International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 4, Jul-Aug 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Legal Risk Identification and Classification Regulation of Data Collection for Household Robots

Zhang Xin-qi

School of Law, Beijing Wuzi University, Beijing, China

ABSTRACT

With the deep integration of domestic robots and artificial intelligence, the traditional online data standardization mechanism is inadequate. Combined with the recent practice of family information leakage, it is urgent to bring the collection of family data into the legal orbit. The core goal of this paper is to take the data classification and protection system established by the Data Security Law as a theoretical tool, systematically identify the legal risks of data collection in the process of domestic humanoid robots providing services, construct a data classification and classification list suitable for embodied intelligence scenarios, and put forward targeted legal regulation paths and compliance suggestions.

KEYWORDS: *Data security Law; Data classification and grading; Household robot; Scene completeness theory.*

How to cite this paper: Zhang Xin-qi "Legal Risk Identification and Classification Regulation of Data Collection for Household Robots" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-4, August 2026, pp.421-426, URL: www.ijtsrd.com/papers/ijtsrd142031.pdf

IJTSRD142031

I. Research background

At the 2026 CCTV Spring Festival Gala, the humanoid robot became the focus of public discussion with its amazing group dance performance, marking the official entry of the household service humanoid robot from the laboratory concept into the public view and the fast track of industrialization. Since then, domestic humanoid robots such as Yushu Technology H1, Youbixuan Walker X and Zhiyuan Expedition have sprung up, the capital market has continued to rise, and the industry has shown an outbreak of "in the ascending-one hundred machines competing".

From a technical point of view, humanoid robots belong to a type of "embodied intelligence". Different from the artificial intelligence software that has completely penetrated into our lives, embodied intelligent robots are a kind of physical entity as the carrier, In short, embodied intelligence is a product that gives AI an entity so that it can have an impact on the physical world. At the same time, the subject of embodied intelligence has become one of the professional disciplines set up in many colleges and universities[1], and deeper theoretical research on it will also become a guide for future practical development Key factors.

From the perspective of daily life, the concept of embodied intelligence is not a new word. Many products that we are familiar with and frequently use have the characteristics of embodied intelligence, which can be said to be the predecessor of embodied intelligence. For example, sweeping robots, voice assistants with whole-house intelligence, intelligent door locks, etc. These early home smart devices provide the development foundation and direction for today's embodied intelligence, but also expose the relevant legal risks.

[1] In 2026, Shanghai Jiao Tong University, Zhejiang University, Xi'an Jiaotong University, Harbin Institute of Technology, Beihang University, Beijing Institute of Technology, Nanjing University of Aeronautics and Astronautics, Beijing University of Posts and Telecommunications, and Northeastern University officially began to admit undergraduate students majoring in embodied intelligence. As an interdisciplinary field, they will confer a bachelor's degree in engineering.

II. Literature review

Regarding the legal and ethical risks caused by data collection by domestic robots, it is generally believed that the traditional data protection rules based on cyberspace are inadequate when facing the intrusive collection of the physical world by embodied agents. The existing legal system, such as the core principles of "informed consent" and "minimum necessary" established in the ***Personal Information Protection Law***[2], is mainly designed for online or static data processing activities. However, the "embodied" nature of domestic robots enables them to deeply integrate into the private space of the home. The continuous data collection of domestic robots is hidden, which makes it difficult for users to give effective consent in a rational and undisturbed way, and may even be guided or manipulated without knowing it, making the data control mechanism useless. As some scholars have pointed out, over-reliance on consent mechanism has become a "FIG leaf" for personal information protection, and users' control over information is seriously weakened before embodied intelligence. This shows that applying the traditional regulatory logic of cyberspace directly to the dynamic interaction in the physical world can no longer effectively respond to the complex challenges posed by domestic humanoid robots. Therefore, based on the data classification and classification protection system established by the Data Security Law, this paper attempts to systematically identify the legal and ethical risks of data collection in the process of domestic humanoid robots providing services, and puts forward targeted legal regulation paths and compliance suggestions.

III. Technical characteristics of data collection for household robots

The core characteristics of domestic robots can be summarized as "self-perception, self-learning, self-decision-making, self-adaptation, self-optimization" five capabilities, of which the most characteristic is the ability of self-perception and self-decision-making. Self-perception enables the robot to collect dynamic data of the environment, its own state and cooperative devices in real time through multi-source sensors such as vision and force sense. Self-decision-making enables it to autonomously plan tasks and generate optimal execution plans based on simulation preview and prediction models. Compared with the traditional artificial intelligence concept, embodied intelligent robot not only has the cognitive ability, but also has the interaction ability of physical entity and environment and the continuous evolution mechanism. Its intelligence comes from the closed-loop coupling of "ontology - environment - intelligence", which can truly realize the complete closed-loop of "perception - decision - action - feedback". In physical tasks, embodied intelligent robots show adaptability, autonomy and evolution capabilities beyond pure software AI.

At the technical level, home robots are often equipped with vision sensors. Such as RGB-D cameras, infrared cameras; Auditory sensors, such as microphone arrays; Tactile sensors and various sensing devices, such as light detection and Ranging (LiDAR), enable comprehensive monitoring of home environments and user activities. For example, through Simultaneous Localization and Mapping (SLAM) technology, the robot can use the fusion data of lidar and vision camera to construct a home layout map in real time, and accurately identify the room structure, furniture location, and obstacle distribution. At the same time, the vision algorithm based on deep learning can perform face recognition, object detection, and human pose estimation, so as to capture the user's identity, behavior habits and even emotional state. The auditory sensor continuously captures the voice interactions and life sounds of family members, and even identifies the user's identity through the voiceprint. In addition, force sensors enable the robot to sense contact strength and adjust actions when performing tasks such as nursing and grasping, but also record the user's physiological reactions and activity states synchronically. It can be seen that the data collected by domestic robots not only covers traditional personal information, but also extends to the physical environment of the home, user's biometric characteristics, behavior trajectory, emotional psychology and other elements. At the same time, the data collection is continuous and hidden, which increases the difficulty of legal regulation.

IV. Legal principles for identifying risks

1. Breakthrough in the principle of informed consent in the process of information collection

Different from traditional Internet platforms or mobile applications, the information collected by home robots in home scenes has several significant characteristics, which make the implementation of the

[2] Refer to Article 6 of the ***Personal Information Protection Law***. The processing of personal information should have a clear and reasonable purpose, and it should be directly related to the processing purpose. It should be carried out in a way that minimizes the impact on personal rights. When collecting personal information, it should be limited to the minimum scope necessary to achieve the processing purpose and excessive collection of personal information should be avoided. Article 14: When processing personal information based on the individual's consent, such consent should be voluntarily and clearly given by the individual after being fully informed.

informed consent principle face more complex situations.

Household robots have significant characteristics in information collection. First, the collection is embedded and continuous. The robot is deeply embedded in daily life, and the information collection behavior is uninterrupted and uninterrupted throughout the entire use cycle. Second, the information type is complex and hidden, which not only collects the information provided by the user actively, but also involves the passive collection of environmental data, such as room structure, activity trajectory, daily work and rest, etc., which may reveal highly sensitive privacy after aggregation. Third, the illusion of physical presence and trust. Robots exist in the most private space as physical entities, which makes users have the illusion of "natural security" and ignore the risk of information processing. Fourth, consent is bundled with services. If users refuse to consent, they cannot use the device, so the bargaining space is extremely limited.

These characteristics will eventually lead to a breakthrough in the principle of informed consent. The privacy policies of home robots are often lengthy and professional, and the information collection is "background". In addition, many data, such as spatial mapping, are collected in scenes that are not perceived by users, and it is difficult for users to match abstract terms with specific experiences. Some scholars have put forward the concept of "digital divide", which refers to the fact that information subjects have great difficulties in understanding the content provided by information processors. In order to avoid responsibility, information processors often use extremely abstract language to describe their user notification terms, and induce the consent of information subjects in technical means. In the home service scenario, many intelligent machine products are purchased to assist the elderly, children and other groups with low digital literacy in the family, and the practical effect of informed consent rules is almost overpowered.

The legal risk of the virtual nature of informed consent has moved from theoretical discussion to practical law enforcement. A case in point is the Federal Trade Commission's case against Apitor Technology, a company whose robot application violated the Children's Online Privacy Protection Act by collecting children's geolocation information without parental consent through third-party SDKS. This case fully demonstrates the ability of domestic robots to enter private Spaces and collect private information, and it has a greater impact on the informed consent rights of younger users. The case finally reached a settlement with Apitor Technology agreeing to delete the illegally collected information, modify the privacy policy and establish a parental consent mechanism. This solution provides a governance idea for the data collection problem that embodied intelligence is difficult to detect in private space. It also makes us feel excited about the rise of domestic robots, but at the same time, it sounds the alarm of legal risks.

2. The contradiction between the principle of minimum necessary and practical needs

Different from the impact on the informed consent principle, there is a fundamental contradiction between the information collection process of domestic robots and the implementation of the minimum necessary principle. First, the "continuity" of information collection is difficult to be compatible with the "minimization of information" requirement. The principle of minimum necessary forms substantial restrictions on information processing activities from multiple dimensions such as "minimum quantity", "minimum type" and "minimum frequency", but home robots, such as sweeping robots, need to work continuously and repeatedly collect environmental information to complete tasks. Its necessity runs through the entire use cycle, and the boundary of the minimum is difficult to be accurately delineated in practice. Secondly, the complexity of information types makes it difficult to apply the "minimum type" standard. Multiple sensors of home robots collect multi-modal data at the same time in a task, and any data may directly affect the quality of the final service. There is a contradiction between the "direct relevance" standard required by the minimum type, which makes it almost impossible to judge "which data types are necessary".

The root cause of the contradiction in this principle is the blurring of the line between the basic functions of the home robot and the extended functions. Manufacturers often package extended functions into the category of necessity for optimizing user experience, which makes the constraint of the minimum necessary principle overblown. Some scholars pointed out that it has to be admitted that the large amount, large-scale, high frequency and wide range of information processing in the digital age is conducive to mining the potential value of information, while strictly limiting the scope and impact of information processing undoubtedly hinders the play of information utility, and is not conducive to the development of big data technology and the maximization of public and private interests. However, in the field of home robots, a large number of voice and behavior data collected by manufacturers

on the grounds of optimizing experience are often beyond the scope necessary to realize basic functions. Whether the rationality of its development of science and technology can support the impact on basic principles is a question that must be discussed.

3. Home data risk from the perspective of scene integrity theory

Contextual integrity theory was proposed by American scholar Helen Nissenbaum in 2010. The core idea is that privacy protection depends on the appropriate norms in a specific social context. This theory provides a powerful analytical framework for understanding whether data collection constitutes a violation of privacy in domestic Settings. In short, what information should be disclosed and whether information flow norms are followed; When these constitutive norms are violated, it constitutes a breach of contextual integrity, and therefore an invasion of privacy. As an independent social context, the basic purpose of family is to maintain common life and care, which must be based on intimacy, trust and information sharing. Therefore, the family has special norms for information flow that are different from those in the public sphere. When this theory is introduced into the data collection and analysis of domestic robots, it can be found that the intervention of robots has fundamentally changed the daily family communication that people are used to. The sender of information is extended from traditional human interaction to machine equipment, and the machine is not the focus. The type of information covers multiple dimensions such as private space, private activity and private information.

This is not scaremongering. In addition to the "Apitor case" mentioned above, in February 2026, an overseas user originally wanted to use the PS5 controller to control his DJI ROMO sweeping robot. When the homemade application he wrote connected to the DJI server, he accidentally found that about 7,000 robot vacuum cleaners in 24 countries around the world responded simultaneously. Through this vulnerability, the overseas user could view the real-time video of the ROMO robot's camera, listen to the ambient sound through the onboard microphone, and even obtain the complete 2D room map generated by the device and the location of the device. In the standard expectation of family scene, the subject of access to family data should be limited to family members or explicitly authorized service providers. However, in the DJI incident, the information subject deviated seriously from the "unauthorized third party". By extracting the private token of his device, the developer gained direct access to the video stream of 7,000 devices around the world without entering any PIN code. This meant that anyone with access to the vulnerability could become a bystander in the privacy of a home, without the device owner's knowledge. This is the most central signal that privacy norms have been violated. From the perspective of robot users, some foreign scholars have conducted a survey of robot users based on the scene integrity theory. The evaluation results show that no matter how sensitive the scene is, participants always prefer to share data only with the users directly related to the data, which is the strictest sharing strategy among all options, which confirms that users have low tolerance for even minor privacy violations.

Based on the above analysis, the privacy invasion of home scenes by home service robots is essentially a reflection of the conflict between technical systems and normative expectations. The theory of scene integrity provides a systematic analysis framework for understanding this conflict. Based on the interpretation of privacy rights in the theory of scene integrity, it is necessary and feasible to construct the classification of data information in the family scene as the core premise of home service robot services.

V. Risk identification under the principle of data classification and classification

Firstly, based on the disassembly of home scenes, the data collected by domestic robots can be divided into the following levels and types. From the perspective of information attributes, the data can be divided into basic environmental data, such as indoor temperature, light intensity, user behavior data, such as sleep time, activity trajectory, biometric data, such as face, voice print, heart rate, and emotional psychological data, such as emotional state and social interaction mode. Among them, the basic environmental data is a necessary element for the service of the robot, which is inseparable from the route planning and sensing tasks. Even if the data is leaked, it is generally difficult to harm the privacy and security of the family, and should be classified as the fair use of the data, or "non-sensitive family data". However, user behavior data, biometric data, and emotional psychological data directly touch the sensitive boundary between user's physiological privacy and mental privacy, which are highly sensitive personal information. Attention should be paid to the implementation of the principle of informed consent when collecting, so as to avoid breaking the normative expectations of family scene information circulation.

Secondly, from the perspective of data aggregation level, a single data point may only constitute general personal information, but after continuous collection and cross-modal fusion, it is easy to stitch together a

complete "family portrait", including family member relationships, living habits, health status and even value preferences. The importance of this aggregated data set has gone beyond the scope of individual rights and interests, and may pose a potential threat to social order or public interests. For example, after the convergence of a large number of family sleep trajectories and indoor layout data, the population activity law of a specific community can be analyzed. And even expose the vulnerable links of critical infrastructure, and then fall into the scope of important data or even core data defined by the Data Security Law. The "Apitor case" repeatedly mentioned above is motivated by such considerations. In the United States, where personal information protection is more important, even sending a location from a child's toy is not tolerated by privacy protection concepts. In the US, children's data is the highest priority. After the Cambridge Analytica incident, American society is extremely sensitive to data breaches. For US regulators, Chinese-made smart toys that involve data repatriation or unauthorized collection could easily trigger the association of data espionage. In China, privacy protection is more regarded as part of social stability and network security, and at the same time in daily life, Chinese consumers are often willing to give up part of privacy for convenience such as intelligent interaction and personalized recommendation. But in the foreseeable future, before a large number of home service robots enter the family scene, it is a wake-up call for China's data protection, and it also enables the data classification and classification constructed by the "Data Security Law" to be implemented.

Starting from the theory of scene integrity, some scholars have further extended the privacy specification expectation in the home scene to the human-computer interaction scene, and extracted nine key sub-dimensions, namely location type. Robot engagement; Number of users; User age; User relationship; Data medium; The protected state of the captured scene; The subject of information flow transmission; Combined with the current research on this theory and the three-level protection system established by the "Data Security Law", the data security classification standard in the family scene can be constructed as follows.

Family core data is the category with the highest level of protection, which corresponds to "data related to national security, the lifeline of the national economy, important people's livelihood, and major public interests" as specified in the second paragraph of Article 21 of the Data Security Law. In the family scenario, such data is mainly reflected in the collection of family information that may affect national security after large-scale aggregation. When tens of thousands of family data are aggregated and analyzed, they may form macro information with strategic value, such as population health maps, regional activity patterns, and infrastructure vulnerability assessments. The aggregation analysis of a large number of household data can expose the vulnerable links of critical infrastructure, and even infer the population movement law in sensitive areas. Such data should be included in the scope of core data, and a stricter management system should be implemented.

Important family data is the second category of protection level, which corresponds to the scope of "key protection for data listed in the catalogue" required in the third paragraph of Article 21 of the Data Security Law. In the family scenario, important data should cover: family layout map and precise location information. Once such information is leaked, family members may face personal safety threats. Biometric information of family members, including unique identifying feature data such as faces and fingerprints; Continuous collection of private activity records of family members, such as bathing, toileting, intimate behavior images or audio. Even if the sensitivity of the context is low, the collection and sharing of these types of data should still be highly vigilant.

Household general data is the lowest level of protection and corresponds to other data that are not listed in the important data directory. This type of data includes basic environmental information, such as indoor temperature, humidity, light intensity and other data that do not directly identify an individual; Anonymized activity information of family members that cannot be recovered; Non-identifiable device operation logs; And information that the user has expressly consented to be disclosed. Research shows that users have a high acceptance of such data collection when it is necessary to maintain the basic function of the robot or complete a specific task. However, when the data collection is beyond the necessary scope or irrelevant to the core function, users will re-judge it as sensitive information.

VI. Conclusion

In conclusion, the construction of security classification standards for family scene data should be based on the three-level protection system established by the Data Security Law, and the nine privacy dimensions of scene integrity theory should be used as the analysis framework to form a classification standard with scene adaptability. This standard should not only follow the basic

requirements of legal norms, but also respond to the particularity of home scenes, and provide operational legal guidance for data compliance of smart devices such as household robots.

References

[1] Wang, Y. (2024). Privacy Risks of Embodied Intelligence Agents and Legal Countermeasures: A Research Based on Humanoid Robots. East Law Review, 3, 64–76. (in Chinese)

[2] Liu, Y. K., Wang, Q. J., Zhu, Z. L., et al. (2025). Embodied Intelligent Industrial Robots: System Architecture, Key Technologies and Case Studies. Computer Integrated Manufacturing Systems, 31(12), 4513–4541. (in Chinese)

[3] China Academy of Information and Communications Technology & Beijing Humanoid Robot Innovation Center Co., Ltd. Embodied Intelligence Development Report. (in Chinese)

[4] An, H., Cui, M., Zhang, P., et al. (2025). Research on Mobile Robot SLAM Based on Fusion of Laser and RGBD Camera. Computer Science, 52(12), 200–205. (in Chinese)

[5] Jiang, S. Y. (2023). Path Selection for Effective Implementation of Informed Consent Rules in Personal Information Protection. Journal of Information Security Research, 9(2), 154–161. (in Chinese)

[6] U.S. Department of Justice, Office of Public Affairs. (2025, September 30). Stipulated Court Order Resolves Allegations That Chinese Toy Maker Apitor Technology Co. Violated Children's Privacy Law.

[7] Jin, L. J., Zhai, Y. (2023). Review on the Minimum Necessity Principle in Personal Information Processing. Journal of Beijing Institute of Technology (Social Sciences Edition), 25(4), 140–150. (in Chinese)

[8] Jiang, Y. (2024). Research on the Application of Minimum Necessity Principle in Personal Information Processing [Master's thesis]. Central South University. (in Chinese)

[9] Nissenbaum, H. (2004). Privacy As Contextual Integrity. Washington Law Review, 1, 20–21.Southern News Network. (2026, March 9). DJI Sweeping Robot Leaks User Camera Footage? Latest Official Response. [Retrieved March 17, 2026].

[10] Sullivan, D., Zhang, S., Li, J., Kirkorian, H., Mutlu, B., & Fawaz, K. (2025). Benchmarking LLM Privacy Recognition for Social Robot Decision Making. arXiv preprint.

[11] Zhang, Y. (2025). Promoting Risk Governance of Personal Information in Household Robots. Procuratorial Horizon, 22, 22–23. (in Chinese)