

# Research on the Personal Privacy in the Health Big Data

Cangcang JIA<sup>1,a</sup>, Han LIU<sup>2,b</sup>

<sup>1</sup>School of Government, Nanjing University, Jiangsu, China

<sup>2</sup>School of Mechanical Engineering, Shaanxi University of Technology, Shaanxi, China

**Abstract**—Data mining, data prediction and all-round digital monitoring of health big data make the dilemma of personal privacy control prominent. The weakening of the control of personal privacy by big data technology, the people's data belief, the diversity of interests and the conflict of interests are the main causes of the problem of personal privacy in the context of the application of health big data. Therefore, in the application of health big data, the solution to the problem of personal privacy in the context of health big data application is to enhance the value transparency of big data technology, return and reshape humanism, and explore common values to reduce conflicts of interest.

## 1 Introduction

With the in-depth application of Internet of things, big data, artificial intelligence, 5g and other new generation information technologies in the field of health industry, health data sets covering electronic medical records, electronic health archives, clinical diagnosis and treatment records, disease monitoring, medical experiments, biological samples and gene sequences, etc., namely health big data, have been produced. The transformation of health industry from "small" data to "big" data has broken through the narrow space of medical institutions, doctors and patients in the past. Therefore, the big data of health has become an important basic strategic resource of the country, which is highly valued by the government. In order to actively create a social environment for the safe application and innovative development of health big data, our government has issued a number of policy documents. In practice, the construction of big data of health in China is also in full swing. Health big data provides residents with prevention and rehabilitation guidance, enterprise drug research and development, and social public health governance services. At the same time, it integrates with medical institutions into clinical decision support system to provide more accurate, personalized and efficient medical services for the public [1].

The application of big data in health plays a fundamental role in deepening the reform of health system, promoting the construction of Health China and the high-quality development of health industry. However, it will completely subvert the existing medical service mode, bring a longstanding structural reconstruction to the field of human social life, and trigger a comprehensive life style, work style and thinking mode of residents. The change even directly threatens the personal health information and privacy security, which is also the key

reason for the prudent development of health big data application.

The existing literature has fully discussed the issue of personal privacy in big data, but there are relatively few discussions on the issue of personal privacy in the health big data, and most of them are based on the perspective of ethics. In addition, some scholars pointed out that the application of health big data faced personal privacy risks such as tracking and monitoring, data disclosure and discrimination tendency, and elaborated privacy protection rules, technical standard formulation and legal policies based on the perspective of government behavior [2]. In the context of healthcare data application, if we do not strengthen the protection of personal privacy, it will cause serious losses to the country, organizations and individuals. The paper intends to explore the new challenges faced by personal privacy in the context of health big data application, and analyze the causes from three dimensions of technology, society and reality and propose practical solutions.

## 2 Privacy in the health big data

### 2.1 Connotation of personal privacy in the health big data

Warren & Brandeis (1890) regards the right of personal privacy as a kind of power of "solitude", "freedom from external interference" and "freedom from infringement" [3], and explains the ideas of the protection of the right of personal privacy by law. However, their understanding of the right of personal privacy is limited to the passive defense of individuals from the outside world, which is not applicable to the information society [4]. In the information society, the right of personal privacy attaches more importance to the control of personal information,

<sup>a</sup>DG1906020@smail.nju.edu.cn, <sup>b</sup>1065398771@qq.com

and the information subject has the initiative decision-making power and control power over personal information. When an individual does not want to be disclosed by information held by others, his privacy right is violated. In the context of health big data application, the right of personal privacy refers to the right that government agencies, medical institutions, insurance companies, pharmaceutical enterprises and their relevant personnel cannot disclose the legitimate information of citizens at will and exclude the illegal invasion of citizens' privacy information when citizens receive health promotion services, as well as the right to give citizens control and the right to use personal information. It can be seen that under the background of health big data application, personal privacy has the dual meaning of passive non invasion and active self-control.

## **2.2 New challenges to personal privacy in the health big data**

Based on the ternary world of "people, machines and things" and its interaction in cyberspace, more health big data has been generated, which provides new development impetus and innovation vitality for the economic society. However, due to the complexity of data type, structure and internal model, fast data generation speed, low value density and dynamic changes of personal privacy with many factors, the application of health big data poses new challenges to personal privacy, which are shown in the following:

(1) Data mining: secondary use of personal privacy. In the context of the application of health big data, relevance replaces causality. Through the data mining of "fragmented" health data of the public, the health information that does not seem to exist originally begins to "appear". This process is quite opaque. Non-rival makes the health big data showing the characteristics of quasi-public goods. Any organization and individual can find valuable knowledge and rules for decision-making, and every time they use the data, they can be saved in the original form. This unique charm makes the data mining of health big data full of possibilities. Each data mining may discover new knowledge, including personal privacy. Because of this, personal privacy is likely to be used by stakeholders continuously.

(2) Data prediction: predicted personal privacy. The core of big data is prediction, that is, data algorithms and models are applied to massive data to predict the possibility of events. "Cambridge analysis event" is a famous case of using big data to predict personal privacy. One of the main purposes of the application of health big data is to predict the onset of disease in order to prevent and treat in time. However, this kind of analysis and prediction based on health big data can insight into people's everything, and the individual basically has no privacy to speak of, so the human nature crisis is bound to be born.

This kind of human nature crisis is manifested in that the stakeholders use the predicted privacy information to screen and classify the residents, so that some people have a lower probability of gaining benefits, or suffer discriminatory treatment and unfair treatment, or even

suffer humiliation and loss of dignity. A classic case of predicting personal privacy by using health big data is "speculate on teenage pregnancy and disclose it to businesses". This represents a weakening of personal control over privacy, and is also the root cause of residents' general concern about the threat of personal privacy in the application of health big data.

(3) Full range of digital monitoring: no place to hide personal privacy. We have to admit the following facts: the deep change caused by big data technology in the field of medical and health industry makes people enter the "super panoramic prison", and every resident is always in a state of continuous panoramic intelligent monitoring and real-time computing. This fact challenges the necessity of personal privacy in the context of healthcare big data application, especially the omni-directional digital monitoring sometimes reflected in the will of the state. All these highlight the omni-directional digital monitoring, which makes the massive health and medical data resources concentrated in the hands of a few people, and personal privacy becomes nowhere to hide.

Therefore, as a basic right of residents, the right of personal privacy must be examined from a new perspective. In the context of the application of health big data, the right of personal privacy is not going to die out, but becomes more related to the establishment of people's identity, equality, security and trust, and has a real impact on the political, economic, social, cultural and ethical basis of the country. In the context of healthcare big data application, personal privacy is not a pure secret, but an intermediate state between secret and full disclosure. The focus of attention should be on how to ensure that personal privacy is "not interpreted" based on technology, and on how to have the ability to identify the "interpreted" results of personal privacy.

## **3 CAUSES OF PERSONAL PRIVACY IN THE HEALTH BIG DATA**

### **3.1 Technical cause: the weakening of the control of personal privacy by big data technology**

Big data technology weakens residents' control of personal privacy. The premise of personal privacy disclosure is to obtain personal health information accurately, comprehensively and timely, which cannot be achieved before big data technology. However, the deep integration of big data technology and medical and health industry makes it easier to collect, share, analyze and predict residents' health and medical data, and also makes residents' control over their own health and medical data be quietly taken away by data managers, who can easily grasp personal privacy by using big data technology to analyze health and medical big data; and the discovered personal privacy has changed from "invisible" to "explicit", and the risk of personal privacy disclosure has increased rapidly.

In addition, big data technology makes the principle of informed consent under the control of personal privacy

virtually exist in practice. The principle of informed consent forces the data collection subject to obtain the informed consent of the collection object in the data collection stage in order to obtain its health data. In this sense, personal privacy is controllable - the parties do not authorize it. Under the support of big data technology, the principle of informed consent has lost its efficacy. On the one hand, the individual's informed consent often has no choice but to do something - without authorization, they can't enjoy the necessary basic health services; on the other hand, the data mining of health big data will produce many unexpected and innovative uses, and on these innovative uses, there are hundreds of millions of data students. It is not practical and expensive for the producer to obtain informed consent. At this time, the data manager often decides the application of health big data by himself.

### **3.2 Social cause: belief in data**

"Dataism" is likely to become a belief of residents in the context of health big data application. Data theory not only reconstructs the functional relationship between "health big data and life existence" at the phenomenal level, but also endows "health big data processing" with a kind of ultimate practical significance at the ontological level. In this way, the belief of data theory indicates the change of residents' world outlook from "human centered" to "data centered". Human identity is in the virtual network. The space is digitally reconstructed.

In extreme cases, the belief of data doctrine will eventually make people become the representation of data, that is, people are more willing to use objective health data to know someone, and ignore the factors such as individual's inner self shaping and expectation, as well as the personality, emotional taste, cultural value preference in human nature. This will seriously affect the human nature and subject status, and it is a kind of alienation of the belief of data theory. This is the social cause of the dilemma of personal privacy control in the context of the application of health big data. The belief in data theory has become a force to suppress and dominate people.

### **3.3 Realistic causes: interest diversity and interest conflict**

Health big data has multiple values, and has potential value for other stakeholders, including business organizations and public sectors, except data producers. The application and development of health big data cannot be separated from specific organizations and individuals, but each subject is utilitarian oriented when using health big data. For example, residents expect to keep their privacy while enjoying the benefits brought by the application of health big data; business organizations focus on the economic benefits brought by health big data, and ethical awareness and humanistic care are placed under the economic benefits. The public sector pays more attention to the social value and public interest of health big data application for the health promotion of all citizens. The diversity of the value of health big data and the diversity of the interests among the subjects cause

conflicts of interest, among which the division of private interests and public interests may lead to the fuzzy control of personal privacy, which is the root of the dilemma of personal privacy control.

## **4 Strategies for personal privacy in the health big data**

### **4.1 Enhance the value transparency of big data technology in the health big data**

Although scholars hold two views on the duality of Technology Value: value neutrality and load value, this paper agrees that technology is load value, that is, technology activities should not only follow objective laws, but also have specific value orientation. Generally, when stakeholders use big data technology to analyze and process health big data, whether in the process of product provision or service provision, there will always be positive and negative impacts - positive and negative values appear at the same time or successively, but the public knows nothing about it or knows later, so it is necessary to enhance the application of big data technology in health big data. Value transparency in data applications.

Understand, recognize and respect the public's fear of the unknown, clearly inform the public of the definite and potential uses in the application of health and medical big data, so that the public can master the value orientation of big data technology in the application of health and medical big data. Return the right of choice to the public so that the public can understand the risks and benefits they need to bear, which is in line with the three basic principles of informed consent, equality and respect, and no harm and benefit in the application of health big data, and it is also conducive to reducing the risk of big data technology in the analysis and prediction of health big data.

### **4.2 Return to and reshape humanism in the health big data**

The basic logic of the formation of data belief is the ontology and epistemology of health big data. The ontological problem is that the public claims that the health big data collected is a "full sample data", but this is a beautiful lie, because even if all the data of human social life are digitized, and want to collect these data is beyond the range of human reason, so fundamentally speaking, the health big data is not comprehensive; and even if collected "It is impossible to understand and grasp the real human society with the whole sample data, because the relationship between the phenomena presented in the analysis of health big data is a statistical correlation, a kind of "weak correlation", which challenges the most basic way of thinking for human society to understand the reality and make decisions, and a kind of blasphemy for free will. The problem of epistemology is that people believe that the health big data collected is "metadata" which is not polluted by human subjective consciousness. However, it is impossible. Any data is generated in a

specific cultural system and influenced by the system values and ideology, so any data has more or less human subjective impression. In addition, the belief in data doctrine also tries to use the neutrality of technological value to ensure the objective neutrality of data collection and processing, but this argument is still untenable, because technology, as a product of human society, will inevitably bear the values of human society.

In order to get rid of this dilemma, we should try to eliminate the people's data belief, return and reshape humanism. Humanism advocates that human freedom is higher than data freedom, human value is higher than data value, and it also strengthens the freedom and privacy rights of mediators [19]. The return and reconstruction of humanism will reestablish the dominant position of human in the era of big data of health, respect human's basic rights and data rights, reconstruct the free relationship between human and data, maintain human freedom and increase human welfare.

#### **4.3 Mining common values and reducing conflicts of interest**

In fact, if every resident has the right to refuse others to use their health data, then the individual's prudent and decentralized choice will limit the formation of highly inclusive health data sets and the ability to capture special samples, and ultimately will not be able to effectively collect the data resources used to promote public health improvement and other people's health promotion and collective action for public health promotion. The inherent logic dilemma of personal privacy control and collective action results in that the privacy protection method of individualism cannot be adopted in the context of health big data.

In the era of big data, the boundary between private domain and public domain of human society is developing in the direction of benefiting public domain. Therefore, in the application of health big data, we should strive to combine individual value and common value, excavate the common value orientation of stakeholders, and produce products and services that meet the common value, so as to reduce the stakeholder's reasons when personal privacy is involved in the action Conflicts of interest arising from the diversity of interests. In view of the inevitable conflict between personal privacy control and public interest in some aspects, we often give priority to ensure the realization of public interest. In order to meet the above requirements, we need to rely on stakeholders to focus on finding value consistency on the

basis of open dialogue and mutual cooperation, and to reach a certain consensus on the issue of personal privacy.

## **5 Conclusion**

In the context of health big data application, the dilemma of personal privacy control is prominent. In order to further promote the development of health big data application, we need to find a balance point between innovation and risk, so as to avoid the problem of personal privacy, respect personal freedom and strive to enhance the consistency of value and action, which is conducive to the high quality development of health big data. Of course, to solve the problem of personal privacy in the context of health big data application is not an overnight success, but also requires good systematic and sound legal policies.

## **Acknowledgment**

About the author: Cancang JIA (1994-), male, Han nationality, native of Xi'an, Shaanxi Province, doctor degree candidate, research direction is social medicine and health management. Han LIU (1998-), male, Han nationality, native of Xi'an, Shaanxi province, bachelor degree candidate, research direction is Mechanical Design manufacture and Automation Major. Contact address: 21 Department, Nanjing University, Gulou District, Nanjing.

## **References**

1. Xinxin GUO, Haiyan WANG. Business Model Health Data Sharing Platform Based on Data Crowdsourcing in the Context of Big Data [J]. *Business Review*, 2019, 31(7): 56-64.
2. Xing LIU, Xiaomin WANG. Ethical Issues in the Construction of Medical Big Data [J]. *Studies in Ethics*, 2015, 80 (6): 119-122.
3. Samuel D. Warren, Louis D. Brandeis. The Right to Privacy [J]. *Harvard Law Review*, 1890, 4(5): 193-220.
4. Richard O. Mason. Four Ethical Issues of the Information Age [J]. *MIS Quarterly*, 1986, 1(1): 5-12.