Personal Data Circulation Mechanisms in Embodied AI: Balancing Privacy Protection and Data Value Release

Yue Ma[a]

Beijing Institute of Technology, Beijing, China

[a]15962969630@163.com

Abstract

Embodied AI, characterized by multi-sensor fusion and autonomous decision-making capabilities, has transcended traditional personal information protection rules, giving rise to multiple privacy risks including physical privacy and mental privacy concerns, and causing legal principles such as "informed consent," purpose limitation, and data minimization to face application difficulties. However, current academic research still primarily focuses on personal information protection as the main approach to regulating embodied AI, without systematically addressing the circulation of personal information-related embodied AI data. With the goal of balancing personal information protection and data value release, a regulatory philosophy distinguishing personal information from personal data should be established. Personal information should be transformed into circulatable data through gradient anonymization processing using technologies such as Differential Privacy (DP) and Federated Learning (FL), shifting from the user "informed consent" model to enterprise proactive prevention obligations for personal information leakage. Institutional safeguards should be constructed from dimensions including privacy tort liability delineation, introduction of Data Intermediaries, and regulatory mechanism innovation, to resolve compliance dilemmas, mitigate data leakage and re-identification risks, and provide theoretical support and practical guidance for compliant and efficient circulation of embodied AI datasets.

Keywords

Embodied AI, data circulation, Privacy-Enhancing Technologies (PETs), enterprise compliance.

1. Introduction

In the digital economy era, data has become a core production factor driving industrial upgrading. With the development of IoT and artificial intelligence technologies, data collection, processing, and analysis have gradually become industrialized, and the strategic value of data factors has become increasingly prominent. China's National Data Work Conference explicitly positioned 2026 as the "Year of Data Factor Value Release," marking the transition from the institutional construction phase to the practical phase of value creation and release, with the goal of promoting data supply, circulation, utilization, and security protection, and fully integrating data into economic value creation.

In January 2026, China's first "Embodied AI Dataset" was listed and traded on the Jiangsu Data Exchange, with 25,000 multi-modal structured data records achieving compliant circulation and commercial monetization, propelling the embodied AI industry from laboratory R&D to large-scale industrial applications. Compared with traditional industries, the embodied AI field more easily forms standardized, tradeable "datasets." Embodied AI emphasizes real-time interaction between intelligent agents and physical environments, forming a complete "perception-decision-execution" chain. Data collection accompanies task execution and can be

synchronized with data labeling, effectively breaking through the bottleneck of high cost, low efficiency, and insufficient standardization in physical world data collection, significantly enhancing the commercial value and tradability of embodied AI datasets.

Data differs from traditional resources in that its value lies in circulation. During the collection of personal information by embodied AI agents, due to their embodied characteristics, they inevitably challenge the informed consent rules, purpose limitation, and data minimization principles stipulated in the Personal Information Protection Law. Current legal scholarship on embodied AI mainly focuses on how to effectively protect personal information during the data collection phase. For example, Hua Jie, based on privacy typology theory and combining embodied AI application scenarios with risk classification, explores its impact on negative freedom and positive freedom types of privacy [1]; Zhao Jingwu points out that in AI intelligent agent scenarios, the "data minimization principle" is not failing but rather requires reconstruction of its application logic [2]; Li Zhi and Chen Yingying also indicate that embodied AI presents data privacy risks across technical, legal, and ethical dimensions, facing legal protection dilemmas such as imbalanced rights-obligations-responsibilities allocation, adverse selection by stakeholders, and large model technology governance challenges [3]. However, beyond the personality rights requiring protection, the commercial value embodied in personal information indicates its enormous commercial potential for market circulation. The construction of high-performance AI models cannot proceed without large-scale, high-quality datasets. Nevertheless, the principles established by the current Personal Information Protection Law all aim to achieve protection through restricting personal information collection, without forming an institutional foundation to promote large-scale circulation of personal information-related datasets, resulting in direct conflict between these institutional arrangements and AI's practical needs for data circulation and sharing. In this regard, how to facilitate data circulation channels between different entities and achieve alignment between personal information protection rules and data circulation needs in embodied AI scenarios represents an urgent theoretical and practical challenge.

2. Particularity of Personal Information Protection in Embodied AI Scenarios

2.1. Ontological Characteristics Derived from Embodied AI Technical Features

Traditional Disembodied AI functions like an isolated "brain," operating primarily through abstract symbolic computation and logical reasoning, capable only of symbolic operations in virtual environments. Embodied AI, in contrast, equips this brain with a "body" enabling entry into the physical world, providing the capability for real-time interaction with real environments and enabling learning and task completion through perception and action [4]. Embodied AI possesses the following characteristics distinguishing it from disembodied AI:

First, Embodiment. Multi-sensor Fusion technology constitutes the material foundation of embodiment. Current embodied AI systems are typically equipped with multi-modal sensing devices such as RGB-D cameras, force sensors, and IMUs (Inertial Measurement Units), enabling multi-dimensional perception of complex environments [5]. Visual sensors capture object features, infrared sensors overcome lighting limitations, and tactile sensors capture physical signals, supporting interaction with objects and humans. When interacting with humans, embodied AI can collect personal information including voice, expressions, and movements, then leverage powerful processor computing capabilities to complete real-time information recording and analysis, providing data support for behavioral decisions.

The essence of embodiment is the symbiotic unity of body and mind, emphasizing the Dynamic Coupling of physical carriers, cognitive algorithms, and environment. Embodied AI must rely on physical entities for existence, achieving environmental perception, behavioral execution,

and value output. Data collection requires physical terminals equipped with multiple sensors, and algorithm implementation requires execution modules coordinated with physical scenarios, all requiring physical carriers to transcend virtual symbolic limitations. The current trend toward anthropomorphic design of embodied AI agents is influenced partly by science fiction culture, and partly to meet their technical characteristics and interaction needs. Leveraging multi-modal information collected by multiple sensors, anthropomorphic design enables embodied AI agents to better align with human interaction habits, combining personalized analysis technology to achieve emotional interaction and enhance human trust and acceptance.

Second, Interactivity. Leveraging Natural Language Processing (NLP) and affective computing personalized analysis algorithms, multi-modal interaction-based emotional resonance can be achieved. Through physical carriers and analysis algorithms, embodied AI can interpret human emotions and behavioral intentions, achieving multi-modal bidirectional interaction. Through anthropomorphic design, highly humanoid appearances make embodied AI agents more readily accepted, promoting emotional upgrading of human-machine interaction. Users interacting with companion robots unconsciously view them as social entities and develop emotional dependence.

Embodied AI stimulates empathy and trust through interaction methods aligned with human social habits, but information sharing during interaction also poses new requirements for personal information protection. Leveraging the scene-embedding advantages of physical entities, embodied AI can achieve covert collection of privacy information through scenario-based guidance. For example, in consumption scenarios, embodied AI agents can extract consumption preferences through guided inquiries; in home companionship scenarios, they can gradually extract private information such as user daily routines. This scenario-adaptive extraction mode highlights the covert nature of embodied AI in privacy collection.

Third, Emergence. Emergence refers to new functions spontaneously formed during system operation through non-linear interaction, dynamic coordination, and continuous learning among modules, exceeding the capability range of individual modules, embodying the complex system characteristic of "the whole is greater than the sum of its parts." In the embodied AI field specifically, emergence mainly manifests as: through Nonlinear Processing modes, systems can continuously learn autonomously based on environmental dynamic data, and when facing complex scenario changes, can autonomously plan data processing flows, optimize decisions, and adjust behaviors without human intervention, demonstrating strong autonomy and adaptability.

Emergence-enhanced autonomous decision-making capability poses higher challenges to data circulation compliance control [6]. Particularly in data processing, multi-sensors simultaneously complete data Preprocessing while capturing environmental data. Thus, embodied AI breaks through the traditional linear progression mode of "collection-analysis-decision-use," presenting dynamic coupling and non-linear interaction characteristics, with blurred boundaries between data collection and processing, exhibiting Edge Computing-driven "collection-as-processing" characteristics. Additionally, in Nonlinear Processing mode, algorithms can conduct Online Learning iteration based on real-time data, autonomously planning data processing flows without human intervention, causing data processing stages to form network structures with multi-directional causal relationships, making traditional linear data control methods difficult to effectively intervene.

2.2. Privacy Risks Arising from Embodied AI Characteristics

1) Physical Privacy Risk

Embodied AI has broad applications in health monitoring, emotional companionship, and other fields. The Autonomous Navigation capability and humanoid form of embodied AI agents

enable them to "covertly intrude" into users' private spaces, conducting panoramic recording of private activities. Through PPG (Photoplethysmography) sensors and infrared thermal imaging sensing systems, embodied AI agents can collect physiological signals such as heart rate and body temperature in real-time, and extract Biometrics including fingerprints and facial features. These characteristics pose significant challenges for privacy protection. Data collection is highly automated and covert; users facing anthropomorphized AI agents tend to lower their guard due to emotional dependence, gradually losing awareness rights over body data. Data collected by embodied AI agents is mostly stored in cloud environments with insufficient security guarantees, and leakage would cause irreversible harm to users [7]. Some operators may use data for model training or unauthorized commercial development, infringing on awareness and control rights. The real-time nature of human-machine interaction makes traditional supervision difficult to intervene, exacerbating privacy protection difficulties [8].

2) Mental Privacy Risk

The deepening interactivity of embodied AI drives the human-machine relationship transformation from "tool dependence" to "emotional symbiosis," spawning mental privacy risks [9]. It points to the individual's inner emotional experiences, stemming from psychological trust induced by anthropomorphic design and the covert permeation of the spiritual world by algorithms.

The formation of psychological trust makes users more inclined to reveal personal privacy and even deep emotional needs to embodied AI agents. Embodied AI agents, through anthropomorphic appearance and empathetic dialogue, break down human-machine barriers, inducing emotional dependence in humans. Under trust, humans actively reveal mental privacy, while intelligent agents convert it into data processing, and individuals relinquish control rights through emotional resonance [10].

Additionally, users' "selective expression" under algorithmic guidance leads to loss of independent emotional judgment and self-awareness. "Disciplinary power" theory reveals that the interaction process is actually the algorithm's covert disciplining of the individual's spiritual world [11]. This "personalized" service places users under invisible "mental surveillance," with algorithms gradually shaping users' emotional expression and value judgments.

The physical and mental privacy risks individuals face reflect the real crisis of privacy self-determination rights being eroded. Large amounts of data are silently collected for commercial development, with individuals lacking awareness and intervention capabilities. The autonomous decision-making capability of embodied AI agents enables them to independently complete the entire process from information acquisition and analysis to decision execution, with unpredictable processes. Even if laws provide "automated decision-making rejection rights," these rights cannot form effective self-determination protection in practice due to excessively high technical thresholds in embodied AI scenarios. Privacy risks are more pronounced in digitally vulnerable groups such as the elderly and children.

2.3. Legal Risks for Personal Information Protection Arising from Embodied AI Characteristics

1) Impact on the "Informed Consent" Principle

The Informed Consent principle should be established on the premise that users have full knowledge of data collection purposes and processing methods. However, in the AI era, contactless, imperceptible automated data collection methods are widespread, impacting the preconditions for effective application of the informed consent principle.

The traditional "informed-consent" model is based on static, one-time authorization, while embodied AI agents affect personal information rights in non-transparent ways. Their data

collection and processing often present dynamic and scenario-based characteristics, causing blurred boundaries between data generation and processing, with users unable to predict in advance the scope, purpose, and subsequent circulation path of data collection. Even if data collection policies can be formally notified, users have difficulty maintaining rational cognition and making timely consent during dynamic interaction. Requiring embodied AI agents to frequently pop up windows during user interaction, separately prompting information collection matters and requiring user consent, would inevitably seriously affect user experience and be inefficient; while simplifying procedures would render rules formalistic, highlighting the limitations of traditional static consent models.

Users' cognitive biases and choice limitations render the "informed-consent" principle a mere formality, essentially transferring data risks to users. The effective operation of the traditional "informed-consent" principle relies on users' rational cognition and judgment of Data Processing Authorization, but in embodied AI scenarios, this premise is no longer tenable. Embodied AI agents obtain user trust through anthropomorphic design, easily manipulating users' consent decisions invisibly. Current embodied AI data collection methods and purposes lack transparency, making it difficult for users to clearly recognize the true purposes and potential risks of data processing. Users can only choose between "consent to all" or "refuse all," and so-called "consent" is essentially the result of forced compromise. This mechanism transfers risks entirely to users, forcing users to abandon substantive control over their personal information while enjoying embodied AI services, hollowing out the rights protection function carried by the traditional "informed-consent" principle. From the perspective of embodied AI industry development, strictly requiring separate consent for each cross-entity data transfer would prevent embodied AI enterprises from achieving cross-scenario data reuse and cross-entity sharing, hindering model training and technology iteration; breaching rules would face legal liability for illegal processing of personal information.

2) Failure of the Purpose Limitation Principle

The Purpose Limitation principle takes information collection as the regulatory starting point, delineates pre-review standards for personal information processing [12], clarifies the relevance between collection purposes and processing behaviors, connects with the notification and consent system to provide a foundation for user full knowledge, and serves as important legal basis for the personal information collection stage.

However, in embodied AI scenarios, this principle faces application difficulties of being breached. It manifests in that embodied AI's data collection purposes are overly broad, making accurate delineation difficult in practice. For example, in home care scenarios, embodied AI agents assume multiple functions, with data processing purposes presenting multiplicity. Precisely delineating all purposes in advance faces practical difficulties. The development goal of embodied AI agents is to achieve humanized interaction, which inevitably relies on continuous, massive data collection. Additionally, the emergence of embodied AI makes it difficult for processors to clearly delineate all processing purposes at the initial collection stage, causing the "clear, reasonable purposes" required by the purpose limitation principle difficult to implement.

Furthermore, the application of the purpose limitation principle in embodied AI scenarios conflicts with the industry goals of embodied AI and big data industry development, further exacerbating this principle's failure. The purpose limitation principle protects personal rights by limiting processing purposes, but this contradicts big data development requirements of "processing data without preset purposes to extract maximum value." Embodied AI machine learning relies on massive data training, autonomously learning through detecting data correlations. Data for training cannot be predicted in advance, and strictly following this principle would hinder technology iteration.

3) Difficulty in Implementing the Data Minimization Principle

Embodied AI presents the characteristic of "more data leads to better model optimization," making "minimum necessary scope" difficult to delineate. The Data Minimization principle and purpose limitation principle require data processing to be limited within reasonable purpose scope, achieving minimum scope and minimum impact.

The autonomous action characteristic of embodied AI agents necessitates continuous data collection to judge intentions, optimize decisions, and execute actions, thus negating the application premise of the data minimization principle. Combined with the foregoing, data processing purposes themselves are broad, and when purposes are unclear, the delineation of "minimum necessary" loses its basis. The traditional application of the data minimization principle is based on the assumption that system data is in a static state, but current embodied AI machine learning systems require continuous iterative optimization. A single user's "minimum necessary" information may be key data for other users' model training; minimization for a single entity would affect overall system performance, causing "minimum necessary" delineation to fall into difficulty.

The scenario diversity and technical development needs of embodied AI exacerbate the difficulty of implementing the data minimization principle. Data needs differ across scenarios. Rescue robots need to broadly collect terrain, environment, and other dynamic data; medical scenarios rely on precise physiological and behavioral signals; home care needs to obtain substantial personal preference and emotional information. Mechanically applying this principle would cause functions to fail to operate normally and hinder technology development.

The legal risks of embodied AI at the data collection end are mainly concentrated in excessive personal information collection. However, large-scale collection and utilization of personal information is an inevitable path for AI industry development. The conflict between the Personal Information Protection Law's legislative principles and embodied AI industry development reflects the dilemma of excessive personal information collection, but massive data collection and infringement of rights are not necessarily positively correlated. The positive and negative externalities of personal information utilization highly depend on specific scenarios. Whether personal information processing brings privacy risks is not determined by whether it constitutes personal information, but by "how it is used" in specific scenarios and whether it meets users' reasonable expectations. Therefore, privacy risk should be used as the standard for measuring processor compliance, shifting focus from the collection stage to the utilization stage, namely assessing the privacy risks triggered by information use.

3. Specific Pathways for Promoting Embodied AI Data Circulation

3.1. Establishing the Regulatory Philosophy of Distinguishing Personal Information from Personal Data

1) Legal Application Dilemma of Conflating Personal Information and Personal Data

Establishing the regulatory philosophy distinguishing personal information from personal data is the prerequisite for resolving data circulation dilemmas. Its necessity stems from the current confusion of the two concepts in legal practice and the strict restrictions of the Personal Information Protection Law on personal information circulation.

From legal practice perspective, information and data have long failed to be effectively distinguished, with conflation being prevalent. China's Civil Code also fails to make clear delineation between the two in its expression. Personal information and personal data belong to different categories: personal information is the information ontology, while personal data is the medium or carrier bearing information. Personal information and personal data have essential differences in legal attributes. Personal information carries personality rights,

concerning personal safety and human dignity, strictly protected by the Personal Information Protection Law and Civil Code; personal data reflects property rights, serving as electronic carriers that can extract economic value through processing and circulation, with regulatory focus on standardizing circulation order. The distinction between the two echoes academic consensus: personal information as "ontology" emphasizes personality interests; personal data as "medium" emphasizes property interests. Although the two are highly interdependent, the values and regulatory orientations protected by law are fundamentally different.

According to relevant provisions of the Personal Information Protection Law, personal information, as information closely related to natural persons' personal rights, lacks legal basis for centralized, large-scale circulation. Its collection, processing, and transfer must all comply with strict legal boundaries to protect natural persons' personality rights from infringement. If personal information and personal data continue to be conflated without clear legal delineation and distinction, personal data at any stage of circulation will be incorporated into the regulatory scope of personal information, thereby subject to strict restrictions of the Personal Information Protection Law. This dilemma is more pronounced in embodied AI scenarios: embodied AI development highly depends on massive personal data collection, processing, and circulation. If personal data cannot achieve legal circulation due to conceptual confusion, it will not only hinder embodied AI large model training and technology iteration, but also prevent effective exploitation of data value, unfavorable to compliant and efficient circulation of embodied AI datasets, contradicting embodied AI development philosophy. Therefore, establishing the regulatory philosophy distinguishing personal information from personal data, clarifying their legal boundaries and regulatory rules, is both a practical need for clarifying current legal application confusion and an inevitable choice for resolving embodied AI data circulation dilemmas and achieving win-win rights protection and industry development.

2) Practical Significance of Distinguishing Personal Information from Personal Data

The regulatory philosophy distinguishing personal information from personal data, promoting personal information transformation into circulatable data through anonymization, is key to resolving data circulation dilemmas.

First, "anonymized" personal data escapes strict regulation of the Personal Information Protection Law, promoting personal data to fully release property benefits, reducing compliance costs for data circulation and utilization in embodied AI scenarios, without needing to fulfill pre-procedures for personal information processing. After personal information is anonymized and transformed into data, identity identification attributes are stripped, no longer subject to Personal Information Protection Law regulation, adapting to embodied AI's "more data leads to better optimization" iteration pattern. Enterprises can use it for Deep Learning model training, algorithm optimization, and scenario adaptation, effectively resolving embodied AI data supply insufficiency and circulation obstruction dilemmas.

Second, the philosophy of distinguishing personal information from personal data clarifies rights boundaries for data circulation, avoiding ownership disputes in personal data circulation under embodied AI scenarios, providing foundation for enterprises to utilize data and obtain legitimate benefits [13]. Conceptual confusion and ownership ambiguity easily trigger disputes. The boundary between the two is reflected at processing nodes: the initial stage obtains personal information, which must comply with the Personal Information Protection Law. After de-identification, integration, and other technical processing, identity identification attributes are stripped, and personal information transforms into personal data. Personal data escapes Personal Information Protection Law restrictions, obtaining legal circulation basis. Post-transformation, enterprises' limited control rights over data can be established, incentivizing their investment in anonymization technology R&D. Additionally, this data control right does not protect enterprise control over information content; other enterprises can equally produce

and control data with identical content, thereby promoting entire industry progress through fair competition in data development and utilization.

3) Tiers and Standards of Enterprise Anonymization in Embodied AI Scenarios

In the process of personal information transformation to personal data, "anonymization" is the primary technical means for stripping personality attributes from personal information, thereby granting legality to personal data circulation. However, "anonymization" in embodied AI scenarios faces significant "re-identification" risks, necessitating exploration of reasonable standards suitable for this application scenario.

Information "identifiability" mainly depends on information's own identification attributes and inter-information linkability, through direct identifier individual identification, quasi-identifier cross-correlation, or machine learning-based Linkage Attack to identify specific individuals. In embodied AI scenarios, the risk of anonymized data being "Re-identified" stems from multiple factors including technology and scenario characteristics [14]. From a technical perspective, information collected by embodied AI covers multi-dimensional content including biometrics, behavioral trajectories, and emotional preferences. Even if one dimension is anonymized, attackers can still achieve identity restoration through cross-matching auxiliary data; from the perspective of embodied AI scenarios themselves, their humanized interaction and autonomous decision-making characteristics require data to be continuously updated and multi-dimensionally correlated, which creates tension with the "de-identification" pursued by anonymization. Excessive pursuit of data utility reduces anonymization degree, increasing "re-identification" risk; excessive strengthening of anonymization weakens data value, hindering embodied AI model optimization. Additionally, in judicial practice, the Zhu Ye v. Baidu case also confirms this dilemma: the court recognized that anonymized data escapes the personal information category, but ignored the reality that Cookie anonymous data can be combined with social content, interaction trajectories, and other auxiliary information to achieve re-identification, reflecting the disconnect between anonymization risks and legal regulation.

Against the background of increasingly prominent "re-identification" risks, "what tier of enterprise anonymization constitutes legal validity" and "how to delineate anonymization standards in embodied AI scenarios" become issues urgently requiring clarification in current practice. Combining China's legislative provisions and judicial practice, enterprise anonymization processing of personal information is essentially "de-identification" labor processing. Enterprise anonymization should present gradation, divided into three tiers. First, Basic Anonymization: targeting general personal information (such as behavioral preferences), adopting suppression, generalization, and other technologies, applicable to ordinary life service scenarios. Second, Intermediate Anonymization: targeting more sensitive information, requiring "de-identification + irreversible" combination technology, deleting direct identifiers and fuzzy-processing quasi-identifiers such as age, occupation, and whereabouts, ensuring identity cannot be restored through single datasets while preventing De-anonymization operations through irreversible technology, applicable to embodied AI health monitoring, light care, and other scenarios. Third, High-level Anonymization tier, targeting critical sensitive personal information in high privacy risk scenarios, such as fingerprints and private body part biometric data, adopting Differential Privacy (DP), GAN (Generative Adversarial Network) synthetic data, Federated Learning (FL), and other Privacy-Enhancing Technologies (PETs), minimizing data association with specific natural persons while preserving statistical value, even generating artificial data statistically equivalent to original data, ensuring re-identification is impossible even combining multi-dimensional auxiliary data, applicable to embodied AI medical rehabilitation, precision interaction, and other high-risk privacy leakage scenarios.

Regarding anonymization standards, China's legislation mainly uses "unable to identify specific natural persons and cannot be restored" established by the Personal Information Protection Law as judgment basis, also reflected in Article 42 of the Cybersecurity Law's exception clause

for personal information informed consent rules of "processed to be unable to identify specific individuals and cannot be restored"; Article 27 of the Data Security Law further clarifies data processors' legal obligation to adopt technical and other measures to ensure data security; and Article 1038 of the Civil Code also makes similar arrangements [15]. However, in embodied AI scenarios, these legal norms are facing challenges due to their mechanical and lagging nature, difficult to adapt to scenario requirements. Re-identification risks cannot be eliminated; anonymization is essentially a matter of risk "degree" rather than "presence." Therefore, absolute standards neither conform to embodied AI technical development logic nor may cause enterprises to reduce data utility through excessive compliance pursuit, or evade anonymization obligations due to excessively high standards. Combining the particularity of embodied AI scenarios, anonymization standards should transform from "absolute security" to "context-adaptive, risk-tolerant" dynamics [16]. This can be delineated from three dimensions: first, Risk-oriented standards, using "acceptable low risk" as judgment criterion, combining embodied AI application scenarios, comprehensively assessing data sensitivity, circulation scope, re-identification technical difficulty and other factors to determine reasonable risk tolerance thresholds. As long as anonymized data has re-identification risk within threshold range, it is recognized as legal anonymization, similar to the relative standard established by HIPAA; second, Context-adaptive standards, adhering to "scenario differentiation" principles, adjusting standards according to specific embodied AI application contexts, avoiding "one-size-fits-all" regulatory models. For example, medical scenario anonymization standards should be higher than ordinary life service scenarios, and cross-border data should adopt "higher standard" principles to adapt to different jurisdiction standards; third, Process-oriented standards, no longer solely based on "unidentifiable, unrestorable" results as the only basis, extending regulatory focus to the entire anonymization processing process, emphasizing compliance, transparency, and traceability of anonymization operations. Requiring enterprises to fully integrate data protection measures into collection, processing, storage, circulation and other stages of anonymization, develop clear internal operating procedures, establish audit tracking mechanisms, disclose anonymization technical means and risk warnings, achieving both anonymization effectiveness and process compliance guarantees.

3.2. Transforming User Informed Consent into Enterprise Proactive Prevention Obligations

1) Practical Necessity of Establishing Enterprise Proactive Prevention Obligations in Embodied AI Scenarios

In embodied AI scenarios, the traditional "informed-consent" mechanism for personal information protection is facing systemic failure, and its status as the legal basis for data collection has been significantly weakened [17]. This also means that the legal basis for embodied AI data collection should no longer primarily rely on user "informed-consent," but should shift more toward enterprise-constructed privacy risk minimization mechanisms. Therefore, placing the proactive prevention obligations of enterprises as data processing core entities at the center of institutional construction can resolve embodied AI personal information protection dilemmas and adapt to current technology development.

Some scholars have pointed out that in embodied AI application scenarios, informed consent should be further strengthened: manufacturers or providers should fully prompt the special capabilities possessed by humanoid robots and clearly inform users where their sensitive information will be used and how. However, from the current application of traditional "informed-consent" mechanisms, embodied AI's humanized interaction, autonomous decision-making, and massive data collection characteristics break the three major premises for effective operation of "informed-consent" mechanisms: user readability, user decidability, and platform provability. Embodied AI's algorithmic black box causes data processing logic to be complex

and difficult to understand; even if relevant information is disclosed to users, most users have difficulty truly understanding the scope, purpose, and potential risks of data collection. Embodied AI's technology and market monopoly advantages leave users lacking real freedom of choice when obtaining services, either accepting bundled consent clauses or abandoning service use, with so-called "consent" becoming passive compromise formality; coupled with embodied AI potentially prompting users to make consent decisions in non-rational states such as emotions and pressure through anthropomorphic induction, emotional manipulation and other means, further amplifying consent mechanism defects, causing user consent difficult to reflect true intentions. It is noteworthy that purchasing embodied AI products is not the same concept as consenting to personal information processing.

Establishing enterprise proactive prevention obligations has important significance. First, it aligns with practical needs for data protection, achieving transformation from "formal compliance" to "substantive protection [18]." Traditional informed consent mechanism alienation causes enterprises to often invest energy in "compliance performance," neglecting construction of underlying security mechanisms such as data encryption, de-identification, and access control. In embodied AI scenarios, personal information leakage risks mainly stem from vulnerabilities in enterprise data processing flows. Enterprises constructing reasonable proactive prevention mechanisms can control privacy risks from the source, achieving substantive personal information protection, which also aligns with "risk prevention and control" content stipulated in the Cybersecurity Law and Data Security Law. Second, it balances data utilization and privacy protection, adapting to embodied AI technology development logic. Embodied AI development highly depends on massive data collection and utilization. If adhering to traditional "informed-consent" mechanisms, it will not only increase enterprise compliance costs but also constrain data factor circulation and technology innovation; establishing enterprise proactive prevention obligations exempts enterprises from excessive dependence on formalized consent, also providing clear legal basis for enterprise data collection and utilization, enabling enterprises to fully exploit data value under controllable risk premises, supporting embodied AI large model training and scenario adaptation optimization. Third, it highlights enterprise entity responsibility, resolving rights-responsibility imbalance dilemmas. Enterprises in embodied AI, as leaders in data collection, processing, utilization, and circulation, possess significant technology and data advantages. Compared with dispersed, weak users, they are better equipped with capabilities and conditions to construct risk prevention mechanisms. Establishing their proactive prevention obligations both practices the "consistency of rights and responsibilities" principle and effectively resolves the unreasonable situation of "users bearing responsibility, enterprises exempt" in current embodied AI scenario personal information protection, strengthening enterprise responsibility awareness and compliance awareness.

2) Construction Pathway for Front-end Enterprise Proactive Prevention Mechanisms

The dilution of "informed-consent" rules in embodied AI scenarios does not mean completely abandoning this mechanism, but rather breaking the traditional logic of "consent equals legality." Combined with privacy risk levels of data leakage, a differentiated consent rule system should be constructed, clarifying that user informed consent can be exempted in general low-risk scenarios, while "informed-consent" rules must still be followed in special high-risk scenarios, achieving balance between privacy protection and data utilization, promoting enterprises to construct and implement proactive prevention mechanisms.

In low-risk scenarios, enterprises need not obtain explicit user consent to legally collect personal information, but the scope is limited to non-sensitive information, and users should be granted "opt-out" modes. On the basis of reasonably informing users of the general scope and purpose of data collection, enterprises can collect relevant data by default while providing users with convenient, understandable rejection channels, allowing users to withdraw

authorization and delete data. In general application scenarios of embodied AI, such low-risk data collection is quite common, such as ordinary home companionship embodied AI collecting users' daily routines, interaction preferences, and other non-unique information. Such information, if individually leaked, would not cause user identity identification. Even if users are unwilling to have such information leaked, remediation can be achieved through back-end remedial measures: users can provide ex-post restraint for their personality rights protection through consent withdrawal rights, deletion rights, and other powers; or reasonably realize their data property interests.

In high-risk scenarios, "informed-consent" rules must still play the main role. Enterprises need to fulfill stricter notification obligations to ensure user consent is an explicit expression based on true intentions, which is also an indispensable bottom-line requirement in enterprise proactive prevention mechanisms. In embodied AI scenarios, highly sensitive privacy mainly refers to critical sensitive personal information, especially unique information. Once such information is leaked or re-identified, it will cause irreversible damage to users' human dignity and personal safety. For such high-risk scenarios, enterprises not only need to obtain users' explicit separate consent but also need to draw on participatory consent models in medical decision-making. Through simple, understandable methods, they should exhaustively disclose to users or their guardians the specific scope, purpose, processing method, and potential risks of data collection, repeatedly confirming user intentions, ensuring users fully understand and voluntarily make consent decisions. Additionally, dynamic consent modes should be established, allowing users and their guardians to update and modify consent content at any time according to their own intentions, ensuring users always enjoy control over collection and processing of sensitive personal information.

3) Establishment and Implementation Pathway of Back-end Enterprise Data Review Obligations

In embodied AI scenarios, enterprise proactive prevention obligations for personal information leakage are not limited to front-end data collection and processing stages but must extend to data utilization end, clarifying back-end review obligations of transferee enterprises in data transactions, forming a "front-end prevention—back-end review" full-process enterprise proactive prevention system, filling risk prevention and control gaps in data transaction stages, fundamentally preventing personal information leakage, implementing enterprise proactive prevention responsibilities, and balancing data circulation with privacy protection.

Back-end transferee enterprises must fulfill strict review obligations when acquiring data, verifying the legality and compliance of acquired data, preventing damage to personal information rights of user source entities due to acquiring illegally collected data or data not meeting anonymization standards [19], while avoiding enterprises themselves falling into compliance dilemmas due to using illegal data. The establishment of this obligation stems from the particularity of embodied AI data transactions: embodied AI data involves multi-dimensional, multi-level personal information and carries anonymization re-identification risks. Data providers may evade front-end prevention obligations, fail to complete anonymization processing according to standards, or disguisedly trade illegally collected personal information. If transferee enterprises do not fulfill review obligations, they will become "secondary carriers" of personal information leakage, exacerbating personal information protection dilemmas in embodied AI scenarios. Additionally, transferee enterprises, as terminal entities of data utilization, their review of data legality is also an important means to reverse-force front-end data providers to standardly fulfill proactive prevention obligations and implement anonymization standards, forming full-chain responsibility constraint mechanisms.

4. Institutional Safeguards for Promoting Embodied AI Data Circulation

4.1. Regulatory Innovation: Construction of Multi-party Collaborative Dynamic Regulatory Mechanisms

The particularity of embodied AI data circulation makes traditional regulatory models difficult to effectively adapt. Current regulation faces problems: first, regulatory lag. Rapid iteration of embodied AI technology makes traditional human regulation difficult to accurately identify hidden risks in data circulation. Second, insufficient regulatory targeting. Uniform, one-size-fits-all regulatory models cannot balance compliance control and industrial innovation. Third, insufficient regulatory coordination. Coordination among regulatory agencies, enterprises, the public, and interdisciplinary forces is inadequate, failing to form comprehensive, multi-level regulatory ecosystems. Therefore, regulatory mechanism pathways for embodied AI scenarios need to be explored to achieve balance among data security protection, enterprise compliant development, and industrial innovation upgrading.

First, construct a dynamic regulatory system combining penetrating regulation with Regulatory Sandbox. For enterprise compliance and regulatory gaps, establish penetrating regulatory mechanisms based on blockchain and Trusted Execution Environment (TEE), breaking information barriers between regulatory agencies and regulated enterprises. Regulatory agencies can establish real-time data sharing mechanisms with enterprises and data intermediary institutions, directly obtaining full-process data circulation status, transforming previous passive reception of enterprise submission reports into active penetrating supervision of each stage of data collection, processing, and utilization, analyzing risk points and conflict of interest points, fundamentally cutting off illegal data circulation channels [20], ensuring regulatory comprehensiveness and timeliness, and strengthening supervision of data intermediary service compliance, ensuring their neutrality and professionalism. Additionally, implement Regulatory Sandbox mode, avoiding excessive regulation that suppresses industrial development. Regulatory agencies, regulated enterprises, data intermediary institutions, and relevant experts jointly determine protection measures and technical limitation rules for data circulation, adjusting regulatory plans timely according to risk changes during testing. If enterprises object to adjustments, they need to provide legitimate reasons. This mode can formulate differentiated regulatory policies for different enterprises and scenarios, achieving risk prevention and control while enhancing interaction between regulatory agencies and enterprises, alleviating regulatory time lag, balancing data security and industrial innovation, particularly suitable for development needs of small and medium embodied AI enterprises, avoiding excessive compliance review pressure.

Second, adhere to moderate regulation principles, balancing regulatory intensity and industrial innovation. Regulation should seek balance between intervening in risks and safeguarding innovation. Regulatory focus should center on risks in embodied AI data circulation, strengthening supervision of manipulation intent based on personal profiling and hidden manipulation behaviors inconsistent with user personal goals, requiring enterprises to proactively fulfill prevention mechanisms. Meanwhile, regulation should consider development differences among enterprises of different scales, implementing differentiated compliance requirements for small and medium enterprises, which can rely on data intermediary services to reduce compliance and regulatory adaptation costs, avoiding one-size-fits-all regulatory models that suppress industrial vitality, while strengthening focused supervision of large enterprises, urging them to fulfill front-end prevention, back-end review, and other obligations, playing industry leadership roles.

Third, construct multi-party collaborative regulatory ecosystems, strengthening interdisciplinary cooperation and public participation. The complexity of embodied AI requires integrating knowledge from law, ethics, engineering, and other disciplines, forming

professional regulatory teams to identify cross-risks at technical, legal, and ethical levels, providing support for regulatory policy formulation. Additionally, improve public participation mechanisms through public consultations, hearings, and other forms, absorbing public opinions and concerns on embodied AI data circulation regulation, enhancing policy-making transparency and social acceptability, strengthening public supervision of enterprise violations and regulatory agency performance. Furthermore, promote collaborative coordination among regulatory agencies, enterprises, industry associations, and data intermediary institutions. Industry associations formulate self-regulatory standards, enterprises fulfill proactive compliance obligations and cooperate with regulation, data intermediary institutions provide technical support and compliance assistance, and regulatory agencies are responsible for supervisory enforcement and violation sanctions, forming "government regulation, industry self-discipline, enterprise self-governance, public supervision" multi-party collaborative ecosystems.

4.2. Intermediary Collaboration: Introduction of Data Intermediary Institutions

Embodied AI datasets are large in scale, diverse in type, and rich in sensitive information, requiring strict adherence to hierarchical classification and anonymization rules. However, anonymization technology R&D requires high investment that small and medium enterprises cannot afford, and large enterprises also need to expend substantial resources addressing standard updates. Additionally, full-chain compliance requirements for data circulation are diverse, requiring connection with multiple laws and regulations, while most enterprises focus on technology R&D and lack professional compliance teams, easily causing compliance gaps and facing tort compensation and administrative penalty risks. Coupled with information asymmetry between data subjects and users, enterprises need to balance rights protection and value mining, further exacerbating compliance burdens. Relying solely on enterprises themselves cannot resolve this dilemma, unable to achieve the purpose of promoting data circulation and utilization.

Data Intermediary institutions provide a feasible solution pathway. Referring to EU Data Governance Act (DGA) definitions, data intermediary institutions are third-party institutions independent of data collection, utilization, and other entities. They do not participate in substantive data utilization and transactions but conduct services through data trust relationships, providing professional compliance supporting services, resolving trust crises and compliance difficulties in data circulation, connecting data subjects, enterprises, and regulatory departments, building compliant circulation bridges. Their main functions include: first, technical support, providing customized anonymization processing services; second, compliance assistance, conducting full-chain compliance review and process optimization, assisting enterprises in implementing compliance requirements; third, rights protection and supervision, serving as trustees of data subject rights, assisting in exercising statutory rights while supervising enterprise fulfillment of proactive prevention obligations.

Data intermediary institutions can precisely respond to multiple dilemmas in embodied AI scenarios: their professional technical capabilities can reduce enterprise anonymization processing costs, assisting enterprises in meeting compliance standards, alleviating technology investment pressure; [21] their professional compliance services can compensate for enterprise compliance capability deficiencies, assisting enterprises in connecting full-chain compliance requirements, reducing compliance gaps and risks; their neutrality can resolve information asymmetry issues, building trust bridges between data subjects and enterprises, reducing enterprise burden of balancing dual demands, achieving positive circulation of privacy protection and data circulation, promoting construction of enterprise proactive prevention full-chain regulatory systems.

4.3. Liability Allocation: Typological Considerations for Privacy Tort Liability

The prerequisite for privacy tort liability determination is clarifying the legal status of embodied AI agents. Although embodied AI agents possess autonomous decision-making capabilities, they essentially remain technical tools lacking free will, unable to be incorporated into the legal subject category within existing legal system "subject-object" binary structure [22]. Excessively granting them legal subject status would instead blur enterprise responsibility. Privacy tort liability allocation should still focus on human subjects, unfolding along the entire embodied AI industrial chain, dynamically defining party responsibilities based on technical logic and data circulation stages. Among these, enterprises, as leaders in data collection, processing, utilization, and circulation, should bear primary responsibility for privacy torts, transforming "proactive prevention" from concept to specific responsibility constraints.

R&D and production enterprises bear primary responsibility. R&D enterprises need to bear primary responsibility for algorithm security and compliance: fulfilling algorithmic transparency obligations in initial stages, disclosing data processing logic; bearing Explainable AI (XAI) obligations after large-scale application, avoiding privacy torts caused by algorithm defects. Production enterprises need to implement anonymization technical standards and security requirements, ensuring E2EE (End-to-End Encryption), TPM (Trusted Platform Module), and other hardware compliance, bearing product liability for leakages caused by hardware defects. Post-sales should continuously push security patches, providing compliance manuals at sales stages. Users, as weaker parties, only bear responsibility for intentional cracking or gross negligence; torts caused by enterprise technical defects should not be attributed to user fault.

Embodied AI data collection enterprises bear intermediate-stage responsibility, connecting front-end production and back-end use. As data controllers, they need to fulfill moderate use obligations and security guarantee obligations, with corresponding civil liability mainly including tort liability and breach liability. Data collection enterprises bear scenario-based notification obligations, needing to clearly indicate data transmission scope, third-party access situations, and fulfill moderate use obligations. According to principles of legality, legitimacy, and necessity in Article 1035 of the Civil Code, enterprises shall not excessively process personal data, shall not adopt default settings such as "collection starts upon boot" to force users to provide privacy information. If inadequate notification causes rights damage, fault liability must be borne, and "user did not refuse" cannot be used to claim exemption [23]. Additionally, enterprises need to continuously fulfill data security maintenance obligations, referring to physical space operator safety guarantee obligations in Article 1198 of the Civil Code and relevant provisions of Article 51 of the Personal Information Protection Law, supervising full-process data circulation, preventing data leakage, De-anonymization, and other risks.

Utilization-end enterprises bear back-end responsibility for data circulation, and should strengthen proactive review and risk prevention obligations. Utilization-end enterprises need to adopt differentiated protection measures according to hierarchical classification results, following anonymization standards, must not exceed circulation limits or de-anonymize, otherwise bear legal liability. Additionally, utilization-end enterprises need to establish data compliance ledgers, recording full-process of reception, storage, use, and destruction, conducting regular privacy risk self-inspections. If inadequate self-inspection causes leakage or abuse, fault liability must be borne. If data is found to have legality defects or anonymization failure, utilization must be immediately stopped and reported to regulatory departments. Failure to fulfill this obligation causing damage expansion requires bearing corresponding liability.

5. Conclusion

Against the backdrop of deep development of the digital economy and the 2026 "Year of Data Factor Value Release," data has become a key production factor driving industrial upgrading and reshaping economic forms. Embodied AI, as a new form of deep integration of AI with robotics and edge computing, its large-scale application is promoting datasets from laboratories to market transactions, with increasingly prominent standardization and tradability advantages of data circulation. However, the technical characteristics of embodied AI-multi-sensor fusion, non-linear processing, and personalized analysis-cause multiple tort risks including physical privacy, mental privacy, and self-determination privacy to lurk in data circulation processes. Traditional personal information protection principles of informed-consent, purpose limitation, and data minimization fall into application difficulties. Moreover, the traditional "informed-consent" model proves increasingly inadequate for dynamic data-trading contexts, and the consent-based framework requires structural redesign to accommodate emerging intelligent systems. Coupled with academic research focusing more on data collection stages while relatively neglecting circulation regulation, embodied AI data circulation has fallen into an imbalanced state of insufficient value release and weak privacy protection, becoming a bottleneck constraining high-quality industrial development and marketization of data factor allocation.

Addressing these dilemmas, this paper breaks through static thinking and informed-consent protection models of traditional regulation, based on scenario particularity and technical patterns of embodied AI, with balancing personal information rights protection and data value release as the core, constructing a comprehensive, differentiated embodied AI data circulation governance framework. This paper establishes the regulatory philosophy distinguishing personal information from personal data, clarifying their legal boundaries and anonymization transformation pathways. Additionally, it constructs a responsibility model of transforming user informed consent into enterprise proactive prevention of personal information leakage, filling research gaps in this field. It strengthens enterprise entity responsibility, establishing "front-end prevention-back-end review" full-process responsibility systems; integrates technical empowerment and institutional safeguards, forming "technical support-institutional constraints-multi-party collaboration" comprehensive governance pathways from three dimensions of regulatory innovation, data intermediaries, and tort liability, providing guidance for compliant dataset transactions and enterprise compliant operations.

Embodied AI technology iterates rapidly, and data circulation forms and risks will continue to evolve. How to achieve coordination among technological innovation, data value, and privacy protection remains an important proposition in the digital intelligence era. Only through interdisciplinary research collaboration, improving institutional design and strengthening technical support, can institutional boundaries for embodied AI data circulation be delineated, promoting data factor value release and assisting healthy development of the embodied AI industry.

References

[1] Hua, J. (2024). The impact of humanoid robots on different types of privacy and legal responses. *Journal of Tongji University (Social Science Section)*, (6), 119–125.

[2] Zhao, J. W. (2025). The specific connotation and judgment criteria of the principle of minimum necessity in AI agent scenarios. *Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition)*, (4), 62–71.

[3] Li, Z., & Chen, Y. Y. (2025). Cooperative governance of data privacy risks in embodied AI agents. *Journal of Shanghai University of Finance and Economics*, (5), 140–147.

[4] Lisondra, M., Benhabib, B., & Nejat, G. (2026). Embodied AI with foundation models for mobile service robots: A systematic review. *Robotics*, 15(3), 55. https://doi.org/10.3390/robotics15030055

[5] Zhang, W., Kong, X., Braunl, T., et al. (2024). SafeEmbodAI: A safety framework for mobile robots in embodied AI systems. arXiv Preprint arXiv:2409.01630. https://doi.org/10.48550/arXiv.2409.01630

[6] Perlo, J. (2025). Embodied AI: Emerging risks and opportunities for policy action. arXiv Preprint arXiv:2509.00117. https://doi.org/10.48550/arXiv.2509.00117

[7] Robey, A., Ravichandran, Z., Kumar, V., et al. (2025). Jailbreaking LLM-controlled robots. In *2025 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 11948–11956). IEEE.

[8] Roberts, H., Cowls, J., Morley, J., et al. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. In A. book editor (Eds.), *Ethics, governance, and policies in artificial intelligence* (pp. 47–79). Springer International Publishing.

[9] Laestadius, L., Bishop, A., Gonzalez, M., et al. (2024). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 26(10), 5923–5941.

[10] Mittelstadt, B. (2022). From individual to group privacy in big data analytics. *Philosophy & Technology*, 35(1), 1–28.

[11] Zuboff, S. (2024). The age of surveillance capitalism: The fight for a human future at the new frontier of power. *Journal of Information Ethics*, 33(1), 84–85.

[12] Mühlhoff, R., & Ruschemeier, H. (2024). Predictive analytics and the GDPR: Collective dimensions of data protection. *Law, Innovation and Technology*, 16(1), 107–133.

[13] Finck, M., & Pallas, F. (2020). They who must not be identified—Distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, 10(1), 11–36.

[14] Sweeney, L. (2002). k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 557–570.

[15] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., et al. (2023). Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: Scoping review. *Journal of Medical Internet Research*, 25, e41588.

[16] Ponomareva, N., Hazimeh, H., Kurakin, A., et al. (2023). How to DP-fy ML: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77, 1113–1201.

[17] Solove, D. J. (2024). Murky consent: An approach to the fictions of consent in privacy law. *Boston University Law Review*, 104, 593.

[18] Bygrave, L. A. (2017). Data protection by design and by default: Deciphering the EU's legislative requirements. *Oslo Law Review*, 4(2), 105–120.

[19] Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review*, 34(3), 436–449.

[20] Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops* (pp. 180–184). IEEE.

[21] Von Ditfurth, L., & Lienemann, G. (2022). The Data Governance Act: Promoting or Restricting Data Intermediaries? *Competition and Regulation in Network Industries*, 23(4), 270–295.

[22] Danaher, J. (2016). Robots, law and the retribution gap. *Ethics and Information Technology*, 18(4), 299–309.

[23] Ma, G. X. (2024). Perfecting the personal information protection system in data trading: Centering on the "informed-consent" rule. *Hebei Academic Journal*, 44(2), 168–178.