Exploration of Scientific and Technological Ethics Governance in the Age of Artificial Intelligence

Peng Deng*

Xinjiang Uygur Autonomous Region Institute of Science and Technology Development Strategy, Urumqi 830011, Xinjiang, China

Author to whom correspondence should be addressed.

Abstract: With the rapid development of artificial intelligence technology, China has entered the era of artificial intelligence. It not only promotes social development and progress but also faces a series of ethical risks. In response, this article starts from ethical risks, first expounds on the ethical turn in the development of artificial intelligence technology, then clarifies the core challenges of scientific and technological ethics governance in the era of artificial intelligence, and finally, proposes a theoretical framework and practical path for scientific and technological ethics governance, hoping to provide some reference for relevant personnel to better cope with scientific and technological ethics risks in the era of artificial intelligence.

Keywords: Artificial intelligence; Scientific and technological ethics; Governance framework; Multi-stakeholder governance; Responsibility ethics

1. Introduction

With the widespread application of artificial intelligence technology, remarkable achievements have been made in fields such as healthcare, education, finance, and transportation, bringing unprecedented convenience to society [1]. However, the development of artificial intelligence has also triggered a series of ethical issues, such as privacy leakage, data security, and algorithmic bias. These problems not only threaten the rights and interests of individuals but may also have a negative impact on social stability [2]. Therefore, how to conduct effective scientific and technological ethics governance in the era of artificial intelligence has become an important issue that needs to be solved urgently [3].

2. The ethical turn in the development of artificial intelligence technology
2.1. The historical inevitability of technological revolution and ethical reconstruction

Since the establishment of the artificial intelligence discipline in 1956, the development of artificial intelligence

technology has gone through three stages: symbolism, connectionism, and deep learning. Especially after 2020, artificial intelligence technology has embraced brand-new development opportunities and breakthroughs [4]. It has gradually shifted from the level of instrumental rationality to the level of value rationality. Moreover, the application scope of artificial intelligence technology is becoming wider and wider, covering fields such as education, healthcare, and automobiles [5]. The behavior choices and value discrimination of the public also rely on the help of this technology. At the same time, how to balance ethical norms and technological breakthroughs has become a key issue in social development. For example, modern labor laws emerged during the Industrial Revolution, and the General Data Protection Regulation was introduced in the Internet era [6]. Therefore, in the current social development situation, the decision-making influence, opacity, and autonomy of artificial intelligence technology have impacted social trust, fairness, and the subjectivity of the public from multiple dimensions [7]. There is a particular need to construct an ethical governance system that is highly compatible with social development trends and requirements.

2.2. The essential characteristics of artificial intelligence ethical issues

First, the virtualization of decision-making responsibility is mainly reflected in the ambiguity of responsibility definition, the unpredictability of decisions, and other aspects. This is mainly because artificial intelligence technology has strong autonomy and complexity [8]. The decision-making process is often disconnected from human control and understanding, increasing the unpredictability of decision-making results. In addition, if a series of adverse effects is caused by wrong decisions made by artificial intelligence technology, it is difficult to clearly define and divide the responsibilities. It is impossible to determine whether they are caused by humans, the system operating environment, data input, etc. Therefore, the virtualization of responsibility not only exacerbates the complexity of ethical governance but also further intensifies human distrust of artificial intelligence technology [9].

Second, the concentration and alienation of power. The over-concentration of power in artificial intelligence systems is mainly reflected in the possible abuse and misapplication of power in some cases, leading to power alienation. For example, if machine learning technology is applied to an important decision-making process without an effective supervision and balance mechanism, it may lead to excessive machine intervention or negative consequences for the decision-making process. This may cause the decision to deviate from the "origin" of the goal and even damage public welfare. The emergence of the "black-box" effect further increases the possibility of power alienation [10].

Third, the cognitive reconstruction crisis of the human-machine relationship. The main reason is the rapid development of artificial intelligence technology, which makes the relationship between humans, machines, and technology increasingly blurred. The traditional human-centered and machine-assisted model has undergone profound changes [11]. With the continuous improvement of the autonomy and intelligence level of artificial intelligence technology, machines and technology have surpassed humans in some aspects and can even influence human judgments and decisions in certain decision-making environments to a certain extent. This poses new challenges to social morality and legal rules [12].

3. The core challenges of scientific and technological ethics governance in the age of artificial intelligence

3.1. The inherent conflict between technological characteristics and ethical norms

First, uncertainty and the lag of rules. The dangerous dynamic changes brought about by the autonomous iteration

of AI systems cannot be addressed by the single-dimensional management of "formulating regulations-matching technology." For example, in the field of autonomous driving, 237 accidents related to Level 4 autonomous driving occurred globally in 2023, resulting in moral decision-making situations similar to those in "artificial intelligence-driven driving problems." However, no country's traffic laws clearly define and divide responsibilities [13].

Second, the tension between the pursuit of efficiency and the value of fairness. Enterprises tend to optimize the commercial value of algorithms rather than social fairness, resulting in imbalances in resource allocation. A 2024 study by Harvard University found that the personalized recommendation accuracy of a certain educational AI system for low-income family students is 27% lower than that for high-income family students, which has exacerbated the digital divide [14].

Third, the dilemma of balancing technological innovation freedom and social stability. The military application potential of general-purpose AI (such as autonomous killing robots) has triggered the panic of "technological out-of-control." The negotiation of the "Convention on the Prohibition of Lethal Autonomous Weapon Systems," signed by 120 countries in 2023, reached an impasse, reflecting the dilemma of the game between innovation incentives and safety bottom lines.

3.2. The structural defects of the existing governance system

First, the fragmentation of institutional supply. There are significant differences in legislation among countries. The EU's "Artificial Intelligence Act" (2021) established a risk-based regulatory framework, the United States adopted a decentralized governance model mainly based on industry self-regulation, and China's "Interim Measures for the Administration of Generative Artificial Intelligence Services" (2023) focuses on technological controllability. A global unified governance rule has not yet been formed [15].

Second, the blurred responsibilities of multiple stakeholders. Taking the "medical AI misdiagnosis case in a hospital in 2024" as an example, it is difficult to define the liability relationship among manufacturers, users, and regulators. The hospital shirked responsibility by claiming "technical difficulties," and the manufacturer shirked responsibility by blaming "user errors," which directly reflects the lack of laws on liability sharing.

Third, the lag in the construction of social consensus. There are significant differences in the ethical perception of AI in terms of age and region. According to research by the Chinese Academy of Social Sciences, among people over 60 years old, the proportion of those who trust AI-based medical decisions is 34%, while among Generation Z, this proportion reaches 72%. The obvious social differences in trust levels increase the governance cost.

4. The theoretical framework and practical path of scientific and technological ethics governance

4.1. The paradigm transformation of governance theories

First, from "post-hoc regulation" to "pre-emptive embedding": Introduce the "ethics-by-design" concept and embed value-sensitive design in the development stage of AI systems. The Microsoft Azure AI Ethics Committee requires all algorithm models to pass the "fairness-interpretability-robustness" triple test. In 2024, the deviation rate of its credit model was 65% lower than the industry average.

Second, from "single-stakeholder" to "multi-stakeholder governance": Build a collaborative governance network involving the government (rule-making), enterprises (technological compliance), society (value supervision), and the public (claiming rights). The EU AI Alliance has attracted more than 2,000 stakeholders to

participate in standard-setting, forming a dynamic mechanism of "regulatory sandbox + technical certification + social feedback."

Third, from "regional segmentation" to "global governance": In response to the borderless nature of AI technology, a transnational governance mechanism is needed. At the 2024 G20 Osaka Summit, the "Common Principles for Artificial Intelligence Governance" was adopted, proposing the goal of "global balance between technological innovation and human well-being" and promoting the establishment of an international mutual recognition system for cross-border data flow and algorithm transparency.

4.2. The construction path of a hierarchical governance system

4.2.1. Technical level: Build the infrastructure for trustworthy AI

Research and development of explainable artificial intelligence (XAI) technology: Develop causal reasoning models to replace pure statistical learning. The "neural-symbolic system" proposed by the MIT team decomposes decision-making logic into verifiable symbolic rules, increasing the acceptance rate of explanations in medical diagnosis scenarios by 40%.

Development of ethical risk assessment tools: Establish an assessment framework for the life cycle of artificial intelligence systems, refer to the NIST "AI Risk Management Framework," and use qualitative and quantitative indicators to assess algorithm bias, confidentiality, and social influence. It is expected that by using this framework, the proportion of discriminatory loan approvals in a financial institution will be reduced from 12% to 1.5% by 2024.

Governance application of blockchain technology: Use a decentralized account system to achieve traceability of information and accountability. The "algorithm review blockchain platform" project was piloted in Shenzhen. By recording the decision-making process of artificial intelligence and forming an evidence chain, the time for resolving disputes was shortened by 70%.

4.2.2. Institutional level: Improve the legal regulation and policy system

Classification of the risk levels of regulatory objects: Referring to the EU AI Act, classify the risks of AI applications into three levels (unacceptable risk, high risk, and low risk). Implement a stricter access mechanism for facial recognition (high-risk), and adopt a tolerant and prudent regulatory strategy for content recommendation (low-risk). In 2024, China issued the "Administrative Measures for the Classification and Grading of Artificial Intelligence (Draft for Soliciting Opinions)," clarifying 22 high-risk application scenarios.

Innovation in liability allocation laws: Establish an "algorithm liability insurance" system. Germany requires autonomous driving enterprises to purchase liability insurance of 20 million euros. Explore the "technical trustee" system, requiring AI system operators to act in the best interests of users. California in the United States is piloting the legislation of algorithmic fiduciary duty.

Design of an innovation-tolerance mechanism: Set up a "regulatory sandbox" to allow technology to make mistakes within a controllable range. The Financial Conduct Authority (FCA) sandbox in the UK has approved 37 AI-related financial projects, achieving a balance between innovation and safety through real-time risk monitoring.

4.2.3. Cultural level: Foster a social consensus on technological ethics

Construction of public participation mechanisms: Establish a citizen consultation platform for AI ethics. Canada's "Algorithm Impact Assessment Framework" requires that major AI application projects must include a public

hearing. In 2023, the Toronto public housing allocation algorithm was revised to correct 3 implicit discriminatory rules through citizen participation.

Education and morality: Incorporate the moral education of artificial intelligence into K-12 basic education. The IEEE in the United States has issued the "AI Ethical Norms," requiring high school students to have the literacy to identify and protect data rights. Universities offer a new interdisciplinary subject of "scientific and technological ethics," and Tsinghua University established the world's first AI Ethics and Governance Research Center.

The appeal of media culture and influence: Shape the community narrative of scientific and technological ethics through film and television works, public speeches, etc. In 2024, the BBC documentary "The Age of Algorithms: The Human Factor" attracted global attention to the ethical issues of AI-based emotion analysis technology, prompting relevant enterprises to actively self-regulate the use of their emotion-monitoring technology.

4.3. Improve the institutional norms of scientific and technological ethics

In the era of artificial intelligence, scientific and technological ethics risks are becoming more complex. Improving the institutional norms of scientific and technological ethics is the primary issue to be addressed and a challenge to be faced. In the current social situation, the institutional norms of scientific and technological ethics are in the initial stage, relying too much on guiding documents and human self-restraint. Although this approach is feasible in the short term, it is difficult to effectively cope with scientific and technological ethics governance risks in the long term. Moreover, post-hoc supervision and response measures have relatively limited effects and may even damage the images of collectives and individuals. Therefore, in combination with China's actual situation, effective countermeasures should be taken to improve the institutional norms of scientific and technological ethics as follows: First, deeply review and adjust the existing ethical supervision system to better address the ethical issues brought about by artificial intelligence technology. This also requires clearly defining the responsibilities of ethical review and supervision, making the review, handling, and punishment processes more standardized. Second, establish a special committee responsible for standard-setting, determining the registration system, and improving the regulatory framework. At the same time, continuously improve the certification and selection mechanisms for committee members to ensure that all members have high professional levels and capabilities, and continuously enhance the effectiveness of scientific and technological ethics supervision. Through these measures, the institutional norms of scientific and technological ethics can be continuously improved, providing an escort for the healthy development of science and technology and enhancing the scientific and technological images of collectives and individuals internationally.

4.4. Cultivate professional talents for scientific and technological ethics governance

To effectively improve the effectiveness of scientific and technological ethics governance, it is necessary to recognize the importance of cultivating high-quality professional talents. In the era of the rapid development of artificial intelligence technology, scientific and technological ethics governance talents not only need to have a solid theoretical knowledge reserve but also strong governance capabilities and experience. Therefore, professional talents for scientific and technological ethics governance can be cultivated from the following aspects: First, give full play to the roles of vocational education, higher education, etc. Add scientific and technological ethics courses to education at all stages to ensure that governance talents master the corresponding knowledge and skills before entering the workplace, deeply understand the connotation of scientific and technological ethics,

and effectively enhance their competitiveness. Second, strengthen talent cultivation efforts. Regularly organize training activities, and the training content can include legislation, judicature, and law enforcement in scientific and technological ethics governance. By participating in these training activities, governance talents can master the knowledge of legislation, judicature, and law enforcement in scientific and technological ethics governance and comprehensively improve the professional level of the ethical governance team. Third, attach importance to on-the-job training. Enterprises should establish a "pre-employment + scientific and technological ethics" training model, making scientific and technological ethics an important part of academic exchanges and job tasks. Require governance talents to strictly follow the requirements of scientific and technological ethics to further improve the level of scientific and technological ethics review and supervision. Fourth, establish a talent incentive mechanism, optimize and improve the processes of professional title evaluation and promotion, fully stimulate the work enthusiasm of governance talents, and continuously improve the quality of talent cultivation from multiple aspects such as education, training, incentives, and market supply, so as to ensure the orderly and efficient development of scientific and technological ethics governance.

5. Conclusion and prospects

In summary, the scientific and technological ethics risks caused by artificial intelligence technology are the contradictions between society and technology. With the continuous development and innovation of technology, these contradictions are becoming more prominent, and scientific and technological ethics governance is becoming increasingly important. Therefore, we should recognize the importance of scientific and technological ethics governance and take measures such as the paradigm transformation of governance theories, the construction of a hierarchical governance system, the improvement of institutional norms of scientific and technological ethics, and the cultivation of professional talents for scientific and technological ethics governance to effectively prevent and address scientific and technological ethics risks and contribute to the sustainable development of human society.

Future research needs to pay further attention to: (1) The governance of existential risks brought about by the super-intelligence of general-purpose artificial intelligence (AGI); (2) New ethical issues triggered by the combination of quantum computing and AI; (3) The ethical expansion in scenarios of technology integration such as the meta-universe and brain-machine interfaces, and construct an inclusive and forward-looking governance system to ensure that artificial intelligence truly becomes a constructive force for promoting the progress of human civilization.

Disclosure statement

The author declares no conflict of interest.

References

[1] Chen Y, 2025, Scientific and Technological Ethics Governance: An Inevitable Requirement for Building a Strong Scientific and Technological Country, Wenhui Daily, March 2, 2025, (005).
[2] Hu X, 2025, The Reasons, Dilemmas, and Innovations of Scientific and Technological Ethics Governance. Practice and Theory of Special Economic Zones, 2025(01): 26–33.
[3] Xie X, He X, Weng X, 2025, Research on the Construction of a Digital Platform for Scientific and Technological

Ethics Governance from the Perspective of Digital Governance. Jiangsu Science & Technology Information, 42(03): 7–12 + 36.

[4] Sun Z, Lu Z, 2025, Research on Countermeasures for Artificial Intelligence-Assisted Scientific and Technological Ethics Governance. Anhui Science & Technology, 2025(01): 31–33.

[5] Liu X, Ma J, Li Z, 2025, The Transmission of Scientific and Technological Ethics Values and Governance Principles in Scientific and Technological Communication. Forum on Science and Technology in China, 2025(01): 162–168.

[6] Yao Y, Meng F, 2025, Governance of Digital Scientific and Technological Ethics Issues: Overseas Experience and China's Path. Leadership Science, 2025(01): 145–152.

[7] Gao Y, Jiang Q, 2024, Academician Mei Hong of the Chinese Academy of Sciences: In the Face of the AI Boom, Scientific and Technological Ethics Governance Must Keep Up. Science News, 26(06): 11–12.

[8] Chen L, Zou X, 2024, An Analysis of Scientific and Technological Ethics Governance in the Age of Artificial Intelligence. Special Zone Economy, 2024(11): 53–56.

[9] Zan S, Gu L, 2024, Scientific and Technological Ethics Issues and Their Governance in the Development of Artificial Intelligence, Guangxi Ethics Society, School of Marxism, Nanning Normal University, Guangxi Marxism Theory Research and Construction Engineering Base (Nanning Normal University). Proceedings of the 2024 Academic Seminar on "Educator Spirit and the Development of Educational Ethics," School of Marxism, Guangxi University of Science and Technology, 272–281.

[10] Zhao J, 2024, The Path of Integrating Scientific and Technological Ethics into the Artificial Intelligence Governance System: Taking the Application Scenario of Autonomous Driving as an Example. Rule of Law Society, 2024(05): 16–28.

[11] Lei H, 2025, On the Agile Governance of the Scientific and Technological Ethics of Generative Artificial Intelligence. Journal of North China University of Water Resources and Electric Power (Social Science Edition), 41(01): 50–58.

[12] Xue G, Yan R, 2024, Ethical Challenges of Disruptive Technologies and Their Responses under the Strategy of Scientific and Technological Ethics Governance. Science and Technology Management Research, 44(10): 205–211.

[13] Ruan R, Zhu Z, Chen W, et al., 2024, Government Governance of Scientific and Technological Ethics and the Good-Tech Orientation of Artificial Intelligence Enterprises. Studies in Science of Science, 42(08): 1577–1586 + 1606.

[14] Yu F, 2022, The Social-Psychological Orientation of Scientific and Technological Ethics Governance: Taking Artificial Intelligence as an Example. National Governance, 2022(07): 26–30.

[15] Wu Y, 2022, Guarding the Ethical Bottom Line and Advancing with Good-Tech, People's Daily, March 28, 2022, (013).