EU Big Data Ethics Governance Practices and Implications for China (Postprint)

Authors: Chen Yi

Date: 2023-04-01T16:15:49+00:00

Abstract

[Purpose/Significance] This study summarizes the experience of the European Union's big data ethical governance to provide a reference for China's big data ethical governance. [Method/Process] It systematically reviews the relevant policies of the European Union concerning big data ethical governance, analyzes the practices of EU big data ethical governance, with a focus on examining the governance measures of the European Economic and Social Committee (EESC). On this basis, it summarizes the EU's governance experience and, according to China's actual circumstances, derives implications for China's big data ethical governance. [Results/Conclusion] The EESC, starting from the human life cycle, summarizes big data ethics into ten main issues including "awareness" and "control," and proposes five governance initiatives from multiple levels of individuals, enterprises, and research institutions, while emphasizing the examination of big data ethical issues from multiple perspectives. Currently, the lagging legal norms for data ethics and the lack of citizens' ethical awareness in China pose substantial ethical risks to big data utilization. China should learn from the EU's governance experience, clarify individual rights in the big data environment, attend to the ethical needs of multiple stakeholders in the data governance process, construct a digital education system, and explore the establishment of a continuous review responsibility mechanism for researchers.

Full Text

The Practice of EU Big Data Ethics Governance and Its Enlightenment for China

Chen Yi School of Information Management, Wuhan University, Wuhan 430072

Abstract

[**Purpose/Significance**] This study summarizes the experience of EU big data ethics governance to provide reference for China's big data ethics governance. [**Method/Process**] The paper systematically reviews relevant EU policies on big data ethics governance, analyzes EU practices with a focus on governance measures by the European Economic and Social Committee (EESC), and summarizes EU governance experience. Based on China's actual conditions, it derives enlightenment for China's big data ethics governance. [**Result/Conclusion**] The EESC identifies ten major ethical issues in big data from the perspective of the human life cycle, including "awareness" and "control," and proposes five governance measures from individual, enterprise, and research institution levels while emphasizing multi-perspective examination of big data ethics. Currently, the lag in China's data ethics legal norms and the lack of citizen ethical awareness pose significant ethical risks to data utilization. China should learn from EU experience by clarifying individual rights in the big data environment, attending to multi-stakeholder ethical needs in data governance, constructing a digital education system, and exploring the establishment of a continuous review responsibility mechanism for researchers.

Keywords: big data; ethics; governance
Classification Number: B829
DOI: 10.13266/j.issn.0252-3116.2020.03.014

Big data is profoundly impacting all areas of society and industries. Leveraging big data to drive economic development, improve social governance, and enhance government services and regulatory capabilities has become a trend. As data has become a crucial strategic asset, the enormous social and economic benefits it brings inevitably lead to ethical issues such as illegal collection, dissemination, and misuse of data, including unauthorized acquisition and storage of personal data, data abuse, weakened control over data by data subjects, data monopolies, unfair data application, and biased data guidance. The application of big data and intelligent technologies also involves complex risks, such as illegal predictive calculations and identity theft. A 2018 review article in *Science* titled "How AI can be a force for good" stated that artificial intelligence is reshaping daily practices, interactions between individuals and professionals, and the environment, making it crucial to harness this power for good. Ethical norms play a key role in this process, ensuring that AI's potential is fully realized while minimizing risks. AI is driven by data and thus faces moral challenges related to data governance, including consent, ownership, and privacy, which AI may exacerbate.

In recent years, the EU's perspectives on "digitalization" have grown exponentially, demonstrating strong forward-looking awareness of ethical dimensions and extensively investigating big data ethics. With the deep development of big data and artificial intelligence, the EU has implemented increasingly strict

ethical norms. In April 2018, the EU released the policy document "Artificial Intelligence," which identified "ensuring the EU has an ethical and legal framework appropriate for AI development and application" as one of the three pillars of AI strategic development. The May 2018 implementation of the General Data Protection Regulation (GDPR) established ethical red lines and principles for the big data era. The "AI made in Europe" program's two key principles begin with "ethics by design," and the April 2019 release of "Ethics Guidelines for Trustworthy AI" further demonstrated the EU's emphasis on protecting individuals' right to choose and be informed when facing technology and data. Clearly, the EU has been at the forefront of global supervision over data and technology use ethics in the big data environment. This paper systematically investigates relevant EU policies on big data ethics governance, summarizes EU practices, and provides beneficial references for understanding ethical issues in China's big data development and utilization.

2. Domestic and International Research Progress on Big Data Ethics

2.1 Specific Manifestations of Big Data Ethical Issues

In September 2012, American scholars K. Davis and D. Patterson published the first international academic monograph on big data ethics, *Ethics of Big Data*, arguing that enterprises should establish moral norms, clarify data's value, and emphasize identity, privacy, ownership, and reputation in data, seeking balance between technological innovation and risk. A report by the Royal Society of the UK identified ethical concerns in surveys and public dialogues as privacy, freedom, autonomy, and self-determination, with some issues specifically targeting new technologies like machine learning. University of Washington law professor N. Richards argued that big data ethics should adhere to four principles: treating privacy as a data rule, maintaining confidentiality when sharing private information, implementing transparent mechanisms for data utilization, and recognizing that data can undermine identity. In 2016, Professor F. Luciano and Dr. T. Mariarosaria from the Oxford Internet Institute organized a special exploration of "what is data ethics," identifying three main areas: ethics of data (how to generate, record, and share data), ethics of algorithms (how AI, machine learning, and robots interpret data), and ethics of practice (designing responsible innovation and professional codes to guide this emerging science).

Chinese scholars Xue Fu et al. argue that while big data drives social progress, it creates new privacy ethics challenges different from the past, manifested in data mining, data prediction, and more comprehensive surveillance. Professor Qiu Renzong, focusing on digital identity, summarized ethical issues in big data technology innovation, research and development, and application, highlighting concerns about "digital identity," "personal privacy," "data accessibility," "data security," and "digital divide." An Baoyang believes that while big data brings positive changes to human society, it also triggers network information

ethics issues such as information alienation, data rights, information privacy, and digital divide.

2.2 Causes of Big Data Ethical Issues

Although big data has such enormous global impact, society, government, and legal systems have not truly adapted to the arrival of the big data era. Current academic consensus holds that data itself is an objective existence and big data is neither good nor bad in itself; the root of ethical problems lies in the unreasonable use of data by humans. The huge commercial value of big data drives the transformation of personal data into analyzable data to obtain more products and profits, while commercial involvement creates moral risks. The maturity and perfection of modern network information technology provide more free and unobstructed space for this profit pursuit, greatly increasing the practical difficulty of early warning and avoiding ethical risks.

Some scholars have specifically analyzed the causes of ethical issues from the negative effects of big data application and legal anomie, as well as the lag of ethics and morality. "Data needs management, but the problem is that people are gradually getting rid of management over data collection." Wang Xiaoyang believes that big data falls into ethical dilemmas for four main reasons: the absence of a big data worldview, blurred boundaries of data rights, imbalance in data market application, and short-sightedness in big data technology development, with the root cause being the weakening of human subjectivity and the loss of ethical boundaries in the big data context. Lu Weihua believes ethical issues arise from five major conflicts: conflicts between data collection/application and personal privacy, conflicts between big data management and information security, conflicts between changes in intellectual property concepts and existing protection, conflicts between big data information abuse and information crisis, and conflicts between big data information pollution and the information ecological environment. Zhao Sujin, from the perspective of interest subjects, believes that each individual's activity is a link in the entire high-tech activity chain, with almost everyone being a direct or indirect creator of ethical risk, placing the subject of ethical risk creation in an uncertain "drifting" state that makes it difficult to determine responsible parties and expands the harm of ethical issues.

2.3 Governance Strategies for Big Data Ethical Issues

The development of information technology has trapped ethical issues in dilemmas that are becoming increasingly complex, and scholars urgently hope to explore applicable ethical norms and standards. L. Judd proposed in 1995 that the information age should adhere to ethical and moral principles: appropriately assuming responsibility, anticipating negative impacts, and pursuing fairness guided by John Rawls' principles of justice. R. Severtson advocated using four principles to solve information ethics problems: respecting intellectual property rights, respecting privacy, fair participation, and non-maleficence.

Relevant academic organizations have also formulated ethical norms, such as the Institute of Electrical and Electronics Engineers (IEEE), the Association for Computing Machinery (ACM), the American Statistical Association, and the British Computer Society. Ethical rules have obvious positive significance in promoting industry self-discipline and policy formulation, regulating information personnel behavior, and helping people make ethical decisions, providing beneficial references for big data ethics governance.

In his book *Delete: The Virtue of Forgetting in the Digital Age*, Viktor Mayer-Schönberger argues that in an environment where forgetting has become the exception and memory the norm, we should keep what is meaningful and discard what is meaningless. A. Zwetter proposed development space in the legal and political fields. D. Nunan et al. suggested implementing three ethical practices: the right to be forgotten, data expiration rights, and social media image ownership to protect individual privacy control. Chinese scholar Yu Yingxiang believes we need to construct a comprehensive big data governance system covering macro-level strategy and planning, meso-level institutions, organizations, and processes, and micro-level software/hardware platforms, standards, regulations, and technologies. Song Jixin believes we should combine big data technology constraints, legal norms, and self-discipline education to improve netizens' network literacy and regulate their online behavior. Chen Wanqiu also believes we should establish a mechanism combining law, ethics, and supervision to regulate big data technology development.

2.4 Review of Research Progress

Literature review reveals that current research mostly provides general summaries of ethical issues in the big data environment. However, big data development and utilization is a dynamic evolutionary process with diverse data subjects, and different subjects have different cognitions of big data ethical issues. Current research has not considered stakeholders. Most governance strategies for big data ethics remain at the appeal stage and are mostly end-of-pipe governance, with unsatisfactory practical effects on regulating big data ethical issues. The EU has always demonstrated strong forward-looking awareness in technology application and ethics. This paper systematically reviews relevant EU policies and practical explorations on big data ethics governance, hoping to provide practical and useful references for China's big data ethics governance and establish guiding behavioral norms and preventive mechanisms to achieve proactive avoidance of big data ethical issues.

3. The Process of EU Big Data Ethics Governance

3.1 Evolution of EU Big Data Ethics Governance Policies

Data has become a key asset for economic development and social progress. While seizing opportunities provided by big data, the EU has not ignored its accompanying challenges. In recent years, the EU has released multiple strate-

gies and plans and enacted relevant policies requiring protection of individual rights when utilizing big data, with attention to data ownership, data security, and digital divide issues.

In September 2012, the European Commission released *Unleashing the Potential of Cloud Computing in Europe*, planning to promote member states' utilization of cloud computing potential between 2014-2020 while stating that the Commission would promptly address general network security challenges and adopt appropriate technical and organizational measures to manage security risks and report significant incidents to competent authorities.

In February 2013, the European Commission released the European cybersecurity strategy *An Open, Safe and Secure Cyberspace*, emphasizing that European authorities should protect fundamental rights in cyberspace—democracy and rule of law. The EU must promote "online freedom" and ensure "respect for fundamental rights online" through authorized access, establishing democratic and effective multi-stakeholder governance systems, and sharing responsibilities.

In March 2013, the European Commission launched the Digital Skills and Jobs Coalition, aiming to improve digital skills for all citizens, enabling everyone to participate in the digital society; enhance labor force digital vocational skills; strengthen career advice and employment guidance; improve high-level vocational skills for ICT professionals; and transform digital skills education systems to establish lifelong learning systems. This plan can effectively improve EU citizens' digital literacy and bridge the digital divide.

From a more pragmatic perspective, the European Commission approved the implementation of the Horizon 2020 research program in December 2013, with "big data" as its core, providing direct funding support for big data-related research and encouraging innovative exploration of big data ethics issues.

In November 2015, the European Data Protection Supervisor wrote Opinion 7/2015, discussing the risks and challenges of big data itself, calling for transparency, user control, and data protection, and emphasizing that big data development and utilization must depend on four elements: organizations must process personal data in more transparent ways; users can better control their personal data; design more user-friendly data protection in products and services; and assume responsibility for data development and utilization.

On May 25, 2018, the General Data Protection Regulation (GDPR) officially came into effect. This strictest EU data protection bill emphasizes personal data rights, no longer treating "individual rights" as one principle of data protection but using an entire chapter to separately regulate it, establishing eight individual rights including the right to be informed, right of access, right to rectification, and right to erasure, enabling people to more effectively control their data.

On April 8, 2019, the European Commission's High-Level Group on Artificial Intelligence (AI HLEG) released *Ethics Guidelines for Trustworthy AI*, aiming to maximize AI's role while minimizing risks. It stipulates that AI system

development and utilization should adhere to principles of doing good, doing no harm, preserving human agency, being fair, and operating transparently.

3.2 EU Big Data Ethics Governance Practices

The EU has also designed specific governance measures for big data ethical issues. In June 2016, the European Economic and Social Committee (EESC) launched a big data ethics investigation that began in July and concluded in December. The investigation clarified what specific issues big data ethics includes and formulated corresponding checks and balances. In the big data environment, various subjects may be both data producers and collectors/users, and data governance should achieve diversified coordination. The EESC also recognized that this issue must be examined from multi-stakeholder perspectives.

3.2.1 Ten Issues Clarifying specific manifestations of big data ethical issues is the first step toward governance. The EESC sorted out ethical dilemmas people encounter from before birth to death from the perspective of the human life cycle and summarized them into ten major issues:

1. **Awareness**: When creating digital identities, people unconsciously provide much personal data, and when using digital identities to access third-party resources, they may simultaneously reveal much personal behavioral data without knowing it, losing necessary informed consent rights.
2. **Control**: When users decide to partially or completely delete data they previously provided to service providers, even if providers comply, data already sold to other companies or processed in large batches will not be affected. People have no actual control over their own data.
3. **Trust**: This is the foundation for users' willingness to provide personal data, interdependent with privacy and awareness. However, people have not yet established trust relationships with computer environments, currently relying mainly on strict technical means to solve trust problems.
4. **Ownership**: This mainly refers to complex ownership relationships facing processed original data. How should original user data be handled after processing? Does it remain user data, or should it belong to the company performing the analysis or the company collecting the data? Ownership disputes exist among users, data analysis companies, and data collection companies, with no more practical legal and policy solutions currently available.
5. **Surveillance and Security**: With increased data sources and analytical technology progress, it has become very convenient to use data to generate valuable information, making it easy to track and monitor someone or infer someone's position using certain methods.
6. **Digital Identity**: People can conveniently use digital identity to obtain network services, but the widespread use of digital identity also makes

individuals' publicly available information widely searchable. People often evaluate someone based on data provided by digital identity, which can cause discrimination—"we no longer judge based on behavior but judge behavior based on data." Interactions between people are usually also placed after analyzing digital identity.

7. **Tailored Reality**: Keywords people search through search engines or search preference data from online shopping may be stored, and analysis and processing of this data can be used later to provide relevant information based on user preferences. Users are experiencing more personalized but narrower online experiences, i.e., "filter bubbles."
8. **De-anonymization**: Traditional anonymization technology mainly makes each entry unidentifiable by deleting (or substituting) uniquely identifiable information, but this is already ineffective today. Various data association analyses can generate powerful insights that can still identify certain individuals to some extent.
9. **Digital Divide**: Some users cannot obtain services through new technologies like the Internet due to lack of relevant information skills, or have difficulty truly understanding how these processes work due to unfamiliarity with them, such as the frequent difficulties elderly people face when job hunting.
10. **Privacy**: Privacy is an all-encompassing topic covering the connotations of the above ethical issues, mainly referring to people's right to have their personal information not used by others without permission, which is currently the most important ethical issue.

3.2.2 Five Measures Based on systematic summarization of big data ethical issues, the EESC proposed five governance measures from multiple perspectives to comprehensively cover the above ten issues and achieve effective governance from individual, enterprise, and research institution levels:

1. **EU Privacy Management Platform**: The idea is to establish a pan-European portal website as a privacy management center. European citizens voluntarily register and have a personal data control center where they can visually see lists of all public and private entities that store, process, and reuse their data and can control them autonomously. Precisely because of the "unconsciousness" and "lack of control" over data collection in the big data environment that repeatedly violates privacy, this platform directly grants citizens the right to control their personal data and provides convenient opt-out services. Users can decide whether to share and reuse personal data, effectively enhancing information security. This is also consistent with the GDPR principle emphasizing that natural persons should control their personal data.
2. **Issuing a Data Ethics Management Protocol**: This aims to improve

transparency, making people understand the degree to which big data holders (public and private) comply with EU law. The concept is to design a comprehensive European certification system for enterprise certification in the data protection field. Enterprises certified for data ethics demonstrate respect for law and citizens' rights, can better gain user "trust," and further enhance corporate reputation and economic benefits.

3. **Issuing a Data Management Statement**: This is based on voluntary enterprise participation, starting from the fact that organizational success increasingly depends on multi-party trust from shareholders, customers, employees, and the public. Therefore, to enhance confidence of internal and external stakeholders, organizations can submit statements on "how they collect, use, or sell data from customers or general business activities." These must clarify data nature, future uses, usage strategies, and explicitly explain encryption and anonymization technologies. From the public perspective, the Data Management Statement can improve transparency of personal data utilization, prevent excessive data collection by institutions, and restrict unethical use of data or user digital identity by third parties, strengthening subject control over data.

4. **Establishing a European Electronic Health Database**: The normative utilization of citizens' health data is a key area for ensuring big data ethics. This measure plans to establish a European database containing health data of European citizens for scientific research. When EU citizens receive publicly funded medical treatment, they are asked whether they agree to have their data collected and stored in this database. With individual consent, data is collected and transmitted according to standard exchange protocols. Scholars and research institutions must submit applications to use the data, including specific information about researchers and projects, reasons for data application, utilization methods, and expected results. Approved scholars or institutions receive anonymized and coarsened data (to a degree that prevents de-anonymization). Meanwhile, citizens can use their legal digital identity information to access the database, manage and change their personal medical data. This standardized health data utilization approach can greatly reduce the probability of illegal collection and utilization of medical data, effectively promote scientific research, increase European citizens' control over their own medical data, enhance trust in medical health services and research, and enrich European citizens' digital identity information.

5. **Constructing a Digital Education System for the Big Data Era**: This initiative aims to establish broader digital culture in Europe, enabling European citizens to have deeper understanding of big data and how it interacts with them throughout the life cycle and affects everyone. To enhance this awareness, education programs should be designed for different age groups. Primary to high school education should mainly introduce big data, including basic concepts, what digital identity is, how

data is collected, and how to avoid excessive disclosure of personal privacy. University degree education should mainly cultivate relevant professional talents through systematic learning of ethical methods for creating and using big data. In communities, different topics should be held according to different age groups, especially for the elderly, to improve residents' digital skills. Constructing a digital education system can effectively improve European citizens' digital literacy, narrow the digital divide, provide clearer cognition of the big data environment, enable more rational cognition of information provided by the network environment, and also increase acceptance and trust in new technologies by having some prediction of the risk that "enjoying digital services may require sacrificing some privacy."

3.2.3 Multiple Perspectives Different individuals, organizations, or institutions are at different positions in the data chain and have different perspectives and cognitions of big data ethical issues. To obtain broad opinions, the EESC conducted an online questionnaire survey and selected 16 representatives from data protection agencies, big data analysis companies, SMEs, consumer associations, EU institutions, academia, and data science experts for in-depth interviews. While all expressed concerns about the current state of big data ethics, their specific attitudes differed. For example, academia and national data protection agencies mainly focused on general issues related to reasonable data use and transparency, believing that overly personalized digital experiences might create bias, while two data experts remained optimistic about data utilization and emphasized big data's potential to generate new insights.

Interviewees also held different attitudes toward the five measures. For instance, data experts believed that issuing the Data Ethics Management Protocol and establishing a certification system might increase corporate burden, while the only company manager among interviewees did not consider this an additional burden. Instead, he believed it was a good tool to raise user awareness of their company's data usage. Regarding establishing a European electronic health database, while most interviewees recognized it would greatly help medical care and research, they did not consider it currently feasible, mainly due to concerns about sensitive data anonymization and potential security. Interviewees were very consistent about the digital education plan, believing this measure plays an important role in the information society and should be implemented as soon as possible.

3.3 EU Big Data Ethics Governance Experience

In recent years, while using big data to create value, the EU has always paid attention to ethical issues. Rather than just making appeals, compared to other "case-based" summaries of big data ethical issues, the EU's 梳理 is more comprehensive, implementing governance measures from multiple dimensions and forming valuable experiences.

3.3.1 People-Centered Approach Big data ethical issues are ultimately related to "people." People even begin participating in the digital world before birth and continue providing and utilizing data in different ways and at different awareness levels throughout various life cycle stages. From the emphasis on "democracy," "transparency," and "individual rights" in various EU policies, to the EESC's identification of ethical issues based on the human life cycle and implementation of specific measures to protect personal privacy data, to AI HLEG's emphasis on "privacy and data governance" and "diversity, non-discrimination, and fairness" to address ethical needs from different positions, and the European Commission's communication on April 8, 2019 with the Parliament, Council, European Economic and Social Committee, and Regional Committee about establishing "human-centric" AI trust—all demonstrate concern for "people" and emphasis on individual rights.

3.3.2 Multi-Perspective Governance Measures In the big data environment, various subjects may be both data producers and collectors/users, with different cognitions of big data ethical issues and different needs for regulating big data ethics. Surveys show that 46% of British adults believe they are "harmed by data collected by large companies," while another study found that 60% of people are "satisfied with the amount of personal information shared with companies," and 47% believe "exchange of personal data is crucial for the smooth operation of modern society." The EU's multiple governance measures start from individual, enterprise, and research institution levels, considering different subject positions, with particular attention to the health data field facing severe ethical issues. At the first AI Alliance Conference held in Brussels in June 2019, AI HLEG also expressed the need to tailor ethics policies for the market and promote cooperation among social, industry organization, public sector, and academic stakeholders, moving beyond single, one-way government regulation to multi-stakeholder collaboration to effectively advance governance.

3.3.3 Emphasis on Digital Education The EU pays close attention to improving residents' digital literacy. The Digital Skills and Jobs Coalition initiated by the European Commission aims to improve digital skills for all people. The EESC advocates constructing a digital education system for the big data era, with community linkages through various types of training to give residents of all age groups more comprehensive cognition of the digital society and cultivate data ethics awareness. Improving cognition of the digital society can effectively improve citizens' participation levels, thereby providing more available data for governments and private companies in an ethical context that respects citizen dignity, promoting healthy development of the big data environment.

4. China's Current Situation and Enlightenment

4.1 Current Status of Big Data Ethical Issues in China

The powerful tension of big data makes all fields pay more attention to society's dependence on big data, emphasizing data's social and economic value. However, frequent occurrences of illegal data acquisition and storage, data abuse, weakened control over data by data subjects, data monopolies, and biased data guidance have made some people begin to pay attention to data ethics issues. However, Chinese citizens have not yet formed clear cognition of big data ethics. According to another survey by the author, among 470 literature pieces using social media user data for research, only one directly mentioned ethics-related issues in the text. Researchers rarely realize potential ethical concerns, yet much of this data already involves user privacy. Researchers' data collection, use of data for prediction and correlation analysis, and data storage have not obtained users' informed consent. China's education system lacks systematic training in this area, leading to serious insufficiency in citizens' relevant cognition and knowledge reserves, which is highly likely to cause benefits to accrue to individuals while risks are left to society, bringing enormous technological risks. This shares commonalities with international research findings: M. Zimmer's survey showed that only 4% of literature using Twitter data for research mentioned ethical issues or related factors concerning research design and data collection methods. In contrast, this problem is even more severe in China, which also reflects the urgency of improving Chinese citizens' data literacy.

From a policy perspective, China officially implemented the *Cybersecurity Law of the People's Republic of China* on June 1, 2017, but this is a principled framework that cannot comprehensively cover data issues in today's complex environment and should be supplemented with many supporting systems and regulations. On January 23, 2019, the Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, and State Administration for Market Regulation jointly issued the *Announcement on Launching Special Governance of APPs' Illegal and Irregular Collection and Use of Personal Information* to regulate personal information collection and use. On January 25, 2019, during the 12th collective study session of the Political Bureau of the CPC Central Committee, General Secretary Xi Jinping further pointed out the need to "standardize data resource utilization and guard against risks brought by new technologies such as big data," and Premier Li Keqiang has repeatedly emphasized that "information abuse, privacy infringement… should be cracked down on and cleaned up according to law." On July 24, 2019, the ninth meeting of the Central Committee for Comprehensively Deepening Reform reviewed and approved the *Plan for Establishing a National Science and Technology Ethics Committee*, and China will accelerate the establishment of a science and technology ethics review and risk assessment system. Clearly, China has fully recognized the importance of ethics construction in technological development, but relevant normative construction has just begun, and there is still a long way to go in regulating big data ethical issues.

The lag in China's data ethics regulation and the lack of citizen ethical awareness undoubtedly bring enormous ethical risks to data utilization, and this "absence" will also constrain future development. In view of this, China must accelerate the construction of specific ethical norms in big data utilization and emphasize the cultivation of citizens' data ethics awareness.

4.2 Enlightenment for China

4.2.1 Clarify Individual Rights in the Big Data Environment As the EESC survey shows, lack of relevant "awareness" is one of the most urgent issues. Currently, individuals present a "vague" state regarding data collection, dissemination, and utilization processes. The lack of clear individual rights leads to insufficient control over their own data, while they may also infringe on others' privacy when using data themselves. The GDPR implemented in May 2018 no longer treats "individual rights" as one principle of data protection but discusses it in a separate chapter, demonstrating the importance of individual rights. "Enabling people to control their data" has become one of the key objectives of data ethics governance in some European countries. China has not yet enacted special legislation for personal information protection, let alone clearly defined what data rights individuals enjoy in the big data environment. China should accelerate the formulation of the *Personal Information Protection Law*, reposition the attributes and protection models of personal information, clarify the individual rights system in the big data environment, and define the scope and content of rights in detail, enabling the public to understand what operations they can perform on their own data in the big data environment, improving control over personal data, and enhancing data ethics awareness among data providers and users. At the same time, we should gradually improve legal norms for classified data information protection, clarify rights and responsibilities in data mining, storage, transmission, release, and secondary utilization, and emphasize that personal information protection is not just about setting rights for individuals but aims to construct a legal framework balancing individuals, information users, and social interests.

4.2.2 Attend to Multi-Stakeholder Ethical Needs In the big data environment, everyone simultaneously plays multiple roles in the data chain, with different data needs at different development stages and different cognitions of big data ethical issues. The above survey results show internal conflicts in people's attitudes toward big data, and the EESC survey also reflects different attitudes of different stakeholders toward big data ethics governance. The UK's Minister for Internet Safety and Security pointed out: "There is a big gap in people's views on data trust. Teenagers give up privacy and are happy to share everything in their lives. But we also have a responsibility to look after their interests, especially their online rights." Different cultures and groups have different concepts of privacy, and privacy boundaries are set differently for different environments and objects. When facing big data ethical issues, China cannot adopt a "one-size-fits-all" approach. We must prevent "excessive development"

but also avoid "excessive protection." The need to protect personally identifiable data must be balanced with the possibility that this data may be used to create public good. According to China's actual conditions, we should consider public value orientation, respect diverse ethical needs, and balance benefits and ethics as much as possible in data development and utilization.

4.2.3 Construct a Digital Education System People are the main users of technology and the key to triggering ethical issues. Professor M. Jonathan, Chair of the UK Nuffield Council on Bioethics, also stated: "Most violations are human, not technical." Talent construction is key. In 2018, the UK invested an additional £406 million in mathematics, digital, and technical education. China should learn from EU experience, increase investment in data education, and design different curriculum systems for different targets. In compulsory education, we should focus on popular science education, enabling the next generation to master basic big data knowledge and establish correct "data ethics" concepts. At the university level, we should establish systematic curriculum systems and training plans to cultivate professional data talents. Social education should also link with school education, leveraging libraries' social education functions to hold special lectures to popularize data utilization knowledge, guide and regulate data dissemination and utilization behavior, and improve acceptance of the digital society and data literacy of social members.

4.2.4 Establish a Continuous Review Responsibility Mechanism for Data Users In addition to beneficial experiences worth learning from, some dilemmas faced by EU big data ethics governance in practice also provide references. The European electronic health database advocated by the EESC aims to avoid ethical issues by standardizing data utilization, reducing non-standard data utilization from the source, but has not been recognized. The main reason is the risk of data leakage in the database. Once researchers are authorized to use certain health data, they may permanently save the data on their computers, which may be intentionally or unintentionally used illegally or beyond its originally designated purpose in the future, potentially triggering public opinion that resists research. While encouraging data open utilization, China should advocate establishing a continuous review responsibility mechanism for data users. In addition to detailed review of data users' qualifications and usage purposes during the data application stage, during the process, regular explanations of data utilization should also be required, especially when there are changes from the data utilization model specified in the application, requiring resubmission and detailed explanation. Moreover, rights and responsibilities must be clarified—data users must be responsible for the reasonable use and storage of authorized data. Once illegal events occur, they must bear full responsibility and be prohibited from applying to use data again.

References

[1] MARIAROSARIA T, LUCIANO F. How AI can be a force for good [J]. *Science*, 2018, 361(6404): 751-752.

[2] FLORIDI L, TADDEO M. What is data ethics? [EB/OL]. [2019-09-14]. https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2016.0360.

[3] Liang Zhiming. Big data governance: The proper meaning of national governance capability modernization [J]. *Journal of Jishou University (Social Sciences Edition)*, 2015, 36(2): 34-41.

[4] An Baoyang. Research on network information ethics governance in the big data era [J]. *Science Research*, 2015, 33(5): 641-646.

[5] European Economic and Social Committee. The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU Policy context [EB/OL]. [2019-08-14]. https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data.

[6] European Commission. Artificial Intelligence [EB/OL]. [2019-08-10]. https://ec.europa.eu/digital-single-market/en/artificial-intelligence.

[7] General Data Protection Regulation [EB/OL]. [2019-11-14]. https://gdpr-info.eu/.

[8] European Commission. AI made in Europe [EB/OL]. [2019-08-09]. https://ec.europa.eu/digital-single-market/en/news/ai-made-europe.

[9] Ethics Guidelines for Trustworthy AI [EB/OL]. [2019-08-09]. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[10] DAVIS K, PATTERSON D. Ethics of big data [EB/OL]. [2019-03-10]. https://www.oii.ox.ac.uk/news/releases/what-is-data-ethics/.

[11] The Royal Society. Data governance: Public engagement review [EB/OL]. [2017-6]. https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-public-engagement-review.pdf?la=en-GB.

[12] RICHARDS N, KING J. Big data ethics [J]. *Social Science Electronic Publishing*, 2014, 1(2): 1-6.

[13] Oxford Internet Institute. What is data ethics? [EB/OL]. [2019-03-10]. https://www.oii.ox.ac.uk/news/releases/what-is-data-ethics/.

[14] Xue Fu, Chen Hongbing. Exploration of big data privacy ethics issues [J]. *Studies in Dialectics of Nature*, 2015, 31(2): 44-48.

[15] Qiu Renzong, Huang Wen, Zhai Xiaomei. Ethical issues of big data technology [J]. *Science and Society*, 2014, 4(1): 36-48.

[16] MOOR J. The future of computer ethics: You ain't seen nothin' yet! [J]. *Ethics & Information Technology*, 2001, 3(2): 89-91.

[17] Hao Tingshuai. The big dataization of contemporary social life: Dilemmas and reflections [J]. *Social Development Research*, 2014, 1(3): 196-211.

[18] Chen Shiwei. Ethical governance of big data technology alienation [J]. *Studies in Dialectics of Nature*, 2016, 32(1): 46-50.

[19] Shanghai Academy of Social Sciences. Finding the ethical balance point in "world dataization" [EB/OL]. [2019-03-20]. http://www.sass.org.cn/Default.aspx?tabid=167&ctl=Details&mid

[20] Wang Xiaoyang. The turn of network ethics governance in the big data context [J]. *Today's Media*, 2015, 23(6): 25-26.

[21] Lu Weihua. Research on information ethics in the big data era [J]. *Modern Information*, 2014, 34(10): 66-69.

[22] Zhao Sujin. Ethical risks and avoidance in the high-tech era [J]. *Seeking Truth*, 2011, 52(4): 30-32.

[23] JUDD L. An approach to ethics in the information age [J]. *Public Relations Review*, 1995, 21(1): 35-44.

[24] SEVERSON R. The principles of information ethics [M]. New York: M. E. Sharpe, Inc., 1997.

[25] IEEE. IEEE ethics and member conduct [EB/OL]. [2019-02-25]. http://www.ieee.org/about/ethics.html.

[26] ACM. ACM code of ethics and professional conduct [EB/OL]. [2019-02-25]. http://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct.

[27] Ethical guidelines for statistical practice [EB/OL]. [2019-02-25]. http://www.stats.uwo.ca/faculty/bellhouse/sta9544_{{{2016}}_{{files}}}/ASA%20EthicalGuidelines.pdf.

[28] Code of conduct for BCS members [EB/OL]. [2019-02-25]. http://www.bcs.org/upload/pdf/conduct.pdf.

[29] BERLEUR J. Ethics and the governance of the Internet: Recommendations of IFIP-SIG9.2.2 [M]// Computers and networks in the age of globalization. Boston: Springer, 2001.

[30] Mayer-Schönberger V. *Delete: The virtue of forgetting in the digital age* [M]. Translated by Yuan Jie. Hangzhou: Zhejiang People's Publishing House, 2013.

[31] ZWITTER A. Big data ethics [J]. *Big Data & Society*, 2014, 1(2): 1-6.

[32] NUNAN D, DOMENICO M. Market research and the ethics of big data [J]. *International Journal of Market Research*, 2013, 55(4): 505-520.

[33] Yu Yingxiang. Review of big data skepticism [J]. *Archives Management*, 2015, 32(1): 17-19.

[34] Song Jixin. Research on network ethics [M]. Beijing: Science Press, 2012.

[35] Tang Xiran. Ethical issues and moral philosophy of big data: A review of the first National Cyber Ethics Seminar [J]. *Ethics Studies*, 2015, 13(2): 138-140.

[36] Unleashing the Potential of Cloud Computing in Europe [EB/OL]. [2019-03-24]. https://www.pdpjournals.com/docs/88053.

[37] An open, safe and secure Cyberspace [EB/OL]. [2019-03-24]. https://ec.europa.eu/home-affairs/what-is-new/news/2013/20130207_{{{01}}_{{en}}}.

[38] The Digital Skills and Jobs Coalition [EB/OL]. [2019-03-24]. https://ec.europa.eu/digital-single-market/en/digital-skills-jobs-coalition.

[39] Opinion 7/25 Meeting the challenges of big data [EB/OL]. [2019-03-24]. https://edps.europa.eu/sites/edp/files/publication/15-11-19_{{{big}}*{{data}}}{en}.pdf.

[40] SWAN M. Philosophy of Big Data: Expanding the Human-Data Relation with Big Data Science Services [EB/OL]. [2019-08-14]. https://www.melanieswan.com/documents/Philosophy_

[41] Big data-privacy principles under pressure [EB/OL]. [2019-03-25]. https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf.

[42] FRED R. The Filter Bubble: What the Internet is hiding from you (review) [J]. *Portal: Libraries and the Academy*, 2011, 11(4): 1009-1011.

[43] European Commission. Communication: Building trust in human-centric artificial intelligence [EB/OL]. [2019-08-11]. https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence.

[44] UK Public Research-Online Privacy [EB/OL]. [2019-03-25]. https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf.

[45] The Direct Marketing Association [EB/OL]. [2019-03-26]. http://data.parliament.uk/writtenevidence/com and-technology-committee/big-data-dilemma/written/20453.pdf.

[46] European Commission. EU artificial intelligence ethics checklist ready for testing as new policy recommendations are published [EB/OL]. [2019-08-11]. https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are.

[47] Li Xia. Science and technology ethics: Unconstrained technology is dangerous [N]. *Guangming Daily*, 2015-07-31(10).

[48] ZIMMER M, PROFERES N J. A topology of Twitter research: disciplines, methods, and ethics [J]. *Aslib Journal of Information Management*, 2014, 66(3): 250-261.

[49] Announcement on Launching Special Governance of APPs' Illegal and Irregular Collection and Use of Personal Information by Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, and State Administration for Market Regulation [EB/OL]. [2019-08-14]. http://www.cac.gov.cn/2019-01/25/c_{1124042599}.htm.

[50] Sun Min. Guard the "gate" of science and technology ethics [N]. *Xinhua Daily*, 2019-08-13.

[51] Xinhua Net [EB/OL]. [2019-02-21]. http://www.xinhuanet.com/fortune/2017-11/21/c_{1121987357}.htm.

[52] GOV.UK. Data Protection Act 2018 [EB/OL]. [2019-03-26]. https://www.gov.uk/government/collections/d protection-act-2018.

[53] Yang Weidong. Effectively respond to ethical issues of big data technology [N]. *People's Daily*, 2018-03-23(7).

[54] GOV.UK. AI Sector Deal [EB/OL]. [2019-03-26]. https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.

[55] The big data dilemma [EB/OL]. [2019-03-25]. https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.

[56] Industrial Strategy: Building a Britain fit for the future [EB/OL]. [2019-03-26]. https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future.

Note: Figure translations are in progress. See original paper for figures.