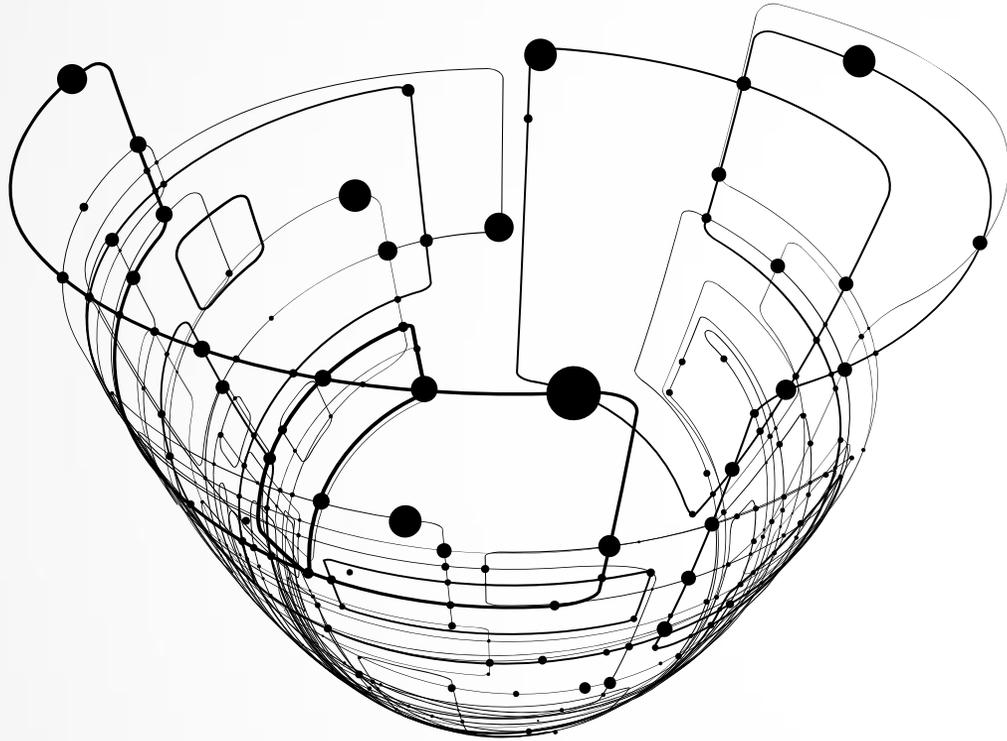
Japanese Artificial Intelligence ( AI ) Society is equipped with an ethics committee inside. High-tech enterprises have also strengthened the ethical responsibility requirements for scientific and technological personnel.

For example, leading enterprises such as Google, IBM and Microsoft put forward the principles of Artificial Intelligence ( AI ) development and established 'Artificial Intelligence ( AI ) Research Ethics Committee' to conduct self-review of algorithms from the whole life cycle of AI, thereby strengthening ethical constraints and guidance for scientific and technological personnel and strengthening ethical factors in the development and application of new technologies.

## II

---

We can see that the ethical governance of Artificial Intelligence ( AI ) is carried out according to internalism and externalism according to the analysis on actions for ethical governance of Artificial Intelligence ( AI ) in various countries.



**A** Internalism approach of Artificial Intelligence ( AI ) ethical governance

**B** Externalism approach to Artificial Intelligence ( AI ) ethical governance

## II ( A ) Internalism approach of Artificial Intelligence ( AI ) ethical governance

---

Artificial Intelligence ( AI ) itself is mainly imagined as technical artificial material with the function of moral persuasion in the internalism path. Moral factors are infiltrated into every link of artificial intelligent technology through value sensitive design, which are materialized to specific Artificial Intelligence ( AI ) technology. We can adopt three modes to materialize morals in the Artificial Intelligence ( AI ) technology materials on the basis of analysis on Verbeek.

The **first** mode refers to the designer's moral imagination, namely the designers place Artificial Intelligence ( AI ) into different contexts through imagination with their professional skills and experience. They monitor and forecast its possible mediation role, then the obtained results are fed back to design, and the technology moral mediation role is perfected through modifying the plan.

## II ( A )

---

The **second** mode refers to expansion of constructive technology assessment. It is argued in constructive technology assessment that the assessment of technical material stakeholders is included in the design, it is emphasized that social factors produce constructive influence on the design plan. Technology is also regarded as the stakeholder of the design plan through expansion of constructive technology assessment, it should achieve own position in democratic design, thereby realizing effective communication between application context and design context;

The **third** mode refers to scenario simulation method, the methods adopted by the users are predicted and analyzed as comprehensively as possible through the effective simulation of use scenarios. It is obvious that the authenticity of scenario simulation should be ensured as the key. Therefore, imagination, assessment and even some technical means can be included. If virtual reality technology undergoes scenario simulation, the accuracy of the mediation prediction can be improved.

The analysis on internalism shows that the ethical governance of Artificial Intelligence ( AI ) is presupposed and preexisting, and therefore it is presupposed.

## II ( B ) Externalism approach to Artificial Intelligence ( AI ) ethical governance

---

The externalism approach refers ethics are regarded as a normative force beyond technical activities, the technologies are ethically reflected, supervised and regulated from the results of technical activities, and it focuses on the application consequences of technologies.

Artificial Intelligence ( AI ) technology is adopted as an example. Image recognition technology provides people with convenient and rapid technical support in the occasions requiring identity authentication such as payment, travel, medical treatment, etc. However, such technology also facilitates financial fraud, identity theft, privacy malicious revealing. Therefore, **unintentional** malicious use and **intentional** malicious use of Artificial Intelligence ( AI ) technology are important contents of Artificial Intelligence ( AI ) ethical governance.

## II ( B )

---

European Commission appointed 52 representatives from academia, industry and civil society to form High-level Expert Group on Artificial Intelligence ( AI ) (AI HELP) to support the implementation of the European Artificial Intelligence ( AI ) strategy in June 2018.

The industry, research and energy committee of the European Parliament issued a report calling on the European Parliament to formulate a comprehensive EU industrial policy on Artificial Intelligence ( AI ) and robotics in January 2019, including the legal framework, ethics and governance of cyber security, Artificial Intelligence ( AI ) and robotics, etc.

## II ( B )

---

EU issued two important documents in April 2019-- *'Ethics Guidelines for Trustworthy Artificial Intelligence'* ('Ethics Guidelines' for short) and *'Governance Framework for Algorithmic Accountability and Transparency'* ('Governance Framework' for short). The ethical framework provides principles, guidance and basic requirements for the design, development, production and utilization of artificial intelligence, thereby ensuring that its operation complies with legal, safety and ethical standards.

It means that the ethical governance of Artificial Intelligence ( AI ) cannot remain at the level of abstract principles, it should be integrated into different subjects and practical activities at different levels, thereby becoming a living mechanism.



# III

---

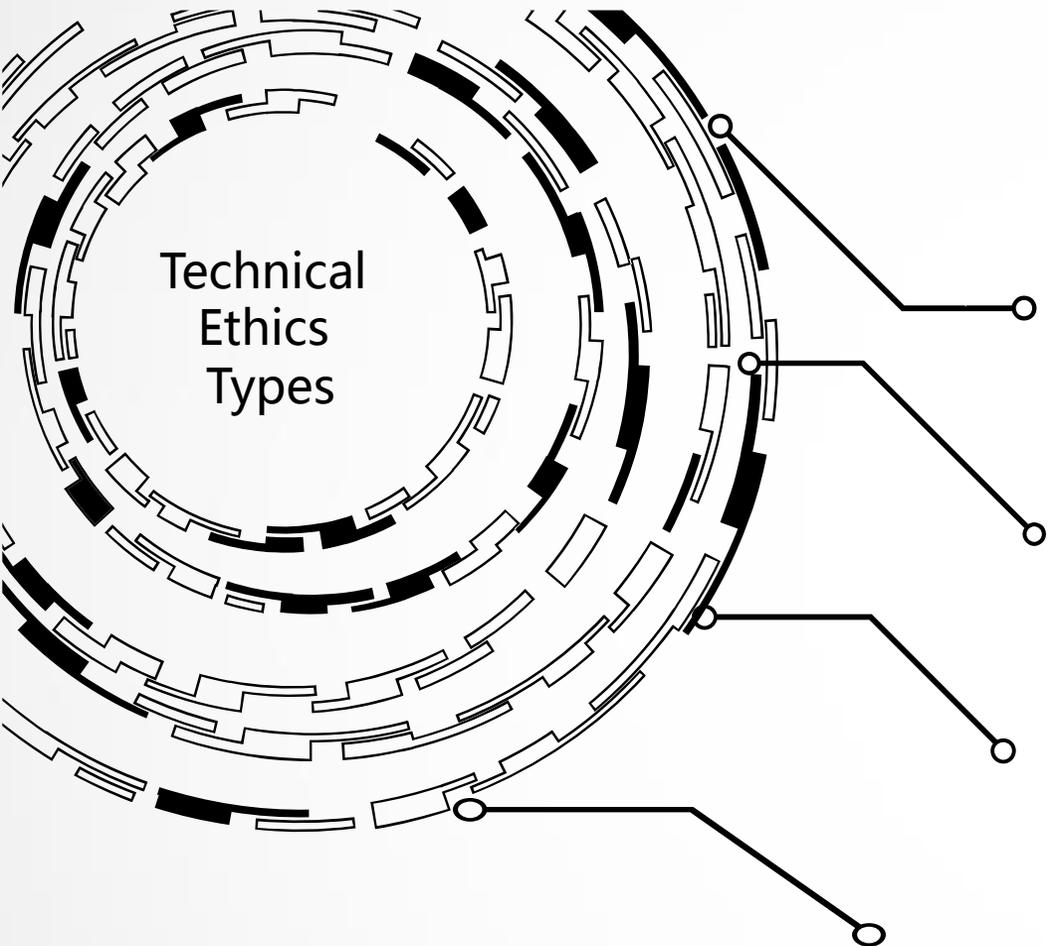
'Technology domination by  
morality': Chinese tradition of  
ethical governance of Artificial  
Intelligence ( AI )

### III

---

Favorable confirmation is provided for ethical governance of Artificial Intelligence ( AI ) from the empirical level according to China technical tradition 'technology domination by morality'.

The so-called 'technology domination by morality' refers to a technical practice mode that specifies and restricts technical behavior and application of technology with ethics and morals. The practice mode is a tradition formed and maintained in China technical activities for thousands of years, which is reflected in the thoughts of Confucianism ( 儒 ), Taoism ( 道 ) and Mohism ( 墨 ).



(A) Technical ethics of Confucian ( 儒 )

(B) Technical ethics of Taoism ( 道 )

(C) Technical ethics of Mohism ( 墨 )

( D ) Chinese technical ethics characterized by 'Technology domination by morality'

### III ( A ) Technical ethics of Confucian ( 儒 )

---

It is emphasized in Confucianism that the function of the technology in line with codes of ethics and benefiting the state and the public should be exerted, thereby limiting and eliminating the application of technology that does not comply with the codes of ethics.

It is recorded in *'Rites· Royal System'* that 'lewd sounds, strange clothing, strange technology and strange devices may doubt the public, and they should be killed.' ( 《礼记·王制》：“作淫声、异服、奇技、奇器以疑众,杀。” ) It shows that the technology of 'diabolic tricks and wicked craft ( 奇技淫巧 )' must be prohibited and restricted. However, agriculture mulberry, water conservancy, construction, metallurgy and similar technologies concerning state long-time stability and the populace basic survival technology must be vigorously advocated.

It is also stressed in Confucianism that 'technology' should be based on benevolence. Mencius said that 'technology must be prudent' ( 术不可不慎也 ).



### III ( B ) Technical ethics of Taoism ( 道 )

---

'DAO ( 道 )' is the highest category in Taoist philosophy. Taoism 'technology domination by morality ( 以道驭术 )' is manifested by the relationship of three moralities and technologies namely 'Dao is similar to technology ( 道近乎技 )' , 'Dao lies in technology ( 道在技中 )' and 'Dao and technology are integrated ( 道技合一 )' .

Multiple meanings were given to ' Dao ' by Laozi:

- It refers to the entity that makes up the world;
- It refers to the force that created the universe; or the laws of motion and change;
- It refers to the code of human behaviors.

### III ( B )

Laozi proposed that human activities, including the application of technology, should conform to the basic spirit and characteristics of 'Dao' and 'Virtue'. Laozi required that the development of technology should be restricted by corresponding ethics, namely 'Dao' .

Technology is inseparable from human conscience and goodness. When technology is integrated into the harmonious development of social culture and ethics, it can make contributions to social progress and get rid of the human misfortune of morality declining or even falling due to technological development. LaoZi observed that technology might destroy human nature and cause harm without moral constraints.



### III ( C ) Technical ethics of Mohism ( 墨 )

Mohism is a typical representative of craftsmen and civilian thought in ancient China

Mohism 'technology domination by morality' is mainly embodied as follows: the technical activities of the group or individual craftsmen should be controlled by the technical ethics. The disciples are required to learn from the spirit of Dayu, who worked hard to control the flood regardless of the wind and rain, Dayu did not care about high official positions and riches, who worked diligently for the interests of the public.

It is introduced in *'Zhuangzi· World'* that the writer in the later generations wear poor clothes and shoes, who worked hard regardless of day and night' . It is also concluded that the practice is wrong, it is not the thought of Dayu, and is it unworthy of writing' . ( 《庄子·天下》篇中介绍说, “后世之墨者, 多以裘褐为衣, 以为服, 日夜不休, 以自苦为极。” 并说: “不能如此, 非禹之道也, 不足谓墨。” ) The the technical ethics and conduct of their disciples and artisans are clearly regulated in these requirements.

### III ( C )

---

Mohism opposes extravagance in the application of technology.

'The ancient emperors built their palaces just for the convenience of life rather than pleasure of viewing' according to '*Mozi·Ciguo*' ( 《墨子·辞过》 ). However, 'the present emperor builds the palace through lavishing money on the people, and seizing their food and clothing, and a great deal of money is consumed, thereby making the country poor, and it is difficult to govern people.

The ancient holy king wore clothes just for the convenience of the body rather than seeking gorgeousness. However, 'the present king collects money from people, strives for seizing wealth, and pursues luxury, thereby forming customs, therefore it is difficult to control the public and restrict the king' , thereby inevitably leading to the world chaos.



### III ( D ) Chinese technical ethics characterized by 'Technology domination by morality'

---

'Technology domination by morality' is a unique technical ethic tradition formed in Chinese society during handicraft stage. The technology itself as well as the subjects of technical activities are bound and specified by ethics and norms in ancient Chinese society based on all social activities, ethics keeps a decision leading role to technology with its integrity, systematicness and stability. Technology is slowly developed towards the direction of benefiting livelihood, stability and harmony between nature and human beings.



# IV

---

Responsible innovation: an  
action framework for  
Artificial Intelligence ( AI )  
ethical governance

## IV

---

Empirical interpretation is made for technical ethical governance by 'technology domination by morality' from perspective of history. It is the action mode of ethics on technology in the Chinese ancient society emphasizing ethics and ignoring technology. Weak technical force is not sufficient to form force contrast and counterbalance with strong ethics and morality from the technical perspective. Technology belongs to passive obedience aiming at ethics. Ethics belongs to a force control aiming at technology.

## IV

---

Now, human beings are brought into the contemporary society emphasizing technology and ignoring ethics by artificial intelligence. The force contrast of technology and ethics is reversed, the border of ethics system is continuously broken by artificial intelligence. Ethics loses the power of the forcible control to technology. It seeks harmonious coexistence pattern with Artificial Intelligence ( AI ) technology as a mediator and adapter. The ethic system must construct a new action framework of good governance on Artificial Intelligence ( AI ) technology with an open, inclusive, covariant attitude in the face of strong Artificial Intelligence ( AI ) technology.

Responsible Innovation (RI) was firstly proposed by Thomas Heisstrom, a German scholar. He believed that the general framework of 'responsible innovation' should be established 'in the context of broader and universal technological development' (Tomas Hellström, 2003). Rene von Schomberg, Richard. Owen, B.Steinhl, J. Stilgoe, Wilford, Hilary and others discussed Responsible Innovation.

## IV

---

Rene von Schomberg believed that responsible innovation is a transparent and interactive process. Social actors and innovators echo each other in the process. The (ethics) acceptability, sustainability, and social desirability of innovation process and its marketable products are fully considered, thereby embedding the progress of science and technology appropriately into our social life.

J.Stilgoe argue that responsible innovation is a process of 'exploring the future through the collective management of existing science and innovation'.

Richard Owen points out that responsible innovation should be regarded as a carrier rather than a new governance paradigm. In the carrier, 'risk management' is changed into 'governance of the innovation itself', 'four dimension' action framework of responsible innovation is further proposed, namely the governance of technology innovation is realized through prediction, reflection, consultation and feedback.

## IV

---

Therefore, the responsible Artificial Intelligence ( AI ) ethical governance includes three paths:

Firstly, 'responsible innovation' centers on human behaviors and activities with innovative characteristics. Innovation is transformed into a harmless driving force, behavior and activity to promote the sustainable development of human economy, politics, culture, society and environment. by embedding the appeal of 'responsibility' into the whole process and all aspects of innovation.

Secondly, scientific and technological innovation is an indispensable and important way to achieve sustainable development, but scientific and technological innovation may also cause serious and harmful consequences. It is necessary to couple it with responsibility ethics. Therefore, the combination of responsibility and innovation is the fundamental way to achieve sound governance of technology.

## IV

---

Thirdly, the high risks implied by the new and high technology and the malignant consequences caused by the abuse and misuse of technology have led to people's distrust and disobedience of scientific and technological innovation and even the government, enterprises, scientific research institutes as well as other innovation and governance subjects. It is a common demand to establish open, transparent and participatory innovation regulation.

## IV

---

China National Committee for Governance of New Generation of Artificial Intelligence ( AI ) released '*Governance Principles of New Generation of Artificial Intelligence ( AI ) -- Development of Responsible AI*' on June 17, 2019, which proposed an action framework for ethical governance of Artificial Intelligence ( AI ) in China.

Eight principles :

- Fairness, ( 公平 )
- Inclusiveness, ( 包容 )
- Sharing, ( 共享 )
- Respect for Privacy, ( 尊重隐私 )
- Security and Control, ( 安全可控 )
- Shared Responsibility, ( 共担责任 )
- Open Collaboration, ( 开放协作 )
- Agile Governance. ( 敏捷治理 )

## IV

---

These eight principles are closely related to the theme of 'responsible' for Artificial Intelligence ( AI ) governance. It should be concluded that the promulgation of the '*Governance Principles*' marks the establishment of the action framework of Artificial Intelligence ( AI ) ethical governance in China. It can be regarded as the inheritance of Chinese tradition 'technology domination by morality' in China on the one hand, it is also the positive response to the new concept of responsible innovation that has emerged in Europe and the United States in recent years.

谢谢

---

**THANKS !**

Welcome comment and question!!!

