Proceedings of the 4th International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/2026.33929

Practical Dilemma and Improvement Suggestions of Personal Information Protection under the Background of Big Data Algorithm

Yeyang Jiang

School of Economics and Law, College of Science & Technology Ningbo University, Ningbo, China
jyysuccessful@outlook. com

Abstract. Nowadays, big data algorithms have been deeply embedded in daily life, which makes the sharp contradiction between the rapid development of artificial intelligence and the invisible threat faced by personal information protection, which is the price behind the improvement of life convenience in the era of artificial intelligence. This issue has been widely discussed by scholars from all walks of life. However, due to the ambiguity of the legal category of personal information, the attribution of the information after the anonymous processing of the algorithm, and the nature of the black box of the algorithm, the notification and consent rule is lacking in operability in reality, the signing of privacy policies is more formalistic, and the algorithm indirectly identifies sensitive information to avoid legal constraints. This paper analyzes the real challenges of big data algorithms in the field of personal information protection. This paper concludes that the deep-seated reasons are the inconsistency of definitions, the complexity and confidentiality of the algorithm black box, the formality of the privacy policy, and the uncontrollability of the algorithm's indirect identification. Based on this, this paper puts forward the following suggestions: first, unify the scope provisions of different legal frameworks on personal information, and issue the corresponding judicial interpretation. The second is to introduce the risk stratification notification system and the dynamic consent system. The third is to integrate different dimensions to define sensitive information, and incorporate the inference combination into the evaluation of sensitive information.

Keywords: Artificial Intelligence, Big Data Algorithm, Personal Information Protection, Personal Information Protection

1. Introduction

The technological dividend of AI is accompanied by privacy risk: it not only improves the convenience of life, but also unconsciously erodes the protection of personal information. For example, big data will collect the location information, browsing records and consumption records left by users on the network, and will piece together accurate portraits based on various hidden information. The portraits will gradually shrink until they are defined as individuals. Then all kinds of information pushed and commercial promotions came one after another, which greatly disrupted

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.33929

normal life. Another example is that the Internet platform, with the help of big data technology, carries out differentiated pricing for the same product or service according to the user's personal information, which does not conform to the price law and normal commercial pricing, and carries out "big data ripening", which greatly infringes the legitimate rights and interests of consumers [1]. The improvement of artificial intelligence technology has brought severe challenges to the definition of personal information categories, the improvement of legal provisions, and the implementation of specific protection practices.

In 2021, China officially promulgated the *Personal Information Protection Law of the People's Republic of China* (hereinafter referred to as the Personal Information Protection Law), which stipulates the rights and interests related to personal information. The first paragraph of Article 4 points out that personal information does not include information processed anonymously, which means that the information processed anonymously by artificial intelligence has exceeded the scope of personal information. However, in the face of the algorithm black box, the principle of "informed consent" required by Articles 13 to 17 and the obligation to inform in a clear and understandable manner are facing difficulties in implementation. At the same time, the algorithm indirectly identifies the behavior of sensitive information by combining non-sensitive information, which essentially circumvents the provisions of Articles 28 to 29 requiring separate consent for the processing of sensitive information. Therefore, through the research method of combining normative analysis and comparative research, this paper puts forward the practical dilemma of its implementation, and puts forward corresponding solutions, to provide some reference for the Personal Information Protection Law to better deal with the impact of artificial intelligence.

2. The practical dilemma of personal information protection caused by artificial intelligence

2.1. Fuzzy protection scope of personal information

The impact of the AI big data algorithm, a new technology, poses a challenge to the definition of personal information.

Up to now, the scope of personal information has been defined in three laws, including the *Personal Information Protection Law, the Network Security Law of the People's Republic of China* (hereinafter referred to as the "*Network Security Law*"), and the Civil Code [2]. By analyzing and comparing the provisions of the three laws on the scope of personal information, it is found that there are some differences. *The Network Security Law* emphasizes the identification of personal identity, and the Civil Code emphasizes the identification of specific natural persons, both of which are called "identification theory". The "*Personal Information Protection Law*" has added the "Relevance Theory" related to natural persons.

In the existing theories, the identification theory and the relevance theory are the two most mainstream theories. Identification theory claims whether information can directly or indirectly identify natural persons; Relevance theory emphasizes that information is included in the scope of protection as long as it is related to identifiable individuals. The common point of the two is that they both judge from the perspective of individuals. The difference is that the identification theory focuses on the ability of information identification, and the correlation theory focuses on the dynamic relationship between information and individuals. In the era of artificial intelligence, two theories are facing new difficulties: on the one hand, AI's powerful analytical ability makes the previously irreversible anonymous information recognizable again; On the other hand, the algorithm combination infers that sensitive information circumvents the provision of separate consent of sensitive information. If it do not unify the definition of personal information in the three laws as

DOI: 10.54254/2753-7048/2026.33929

soon as possible, more and more personalized portraits and personalized recommendations will be able to avoid legal constraints [3]. Whether the same information belongs to the category of personal information under two different legal frameworks has different conclusions, which is obviously an urgent problem to be solved.

2.2. Algorithm black box makes it difficult to effectively implement the rule of informed consent

An algorithm black box means that the operation of the algorithm is like a black box; only the results of input and output can be seen. The intermediate process is not only complex, but also will not be disclosed to the public as a trade secret. For example, the algorithm only recommends videos of interest to users, but how to judge them is unknown. Another example is that the algorithm will score and recommend takeout, film and television. What is the basis and formula behind the scoring is also unknown. The existence of an algorithm black box also makes it difficult for users to be held accountable.

Article 17 of the *Personal Information Protection Law* imposes extremely high requirements on the language used by information processors to inform individuals. However, due to the complexity and confidentiality of the algorithm black box and the cyclic use of the processed information, it is difficult to effectively implement Article 17 in practice.

First, tell users that it is difficult to be clear and understandable. The operation mechanism behind generative AI is extremely complex. What users can see is only simple input and output. How the intermediate process runs concretely cannot be shown to users [4]. Even developers know nothing about the intermediate process, and it is impossible to tell individuals in clear and understandable language. In other words, even without considering the nature of trade secrets, enterprises agree to disclose and show the running code to the public, which still has high barriers to non-professionals.

Secondly, the user's informed consent is reduced to form. In view of the requirements of Article 17, personal information processors usually use the privacy policy to inform [5]. The lengthy privacy policy seems clear and definite, but in fact, it puts the key information on the last page. Users are often eager to use it and sign the informed button when it is difficult to read the full text. What's more, if some app users do not sign and agree to the privacy policy, the app will not be able to be used normally, so the user has to agree. The Diablo model claimed by Harrybriginu is incisively and vividly displayed in the carefully designed notification form of the information processor. "Bundled consent", "inadvertently checked consent", "default checked," and other forms simply cannot represent the true psychology of users [6]. In name, it meets the principle of informed consent, but in fact, it becomes formalism.

In addition, it is inevitable to use it for other purposes after consent. Article 14 it is required to obtain personal consent again after the purpose of personal information processing is changed. Data is the basis of an algorithm, and an algorithm is the essence of data. The personal information processed by artificial intelligence has become one of the huge amounts of data, which will inevitably provide pre-training and reference for thousands of different outputs. It is obviously unrealistic to tell individuals every time. On the contrary, if the individual is not informed every time, he will return to the package authorization, which violates the provisions of Article 17.

2.3. Indirect recognition of sensitive information by algorithm portrait

The algorithm infers sensitive information through the combination of non-sensitive information, which may avoid the legislative intent of Articles 28 and 29 of the *Personal Information Protection*

DOI: 10.54254/2753-7048/2026.33929

Law. The law only requires personal consent for processing sensitive information, but it does not explicitly prohibit or restrict the behavior of combination inference. For example, users frequently browse health content, purchase specific drugs, and search keywords at specific times - these three types of general information are not sensitive information alone, but after the organic integration of the algorithm, the sensitive information of users' health status can be inferred. The current law lacks clear provisions on such situations, which is one of the legal challenges brought by the development of artificial intelligence technology.

This has to talk about a milestone case: the New York Times AOL dug clues from search terms in 2006. Search "Ga gardener" and "community house sales" to locate Lilburn, Georgia. Search "Arnold", "60-year-old bachelor", and "numb fingers" many times, pointing to a 62-year-old widow, keeping a dog, and paying attention to health. Through cross-validation of public information, Selma Arnold was locked. Called on the door, causing industry shock. The party said: "My whole life has been stripped away. I never knew that someone was staring at me [7]."

Another example is the Cambridge Analytica scandal exposed in 2018. With the help of the psychological test app, the British company illegally obtained the data of 87million Facebook users, analyzed the political tendencies of voters, and affected Trump's election campaign. This incident directly triggered the worldwide privacy legislation reform, and this large-scale abuse of data to manipulate politics became a typical case that caused a sensation all over the world [8].

The above two cases have strongly proved that the combination of non-sensitive information infers that sensitive information not only violates personal privacy, but may even manipulate politics, which once again rings the alarm for legislators. Algorithm inference is not a one-time behavior, but becomes powerful in continuous data accumulation. At first, it was just a vague group portrait. With the continuous enrichment of data, the blurred image pixels gradually increased until individuals could be locked. This process makes "when constitutes sensitive information identification" an uncertain issue. Which link constitutes sensitive information identification? If it focuses on the final identification results of sensitive information, it means that the inference process of the algorithm is not constrained, but the fact is that there is a problem in this process. If it is attributed to the information collection stage, the platform is required to foresee in advance, which requires too much technology and is difficult to achieve. Professor Dingxiaodong pointed out that the scope of sensitive personal information itself is uncertain. Much seemingly non-sensitive personal information can also be inferred from sensitive personal information. This uncertainty is dramatically magnified in the scenario inferred by the algorithm [9].

3. Suggestions on the improvement of enabling personal information protection

3.1. Unified personal information range

Based on the definition of personal information in Article 4 of the Personal Information Protection Law, it can discuss how to keep consistent with the basic definition when interpreting the relevant concepts in the *Network Security Law, the Civil Code*, and other laws in the future, so as to help improve the authority of the *Personal Information Protection Law* in the personal information protection standard system.

As for the expression "related to natural persons" in this article, it may be necessary to further clarify its judgment standard through judicial interpretation or guiding cases in judicial practice. Because the legal provisions are generally macro and there is a certain ambiguity in terms, different courts may have different understandings of "relevance", and then there will be different judgments in the same case. Therefore, when formulating a unified reference rule for handling cases, it is worth

DOI: 10.54254/2753-7048/2026.33929

considering how to define the "relevance" between information and natural persons. For example, it can discuss such a judgment idea that if a certain information cannot be independently identified or associated with a specific natural person without the assistance of other information, the information may not be classified as personal information. Such refined criteria help Judges should correctly apply the law when dealing with new situations.

3.2. Introduction of risk stratification notification system and dynamic consent system

Some scholars put forward the scene integrity theory. The core point of this theory is that it emphasizes that the processing of personal information must be limited to the scene of obtaining information. The authorization obtained in this scene should not be used in the scene replaced later. The risk of the replaced scene is different from that of the original scene. If the authorization is continued, it is obviously impossible to guarantee the reasonable expectation at the time of the original authorization [10].

To this end, it can introduce the risk stratification notification system and dynamic consent system. The platform is required to classify different scenarios and set different warning levels for different risks. For the processing of high-risk personal information, such as sensitive information, the "forced pop-up system" is adopted, and users' attention is increased with the setting of a vibration alarm. The most important information in the user privacy agreement shall be marked in bold red letters and placed in the most conspicuous position of the privacy policy for attention. For low-risk personal information processing, such as general information, users can only be informed by SMS. If users are troubled by the frequent notification of low-risk personal information, they can choose the flexible design option of not receiving notifications about low-risk personal information, and they can choose whether to receive the notification.

Paragraph 2 of Article 14 of the *Personal Information Protection Law* stipulates that when the personal information processor changes the purpose, method and type of processing, it shall apply to the user for authorization at any time. On this basis, it should also set up a classification, early warning and user-independent selection mechanism. Users are allowed to choose which categories need to be notified. This not only helps to avoid frequent pop-up windows of authorization interfering with the user experience, but also avoids the problem of excessive information collection caused by package authorization.

3.3. Beware of indirectly identifying sensitive information

3.3.1. Four-dimensional comprehensive consideration to define sensitive information

Considering the identity of the information subject, the processing subject, the information processing method and the recognition probability. Different information subjects, such as minors, the elderly, the disabled and other vulnerable groups, have weak risk defense and are more likely to be inferred as sensitive personal information. Different processing subjects, such as small companies and large enterprises, have significant differences in the processing of browsing information. Different recognition types have different effects on sensitivity; for example, cosmetics companies and food companies have different sensitivities to the same information. Cosmetics companies are obviously more sensitive to users' browsing of cosmetics. The greater the possibility of information being re-identified, the stronger the sensitivity. Only by integrating the four factors can define sensitive information more scientifically. To prevent common information from being misjudged as sensitive information, and to avoid generalizing sensitive information. Better realize high risk, high

DOI: 10.54254/2753-7048/2026.33929

prevention and control, low risk and low regulation. It not only considers privacy and security, but also does not hinder the rational circulation and use of information.

3.3.2. Incorporate inferred portfolio into sensitive information assessment

To redefine sensitive personal information, it should focus on the purpose interpretation, supplemented by the textual interpretation. Specifically, whether it infringes human dignity and personal property is taken as the substantive standard for judgment, and the enumerated definitions, such as biometrics, are taken as the formal standard for reference and clarification. In this way, the sensitive information derived from the combination of artificial intelligence inference can be included in the legal regulation to prevent the original general information from circumventing the legal provisions in the enumeration definition.

4. Conclusion

In the era of artificial intelligence, big data algorithms bring severe challenges to personal information protection. China's *Personal Information Protection Law* does provide a legal framework for its protection, but this paper finds that there are still loopholes in the delimitation of the scope and the effective implementation of the rules. Data is the basis of an algorithm, and an algorithm is the essence of data. Artificial intelligence can be more powerful under the training of a large amount of data, and there is inevitable tension between the development of artificial intelligence and personal information protection. Only by constantly improving and optimizing the relevant legal provisions can achieve a win-win situation between personal information protection and the development of artificial intelligence.

References

[1] Jin, Y.F., Wang, K.L. and Zhang, T.H. (2021) From the Perspective of Personal Information Protection Law, the Legal Regulation of 'Big Data Killing'. Journal of Zhejiang University of Science and Technology (Social Science Edition), 46, 693-701.
[2] Wang, Y. (2021) The Legal Dilemma and Improvement Measures of the Protection of Personal Information Rights. Young People, 21, 75-77.
[3] Wang, L.M. and Ding, X.D. (2021) On the Highlights, Characteristics and Application of the Personal Information Protection Law. Jurist, 6, 1-16+191.
[4] Huang, K. (2024) The Challenge and Risk Regulation of Generative AI to Personal Information Protection. Modern Law, 46, 101-115.
[5] Fan, M.Z. and Lu, Y.C. (2024) On the System of Personal Information Protection Rules in Algorithm. Digital Rule of Law, 1, 93-111.
[6] Yang, W.Q. (2024) The Institutional Logic, Normative Interpretation and Reinforcement of the Consent Rule for Sensitive Personal Information Disclosure. Financial Law, 1, 100-115.
[7] Barbaro, M., Zeller, T., & Hansell, S. (2006). A face is exposed for AOL searcher no. 4417749. New York Times, 9(2008), 8.
[8] ur Rehman, I. (2019). Facebook-Cambridge Analytica data harvesting: What you need to know.
[9] Ding, X.D. (2025) The Interpretive Reconstruction of Sensitive Personal Information Protection System. Chinese Law, 2, 168-187.
[10] Nissenbaum, H. (2004) Privacy as Contextual Integrity. Washington Law Review, 79, 119-148.