Received: 17-08-2025
Accepted: 21-11-2025
Published: 21-04-2026

Progress, Challenges, and Future Path of AI Ethics Governance in China

Hongxu Zhou
Associate Professor, Yunnan Administration College
eileenzhou@aliyun.com

How to cite: Zhou, H. (2026). Progress, challenges, and future path of AI ethics goverance in China. *Journal of Information, Knowledge and Libraries 1*(1) 40-47.

Abstract

The rapid development of artificial intelligence (AI) presents both opportunities and challenges, particularly in navigating its ethical implications. While AI offers transformative potential across industries, concerns regarding bias, transparency, accountability, and societal impact necessitate robust ethical frameworks. This study investigates China's approach to AI ethics governance, aiming to analyze its effectiveness and identify areas for improvement. The research question guiding this study is: What are the key challenges and opportunities facing China in establishing a comprehensive and effective framework for AI ethics governance? Building on existing literature examining global AI ethic frameworks and China's policy landscape, this research adopts a mixed-methods approach. This includes analyzing relevant policies and regulations, conducting case studies of AI applications in various sectors. The expected outcome is a comprehensive understanding of China's AI ethic governance landscape, highlighting its strengths, weaknesses, and potential future directions. This research contributes to the ongoing global discourse on AI ethics by providing valuable insights into the experiences and approaches of a major player in the field.

Keywords: china, artificial intelligence, ethical aspects, artificial intelligence ethics, AI governance

Introduction

As an important driving force for leading a new round of scientific and technological revolution and industrial transformation, the profound expansion of artificial intelligence technology not only profoundly changes people's production, life and learning methods, but also brings major opportunities for human society to achieve human-machine collaboration, cross-border integration, co-creation and sharing. The large model of artificial intelligence has driven the rapid development of the generative artificial intelligence industry, which has brought huge innovation opportunities in many fields such as scientific exploration, technological research and development, artistic creation, and enterprise management. Global investment in AI is expected to reach nearly $200 billion by 2025 (Goldman Snac, 2023). At the same time, the rapid development of artificial intelligence has also brought various uncertain risks and global challenges such as data competition, data security risks, intellectual property infringement risks, false information or vicious content risks, and technological hegemony risks. The breakthrough development of artificial intelligence technology has caused profound ethical problems and become a hot issue of global concern. China is one of

the important participants in the field of artificial intelligence in the world. China has promoted the ethical governance of artificial intelligence for years and achieved positive progress. However, it is clear that progress and challenges are very closely associated with each other. At present, China has become the second largest AI country in the world, and the research on the status quo, challenges and future path of China's AI ethical governance may provide useful references for the international community to deal with related issues.

The conceptual definition of artificial intelligence ethics

"Ethics" is the code of human behavior, the truth and order of dealing with the relationship between people and the relationship between people and society. In science and technology activities, ethics guides technology development and application at the level of value guidance and practice norms. Ethics, also known as moral philosophy, is a branch of philosophy that studies the norms of behavior and values accepted by human society, and then constructs moral guiding principles for judging behavior from right and wrong, good and evil, beauty and ugliness, wisdom and ignorance, justice and crime. Ethics can be roughly divided into four fields: meta-ethics, normative ethics, descriptive ethics and applied ethics, and the ethics of artificial intelligence basically belongs to the category of applied ethics (Jiangsheng Yu,2022).

As far as the relationship between technology and human morality is concerned, the ethical governance of artificial intelligence is an age-old problem. Artificial intelligence has been around for more than 70 years, and concerns about the ethics of AI have been around since the mid-20th century (Kaynak, 2021). At the current stage, artificial intelligence not only inherits the ethical problems of previous information technology, but also has new characteristics due to the opaque, difficult to explain, adaptive, and widely used characteristics of some artificial intelligence algorithms such as deep learning which may bring a series of ethical risks in many aspects. The bias of artificial intelligence algorithms in target demonstration, algorithm discrimination, and training data may bring about or expand discrimination in society and violate citizens' right to equality. The abuse of artificial intelligence may threaten citizens' right to privacy and personal information. Complex artificial intelligence algorithms such as deep learning can lead to algorithmic black box problems, making decisions opaque or difficult to interpret, thus affecting citizens' right to know, due process, and citizens' right to supervise. The abuse and misuse of artificial intelligence technologies such as accurate information push, automated fake news writing and intelligent targeted dissemination, and deep counterfeiting may lead to problems such as information cocoons and the proliferation of false information and may affect people's access to important news and democratic participation in public issues. The precise push of false news may also increase the influence of people's knowledge and opinions on facts, which may incite public opinion, manipulate commercial markets, and influence politics and national policies.

According to the OECD AI Incidents Monitor, there have been 7,758 AI incidents worldwide since January 2014. It peaked in November 2023 with 510 cases (OECD, 2024). AI ethical issues cover almost every stage from technology development, product development and application, and can be found in technology, value, innovation and order systems. In these systems, the basic form of ethical problems can evolve into risk of loss of control, behavioral disorders, and ethic obstacles. In the specific application scenarios, it is more prominent in the fields of graphic generation, automatic driving, intelligent medical treatment, etc. In summary, AI ethics is the value concept and code of conduct that should be followed in carrying out AI research, design, development, service and use.

The category of artificial intelligence ethical governance

The much-quoted definition is: "Governance is the sum of the many ways individuals and institutions, public and private, manage their common affairs. It is a continuing process through which conflicting or diverse interests may be accommodated, and cooperative action may be taken. "According to

UNESCO's recommendation on the Ethics of Artificial Intelligence, Article 75 states, "Member States should promote open data," and Article 76 states, "Member States should promote and facilitate the use of quality and robust datasets for training, development and use of AI systems, and exercise vigilance in overseeing their collection and use." According to UNESCO's recommendation on the Ethics of Artificial Intelligence, Article 131 states, "Member States should, according to their specific conditions, governing structures and constitutional provisions, credibly and transparently monitor and evaluate policies, programmes and mechanisms related to ethics of AI, using a combination of quantitative and qualitative approaches……(d) strengthening the research- and evidence-based analysis of and reporting on policies regarding AI ethics; (e) collecting and disseminating progress, innovations, research reports, scientific publications, data and statistics regarding policies for AI ethics……" On 22 September of 2024,during the UN Future Summit World leaders adopted a Pact for the Future that includes a Global Digital Compact and a Declaration on Future Generations. The Global Digital Compact, annexed to the Pact, is the first comprehensive global framework for digital cooperation and AI governance. At the heart of the compact is a commitment to design, use and govern technology for the benefit of all. This includes commitments by world leaders to:

- Connect all people, schools and hospitals to the Internet;
- Anchor digital cooperation in human rights and international law;
- Make the online space safe for all, especially children, through actions by governments, tech companies and social media;
- Govern artificial intelligence (AI), with a road map that includes an International Scientific Panel and a Global Policy Dialogue on AI;
- Make data more open and accessible, with agreements on open-source data, models and standards.

Different from other emerging technologies, artificial intelligence has complex attributes such as concealability of technical core, anthropomorphism of technological form, cross-field application opportunities, interweaving of interest subjects, multi-dimensional technological risks, and complexity of social impact. It has triggered risks and challenges in different dimensions such as society, enterprises and individuals, and thus put forward the urgent need for professionalization, diversification, agility and globalization of governance (China Academy of Information and Communications Technology and China Artificial Intelligence Industry Development Alliance, 2020). Therefore, the concept of artificial intelligence governance came into being and has become the direction of extensive attention and research in the international community. Artificial intelligence governance should not only make reasonable use of the advantages of artificial intelligence but also be good at avoiding the shortcomings of artificial intelligence, so as to promote the social well-being of all mankind.

Ethic concerns in the governance of emerging technologies provide rich theoretical resources for the ethical governance of artificial intelligence. At present, the research on artificial intelligence ethic governance focuses on these key aspects. Research on governance criterion, regulation method and action path of artificial intelligence ethics, ethic governance and countermeasures in different fields of big data and artificial intelligence, embedded and regulatory ethic governance of autonomous vehicles explores the ability of autonomous driving to cope with complex societal risks, classifies (Wang,2022).

Recognizing the enormous potential and obvious risks of AI, governments, intergovernmental organisations, research institutions, non-profit organizations, science and technology societies, and businesses around the world have proposed more than 70 proposals on the ethical principles of AI. The proposals focus on 10 themes: people, cooperation, sharing, fairness, transparency, privacy, external security, internal security, accountability, and long-term application. On 24 November 2021, UNESCO published recommendations on the Ethics of AI, the first global normative framework for

the ethical governance of AI, which was adopted by all 193 member states. The recommendation identifies 10 principles and 11 areas for action to regulate AI technology. The proposal also states that the development and application of AI should reflect four values（UNESCO, 2021）:

- Respect, protect and promote human rights and enhance human dignity;
- Promote the development of the environment and ecosystems;
- Promote diversity, inclusion and equity in the workplace;
- Building peaceful, just and interdependent human societies.

However, translating these ethic principles into practice is a complex and challenging exercise that requires a more systematic approach in identifying problems, choosing solution paths, and assigning roles to relevant stakeholders. Therefore, each country needs to find its own path of AI ethic governance, and implement AI governance strategies such as ethic embedding, evaluation, adaptation, and construction in technology life cycle stages such as R&D, design and manufacturing, experimental promotion, and deployment and application.

Research on China's artificial intelligence ethics governance

The development process of China's AI ethics research has roughly experienced three stages: robot ethics, data ethics and AI ethics, and has gradually changed from taking human as the basic research object to multiple research objects. Chinese scholars' research on the governance of artificial intelligence ethic risks generally includes three perspectives: principle norms, theoretical orientation and ethic concepts (Lu and Cui, 2022). Principle norms include the formulation of principles and the establishment of legal systems. Theoretical orientation refers to the exploration of ethical governance countermeasures of artificial intelligence based on a certain theory, and the governance methods of ethical issues of artificial intelligence from the perspective of philosophy. Some scholars start with the internal approach to ethical governance methods. Some research focuses on artificial intelligence ethics risk, and argues that the uniqueness and openness are characteristics of artificial intelligence ethics risks, and there are three challenges in terms of theoretical level, institutional level and governance subject level. In specific fields, a white paper on Empowering the healthcare industry with large AI models, written by the China Academy of Telecommunication Research of MIIT (CAICT), discussed the ethical problems while large models of artificial intelligence empower healthcare, such as large AI models exacerbating the problem of medical bias and the spread of harmful disinformation. One of the most discussed areas of AI ethics is autonomous driving; some experts discussed ethical dilemmas and legislative responses to autonomous vehicles (Zheng Zhifeng, 2024). While the application of digital technology is increasingly widespread, scientific and technological ethical irregularities such as the digital divide, big data killing, algorithm discrimination, privacy disclosure, ageing of financial products and lack of barrier-free construction are common (Shan Qiang, 2023).

Progress of ethic governance of artificial intelligence in China

In recent years, China has continuously promoted the process of artificial intelligence governance, which can be understood from several dimensions of governance concept, establishment and improvement of mechanisms, detailed measures and industry norms, and international cooperation.

The position of ethical governance of artificial intelligence in China

China released its position paper on strengthening the ethical governance of artificial Intelligence on November 17th of 2022. In the document, it has been clearly mentioned:

We can find that "putting people first" and "smart for the good" are the core of AI ethic governance. If we follow up China's AI governance process, ethic always goes firstly, and all plans, policies are pointed to goodness for all man kinds.

The industry actively explores AI ethic governance landing measures

In recent years, many enterprises of China have actively implemented the ethic management requirements of artificial intelligence and explored relevant self-regulatory measures, which mainly consist of Publishing AI ethic principles, establishing internal AI governance organizations (such as ethics committees), conducting ethic reviews or security risk assessments for AI activities, disclosing information related to algorithms, and promoting openness and transparency of algorithms, exploring technical solutions to AI ethic issues (such as detection and identification tools for synthetic information content, AI governance assessment tools, privacy computing solutions such as federated learning), and so on. Taking Huawei as one example, as a leading global provider of information and communications technology (ICT) infrastructure and smart devices, Huawei has its own AI ethic governance principles:

Firstly, AI can only serve humans, not determine humans, which is a fundamental principle that must be realized. Secondly, the unfairness that artificial intelligence may promote development should be corrected. AI products must have quality and security guarantees, including that the data used in the training process should not violate privacy and corporate intellectual property rights; In addition, the training process and reasoning process of AI large models should be traceable so that responsibility can be clarified when problems occur.

Actively participate in international cooperation on AI ethic governance

At the regional level, in August 2023, during the 15th BRICS Leaders' Meeting, the BRICS countries have agreed to launch the work of the artificial intelligence research group as soon as possible, expand artificial intelligence cooperation, form a broad consensus on artificial intelligence governance framework and standards, and improve the safety, reliability, controllability and fairness of artificial intelligence technology. At the Bilateral level, by facing the risks and challenges posed by the rapid development of artificial intelligence (AI) technology, the industry has been looking to China and USA which are the two leading countries to work together. Recently, the cooperation trend between China and the United States in the field of artificial intelligence has also become the focus of global attention. In addition, China is taking a more proactive attitude to be an active promoter, participant and contributor in the field of global AI governance. In 2023, China issued the Global AI Governance Initiative, which highlights the ethical principles as follows:

Challenges of AI ethics governance in China

Global AI governance faces many challenges that hinder the process of AI ethic governance, some external and internal factors which also hinders AI ethic governance in China.

External and Internal factors: In terms of technical safety, problems such as difficulty, lag and weak binding of standards are evident; in terms of technological innovation, there is a lack of trust among countries, and strategic competition replaces open cooperation in technology; problems of digital divide and unequal development have emerged in the application of technology.

Developing countries, including China, still have the problem that their participation in the global governance and international rulemaking of AI ethics is not deep and the scope of participation is narrow.

There are also three constraints on China's AI development. From a data point of view, artificial intelligence cannot be separated from stable data. Systems must input large amounts of data so that they can "train" themselves to continually improve and refine their output, thereby facilitating the adjustment and regulation of the ethic dimension of AI. However, there is still a gap between China and developed countries in creating a data-friendly ecosystem with unified standards and cross-platform sharing. Second, there is still a need for more openness in opening up government datasets

to spur private-sector innovation and cross-border data flows. From the algorithm point of view, at the application level, China is on par with other countries in algorithm development. In particular, breakthroughs have been made in developing artificial intelligence algorithms for speech recognition but China lags behind countries such as the US and the UK in the basic research that is driving the development of artificial intelligence.

One of the main reasons is the shortage of talent. Both the quantity and quality of relevant talents in China are far from meeting the needs. With the wide application of artificial intelligence in various industries, the demand for talents in this field has exploded, but the total supply of talents is seriously insufficient, which is of course a common problem faced by most countries in the world. It is expected that the shortage of AI professionals in China will reach 4 million by 2030.

The theory and practice of artificial intelligence ethic risk prevention need to be more closely integrated. In fact, although the current research on artificial intelligence ethic risk prevention has covered many aspects in the horizontal aspect, it is mostly theoretical discussion, lack of operability, and difficult to fall to practical practice. Because there is no safeguard measure to implement the theory and principle of risk prevention, resulting in the dilemma of "lack of practicality" in artificial intelligence ethic risk prevention. In this sense, in order to prevent ethic principles from becoming abstract concepts, all the ethical principles must be permeated and reflected in all aspects and the whole process of artificial intelligence technology research and development to application.

The supply of policies and regulations is insufficient and unbalanced. With the increasingly profound impact of artificial intelligence on human production and life, some organizations and institutions have conducted discussions on related issues and issued a number of suggestions on artificial intelligence ethic guidelines. However, in general, these policies focus on macro-level industry promotion, and there is a lack of guidance for specific areas. For example, for products and fields that have emerged, the number of relevant policies and regulations is relatively small, whether at the national or local level.

There is currently a disconnect between the teams that really understand large AI model technology and the people who apply large models.

In addition, although there is a certain foundation for the construction of science and technology ethics standards, there is still room for improvement to promote the implementation of standards through such measures. In the fields of drones, autonomous driving, areas of intelligence applications, whether it is the promotion policy of artificial intelligence industry at the macro level, or the guidance policy in specific fields, there is an unbalanced supply situation.

Suggestion to improve AI ethical governance in China

Make unremitting efforts to develop the artificial intelligence industry and promote artificial intelligence ethics in the process of development. China can build a stronger data ecosystem by developing and implementing data standards, further opening public sector data to private sector innovation and exploration, and encouraging cross-border flows of data. In the development of artificial intelligence, the effectiveness of scientific and technological ethics governance should be enhanced, and industrial innovation should be coordinated with ethical risk prevention.

Artificial intelligence itself is highly interdisciplinary, and interdisciplinary perspectives and joint efforts are essential, as well as attempts to achieve cross-disciplinarity and diversity in practice. Building a network channel that can communicate promptly and based on strengthening the consultation and communication between the government and research institutions, technical experts, social organizations and the public, discuss the relationship between the rights and responsibilities of the relevant responsible parties. Laws and regulations related to the ethical governance of artificial intelligence should be strengthened.

Governments need to invest in AI-related education and research programs, re-adjusting the education system, support universities to set up science and technology ethics courses to guide provincial and scientific workers to improve artificial intelligence ethics. Strengthen science and technology ethics education for the public.

Conclusion

Building on existing literature examining global AI ethics frameworks and China's policy landscape, this research analyzing China's AI ethic governance landscape. The construction of artificial intelligence ethic laws and regulations has been accelerated in China. Developing countries, including China, still have the problem that their participation in the global governance and international rule-making of AI ethics is not deep and the scope of participation is narrow. The monopoly of developed countries on governance mechanisms and international rules further strengthens the dilemma that developing countries cannot participate effectively in international AI ethic governance. There is still a gap between China and developed countries in creating a data-friendly ecosystem with unified standards and cross-platform sharing, China lags behind countries such as the US and the UK in the basic research that is driving the development of artificial intelligence, the quantity and quality of relevant talents in China are far from meeting the needs. The theory and practice of artificial intelligence ethic risk prevention need to be more closely integrated, the supply of policies and regulations is insufficient and unbalanced, and unclear boundaries of responsibility make ethic review more difficult.

The feasible countermeasures and suggestion on improving AI ethic governance in China have been discussed in this paper. Such as making unremitting efforts to develop the artificial intelligence industry and promote artificial intelligence ethics in the process of development, improving AI ethic governance framework, enhancing the ability of various entities to deal with the ethic risks of AI, and strengthening international cooperation on AI governance by promoting the Global AI Governance Initiative.

In the future in the living world of human-machine coexistence, new ethic problems and moral dilemmas may continually emerging, and how to build a responsible artificial intelligence governance system has become a key issue in the development of artificial intelligence and ethical governance of science and technology. At present, AI ethic governance is transiting from theoretical discussion to regulation and framework design, and although countries overlap in some respects there are still significant differences. We need to have universal rules across the globe but also allow each region to adapt the rules to its own culture and needs. China's exploration and practice of AI ethic governance may provide some references for other developing countries.

References

China Academy of Information and Communications Technology, & China Artificial Intelligence Industry Development Alliance. (2020, September). *White paper on artificial intelligence governance*. http://www.caict.ac.cn/kxyj/qwfb/bps/202009/P020200928368250504705.pdf

China Artificial Intelligence Industry Development Alliance. (2019, June 1). *Notice on the "Artificial Intelligence Industry self-discipline convention (draft for comment)" for public comment*. https://www.sohu.com/a/317958550_735021

Goldman Sachs. (2023, August 1). *AI investment forecast to approach $200 billion globally by 2025*. https://www.goldmansachs.com/intelligence/pages/ai-investment-forecast-to-approach-200-billion-globally-by-2025.html

Ministry of Foreign Affairs of the People's Republic of China. (2022, November 17). *China's position paper on strengthening the ethical governance of artificial intelligence*. https://www.mfa.gov.cn/web/wjb_673085/zfxxgk_674865/gknrlb/tywj/zcwj/202211/t20221117_10976728.shtml

Ministry of Foreign Affairs of the People's Republic of China. (2022, June 14). *Artificial intelligence governance policies*. https://www.mfa.gov.cn/wjb_673085/zzjg_673183/jks_674633/zclc_674645/rgzn/202206/t20220614_10702838.shtml

Stanford University, Institute for Human-Centered Artificial Intelligence. (2023). *AI index report 2023*. https://hai.stanford.edu/assets/files/hai_ai-index-report_2023.pdf

Yu, J. (2022). *Ethics of artificial intelligence*. Tsinghua University Press.

Kaynak, O. (2021). The golden age of artificial intelligence. *Discover Artificial Intelligence, 1*(1), 1–7. https://doi.org/10.1007/s44163-021-00001-8

Lu, Y., & Cui, Z. (2022). Research progress and trends on artificial intelligence ethics in China. *Science & Technology Review, 40*(18), 69–78. https://doi.org/10.3981/j.issn.1000-7857.2022.18.009

Ministry of Science and Technology of the People's Republic of China. (2022, March 23). *Opinions on strengthening ethical governance of science and technology: Press conference*. https://www.most.gov.cn/xwzx/twzb/fbh22032301/

OECD. (2024, January). *The OECD AI incidents monitor (AIM)*. https://oecd.ai/en/incidents

UNESCO. (2021, November 24). *Recommendation on the ethics of artificial intelligence: Shaping the future of our societies*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Zheng, Z. (2024). Ethical dilemmas and legislative responses to the "trolley problem" of autonomous vehicles. *Law Science, 516*(11), 107–123.