# The Challenge of Generative AI to Personal Information Protection and Its Risk Regulation

**HUANG Pei***

**Abstract:** The technical characteristics of Generative AI pose a challenge to the personal information protection system established by China's Personal Information Protection Law, including: although Generative AI has solved the problem of pre-training language models with large-scale unlabeled data, the adopted technical route also makes Large Language Models (LLMs) a complete technical black box, making it difficult for developers to comply with the rules of informed consent for personal information processing; whether based on the principle of purpose limitation or the theory of scenarios, the technical characteristics of Generative AI make it difficult to meet the statutory requirement of processing personal information within a "reasonable scope" in accordance with the law; the technical characteristics of Generative AI make it possible for LLMs to infringe on the sensitive personal information rights and individual privacy rights of information subjects at both the input and output ends. We should base our basic risk regulatory philosophy of "inclusive prudence" on this and take measures such as adjusting the application of informed consent rules in the field of Generative AI, reshaping the rules for processing publicly available personal information in the field of Generative AI, and establishing administrative regulatory measures for protecting the personality rights of personal information in the field of Generative AI, to achieve a balance between innovative technology development and personal information protection.

**Keywords:** Generative AI; ChatGPT; Personal Information; Risk Regulation

## I. INTRODUCTION

Generative AI refers to a subcategory of artificial intelligence technology, defined as "artificial intelligence systems capable of autonomously generating new text, images,

* HUANG Pei (黄锫), Professor, Law School of Tongji University.

audio, and other content."[1] *China's Interim Measures for the Administration of Generative Artificial Intelligence Services*, which took effect on August 15, 2023, define generative AI technology in Article 22 as "models and related technologies with the capability to generate content such as text, images, audio, and video." Generative AI differs fundamentally from traditional decision-making AI. The latter analyzes large-scale datasets, studies conditional probabilities within these datasets, identifies relatively stable patterns, and makes predictions about the future to assist decision-making processes. In contrast, generative AI uses algorithmic models to create entirely new content based on patterns identified through big data analysis. Simply put, while the primary function of traditional decision-making AI is "making predictions," the defining capability of emerging generative AI is "producing content."[2]

These functional differences lead to distinct application scenarios for each type of artificial intelligence. Decision-making AI is typically deployed in contexts requiring predictions of user needs, environmental conditions, and risk probabilities—such as personalized recommendation services on e-commerce platforms, environmental monitoring systems in autonomous vehicles, and investment or default risk assessments in the financial sector. Conversely, generative AI excels in domains requiring efficient, rapid, and autonomous creation of specific digital content, with prominent examples including ChatGPT for text and code generation, Midjourney for image creation, Sora for video production, and MuseNet for music composition.

The emergence of generative AI marks a watershed moment in technological development and represents a revolutionary stage in the evolution of artificial intelligence. It promises to become not only a powerful engine driving rapid digital economic growth but also a transformative force fundamentally reshaping numerous aspects of human life. However, alongside its rapid development, several implicit risks have become increasingly apparent: First, due to the sophisticated anthropomorphic qualities of generative AI, users often develop unwarranted levels of trust in these systems, enabling platforms like ChatGPT to "efficiently, massively, and covertly manipulate, persuade, and influence users in contextualized and personalized settings through their sophisticated

---

1 Guo Chunzhen, *The Coherent Legal Governance of Generative AI: Taking the Generative Pre-Training Model (GPT) as an Example*, 45(3) Modern Law Science 88, 88 (2023).

2 Philipp Hacker, Andreas Engel & Marco Mauer, *Regulating ChatGPT and Other Large Generative AI Models*, in PROCEEDINGS OF THE 2023 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, at 1113.

interactive capabilities."[3] Second, the large language models underpinning generative AI incorporate substantial amounts of intellectual property-protected content in their pre-training datasets, potentially leading to copyright infringement in their generated outputs.[4] Third, the advanced deep synthesis capabilities of generative AI enable the production of deceptively realistic misinformation.[5] Fourth, since generative AI systems are primarily pre-trained on text corpora from mainstream populations, discriminatory perspectives against minority groups may be inadvertently reflected in their outputs.[6] Fifth, when integrated with downstream internet platforms, generative AI could potentially function as a tool for network centralization in the Web 3.0 era, exacerbating the risk of re-centralization of platform power.[7] Sixth, the pre-training of algorithmic models in generative AI necessitates massive data inputs, including extensive volumes of personal information, thereby creating potential infringement risks related to the breadth of data collection, depth of processing, and use of the results.[8]

Among these diverse risks, this paper specifically examines the challenges that generative AI presents to personal information protection and the associated regulatory issues. This focus is particularly relevant because China implemented its Personal Information Protection Law on November 1, 2021, thereby establishing comprehensive and systematic provisions for personal information protection in the digital era. However, generative AI only emerged prominently at the end of 2022, meaning that considerations related to this technology were not incorporated during the drafting process of the legislation. Consequently, numerous challenges have arisen concerning the law's provisions on personal information protection in the wake of generative AI's rapid development. These challenges potentially create not only new risks for personal information protection in China but also legal constraints on the development of China's generative AI industry. Therefore, the current rise of generative AI necessitates an urgent study of the challenges it poses to China's personal information protection legal framework and, on this basis, the exploration of appropriate risk regulation approaches.

---

[3] Zhang Xin, *Algorithmic Governance Challenges and Regulatory Governance of Generative Artificial Intelligence*, 45(3) Modern Law Science 108, 112 (2023).

[4] Liu Xiaochun, *"Non-Work Use" Nature of Generative Artificial Intelligence Data Training and Its Legitimization*, 39(3) Legal Forum 67, 67 (2024).

[5] Zhang Linghan, *Logic Update and System Iteration of Deep Synthesis Governance: China's Path to Governance of ChatGPT and Other Generative Artificial Intelligence*, 41(3) Science of Law 38, 48 (2023).

[6] Yu Xingzhong, Zheng Ge & Ding Xiaodong, *Six Legal Issues of Generative Artificial Intelligence: Taking ChatGPT as an Example*, 50(2) China Law Review 1, 17–18 (2023).

[7] Chen Quanzhen, *Generative Artificial Intelligence and Re-centralization of Platform Power*, (3) Oriental Law 61, 61 (2023).

[8] Liu Yanhong, *Three Major Security Risks and Legal Regulation of Generative Artificial Intelligence: Taking ChatGPT as an Example*, (4) Oriental Law 29, 32–33 (2023).

The subsequent sections of this paper will analyze the challenges that generative AI poses to three key areas: informed consent requirements for personal information processing, rules governing publicly available personal information, and protections for sensitive personal information, while also examining the technical causes behind these challenges. Furthermore, grounded in the fundamental regulatory concept of "tolerant yet prudent," this paper will explore specific regulatory pathways to address potential personal information infringements by generative AI, aiming to make an intellectual contribution to both the revision of the Personal Information Protection Law and the formulation of the forthcoming "Artificial Intelligence Law."[9]

## II. CHALLENGES POSED BY GENERATIVE AI TO INFORMED CONSENT RULES AND THEIR CAUSES

The "notification–consent" principle constitutes a fundamental cornerstone of China's Personal Information Protection Law concerning the processing of personal information.[10] This principle requires personal information processors, absent applicable statutory exceptions, to notify data subjects and obtain their consent prior to lawfully processing personal information—commonly referred to as the "informed consent rule."[11] Data containing personal information is one of the most critical components of training datasets for generative AI in large language model pre-training. Consequently, generative AI developers qualify as personal information processors under the Personal Information Protection Law[12] and must therefore comply with the informed consent rule when processing personal information. However, the intrinsic technical characteristics of generative AI pose significant challenges to implementing this foundational rule of personal information protection.[13]

---

[9] General Office of the State Council, *Notice on the State Council's 2024 Annual Legislative Work Plan* (SCGO [2024] No. 23, May 9, 2024), proposing to "prepare to submit the Artificial Intelligence Law draft to the Standing Committee of the National People's Congress for deliberation"; Standing Committee of the National People's Congress, *2024 Annual Legislative Work Plan* (2024), including preparatory deliberation items on the "healthy development of artificial intelligence."

[10] LONG WEIQIU ed., INTERPRETATION OF THE PERSONAL INFORMATION PROTECTION LAW OF THE PEOPLE'S REPUBLIC OF CHINA, at 57 (Beijing, China Legal Publishing House, 2021).

[11] Cheng Xiao, *On the Personal Information Processing Rules in China's Personal Information Protection Law*, 15(3) Tsinghua Law Journal 55, 61 (2021).

[12] Article 9 of the Interim Measures for the Administration of Generative Artificial Intelligence Services stipulates that "generative artificial intelligence service providers shall lawfully assume the responsibilities of personal information processors."

[13] Ding Xiaodong, *On Data Institution that Promotes Artificial Intelligence*, 54(6) China Law Review 175, 177 (2023).

## A. Compliance Challenges Faced by Generative AI Developers with the Informed Consent Rule

According to Article 13 of China's Personal Information Protection Law, except under six statutory circumstances, personal information processors must obtain consent from data subjects (individuals) before processing personal information. Additionally, pursuant to Articles 14 and 17, when personal information is processed based on individual consent, such consent must be given after the data subject is fully informed. Personal information processors must truthfully, accurately, and comprehensively inform the data subject of the purposes and methods of personal information processing. Moreover, if the purpose or method of personal information processing changes, the data subject's consent must be obtained again.[14]

The massive volume of training data used in the pre-training of generative AI large language models contains extensive personal information. This enormous training dataset is derived partly from data purchased by developers or extracted from the internet using web crawler technology. For instance, during ChatGPT's large language model pre-training phase, more than 300 billion words of data were harvested from the internet,[15] including substantial amounts of personal information. Another source consists of authentic human–machine interaction data collected by developers in the course of providing large language model services. For example, the human–machine interaction data generated by ChatGPT's global user base serves as training data for OpenAI's model upgrades and iterations, which necessarily include considerable personal information provided by users themselves.

When generative AI developers utilize data containing personal information for the pre-training of large language models, they necessarily engage in activities such as the collection, storage, use, and processing of personal information—activities that align directly with the personal information processing behaviors defined in Article 4 of the Personal Information Protection Law. Consequently, generative AI developers qualify as personal information processors as defined by the legislation and are subject to the

---

[14] Article 7(2) of the Interim Measures for the Administration of Generative Artificial Intelligence Services requires generative artificial intelligence service providers to obtain individual consent when using data involving personal information in data processing activities such as pre-training and optimization training of large language models.

[15] ALEX HUGHES, CHATGPT: EVERYTHING YOU NEED TO KNOW ABOUT OPENAI'S GPT-3 TOOL, Science Focus, at https://www.sciencefocus.com/future-technology/gpt-3 (last visited March 19, 2024).

informed consent requirements set forth in Articles 13, 14, and 17. This means that when generative AI developers process data containing personal information during the pre-training of large language models, barring statutory exceptions, they must, in principle, truthfully, accurately, and comprehensively inform data subjects about processing purposes, methods, and other relevant details, and may only use such data for pre-training activities after the data subjects have been fully informed and have provided consent.

However, despite these legal requirements, the technical characteristics of generative AI render developers effectively unable to fulfill their obligation to truthfully, accurately, and comprehensively inform data subjects about processing purposes, methods, and other relevant information—thus making compliance with the statutory requirements of the informed consent rule practically infeasible. In essence, the technical characteristics of generative AI present fundamental challenges to the applicability of the informed consent rule to personal information processing in the era of artificial intelligence. The following analysis examines the technical causes of this dilemma using the GPT model (Generative Pre-trained Transformer)—the leading model in generative AI—as a case study.

## B. Technical Causes of Generative AI's Challenge to the Informed Consent Rule

The GPT model, developed by OpenAI, is a sophisticated large language model designed to enhance computational understanding and generation of natural language text in complex scenarios.[16] The widely recognized ChatGPT application is built on this foundational model. The GPT model has achieved remarkable accuracy in the computational recognition of human natural language, with one of its key innovations being its ability to use massive volumes of non-annotated data for pre-training.

Data forms the essential foundation of language model pre-training. Theoretically, larger data volumes enable more precise model learning and greater overall effectiveness. Traditional language model training primarily relied on manually annotated data for pre-training, as such data could be readily processed by computers and incorporated into model development. However, manually annotated data presents significant limitations: it is prohibitively expensive to obtain and inherently limited in scale, thereby severely constraining the learning potential of language models. Conversely,

[16] OPENAI, GPT-4 TECHNICAL REPORT, at https://arxiv.org/abs/2303.08774 (last visited March 19, 2024).

the internet contains virtually limitless volumes of non-annotated data, which continue to expand exponentially. The effective use of this non-annotated data for language model pre-training dramatically reduces pre-training costs while substantially improving effectiveness, thereby facilitating the evolution of conventional models into large language models with unprecedented capabilities.

The GPT model employs a sophisticated technical approach to address the challenge of using non-annotated data at scale. This approach conceptualizes each word[17] as a distinct point within an expansive vocabulary space, where semantically related words occupy proximate positions and unrelated words remain distant. Each word undergoes vectorization through a process analogous to geographic coordinate mapping, transforming its position in this multidimensional vocabulary space into a numerical sequence potentially tens of thousands of digits long. These numerical representations encode the complex network of potential associations between each word and all others in the vocabulary. The model then employs massive computational resources to identify statistical regularities in natural language—specifically, to uncover the probabilistic distribution patterns governing word relationships.[18] Subsequently, when provided with user prompts, the GPT model leverages these learned statistical patterns to predict successive words through an autoregressive process, thereby generating coherent and comprehensible content for users.[19] Crucially, unlike traditional search engines that retrieve pre-existing information from stored databases in response to queries, the GPT model dynamically generates information by "predicting the next word" based on probability distributions derived from its pre-training on natural language patterns.[20] This innovative approach enables the GPT model to effectively use virtually all non-annotated data available on the internet for pre-training purposes, achieving unsupervised learning while significantly reducing data preparation costs and enhancing natural language processing capabilities.

---

[17] Strictly speaking, it should be expressed as "token." In the pre-training technology of large language models, a "token" could be a word, a character, or a text fragment, and the choice depends on the specific application scenario. For ease of understanding, this article uniformly uses the term "word."

[18] Yu Shiwen, Zhu Xuefeng & Gheng Libo, *Natural Language Processing Technology and Deep Language Computing*, (3) Social Sciences in China 127, 129–130 (2015).

[19] Amy B. Cyphert, *A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law*, 55 UC Davis Law Review 406, 406–407 (2021).

[20] Laura Weidinger *et al.*, *Taxonomy of Risks Posed by Language Models*, in PROCEEDINGS OF THE 2022 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, at 215–216 (2022).

However, while dramatically improving the effectiveness of language models, this technical approach introduces a fundamental problem: the resulting large language models function essentially as impenetrable technical black boxes. Although generative AI successfully captures implicit statistical patterns in natural language through its pre-training process, these patterns reside within model parameters numbering in the hundreds of billions[21]—a representation fundamentally different from, and less transparent than, conventional data storage. These patterns remain largely inscrutable, even to the developers themselves. This inscrutability reflects an inherent cognitive opacity attributable to the multi-layered neural network architectures that underpin contemporary artificial intelligence.[22] The emergence of this comprehensive technical black box means that personal information processing during generative AI pre-training is similarly opaque. Even the developers themselves cannot precisely determine how personal information will be processed within the model, what linguistic patterns might be derived from such information, what outputs might be generated in downstream applications, or for what purposes—thereby significantly complicating the assessment of rights implications and privacy risks associated with personal information used in large language model pre-training.[23]

Given that even developers cannot comprehend how personal information is processed within the technical black box of large language models, providing detailed notification to data subjects becomes effectively impossible, rendering compliance with the informed consent rule for personal information processing practically unattainable. Indeed, even under the most expansive interpretation of notification requirements—broadly defining the purpose as simply "for use in the pre-training of generative AI large language models"—developers would still encounter substantial compliance difficulties due to the sheer volume of training data and the vast amount of personal information embedded therein, making it objectively impossible to notify each affected data subject and obtain their consent.

The evidence clearly indicates that generative AI's distinctive technical approach to processing massive volumes of non-annotated data during large language model pre-training creates a technical black box that makes it objectively difficult for developers

---

[21] Sun Meng'ge *et al*., *Analysis of the Impact of GPT Technology Revolution on Basic Scientific Research*, 38(8) Bulletin of Chinese Academy of Sciences 1212, 1213 (2023).

[22] Dong Chunyu, *The Nature of AI and Its Limits in the Light of the Opacity of Machine Cognition*, (5) Social Sciences in China 148, 159 (2023).

[23] Yuan Zeng, *On the Legal Capacity of Generative Artificial Intelligence*, (3) Oriental Law 18, 24 (2023).

to fulfill their legal obligation to truthfully, accurately, and comprehensively inform data subjects about processing purposes, methods, and related details—let alone ensure that data subjects can exercise meaningful informed consent rights. This technical reality fundamentally challenges the practical application of the informed consent rule established by the Personal Information Protection Law in the context of generative AI.

## III. CHALLENGES POSED BY GENERATIVE AI TO RULES FOR PROCESSING PUBLICLY AVAILABLE PERSONAL INFORMATION AND THEIR CAUSES

Within the vast training datasets used by generative AI developers to construct large language models, alongside personal information requiring prior consent from data subjects, there exists a distinct category of personal information that may be processed without such consent—specifically, personal information that has been lawfully made public ("publicly available personal information"). According to Articles 13(6) and 27 of China's Personal Information Protection Law, data processors have the right to process public personal information (including information disclosed by individuals themselves or through other lawful channels) within a "reasonable scope" without obtaining consent from data subjects, thereby establishing a default rule governing such processing.[24] Consequently, generative AI developers may directly process public personal information within reasonable parameters without being subject to informed consent requirements. The determination of what constitutes a "reasonable scope" for processing public personal information has traditionally relied on two primary standards in legal theory: the purpose limitation principle and contextual integrity theory. However, within the technical framework of generative AI, both standards face formidable challenges, rendering compliance exceedingly difficult for generative AI development and application.

### A. Technical Causes of Generative AI's Challenge to the Purpose Limitation Principle

The purpose limitation principle stipulates that the processing of public personal information should be confined to the original purpose for which the data subject disclosed such information, with this purpose-defined scope constituting the reasonable

[24] Zhang Weiwei, *The Default Rule for Processing Public Personal Information: Based on the First Sentence of Article 27 of the Personal Information Protection Law*, 41(3) Science of Law 62, 65 (2023).

parameters for processing.[25] Applied to generative AI, this principle would require developers conducting large language model pre-training to restrict their processing of public personal information exclusively to the original purpose for which it was disclosed in order to remain within a reasonable scope. However, the inherent technical characteristics of generative AI make compliance with this requirement virtually impossible.

As previously discussed, a fundamental technical principle of generative AI, exemplified by the GPT model, involves developers leveraging immense computational power and massive training datasets to pre-train large language models, enabling them to internalize the statistical patterns inherent in natural language. Subsequently, based on user prompts, the model generates outputs conforming to these learned patterns through its "next-word prediction" mechanism, facilitating human–machine interaction via natural language interfaces. The linguistic statistical patterns assimilated by large language models through this technical approach exhibit extraordinary complexity. To conceptualize this complexity, we might metaphorically equate the number of parameters in a large language model to the number of potential pathways between consecutive words, with more parameters representing more pathways. ChatGPT, for instance, incorporates 175 billion parameters, effectively establishing 175 billion potential connections between any given word and its possible successors. Moreover, this represents only the pathways between two consecutive words; pre-training datasets typically contain billions of words, with comparable numbers of pathways potentially existing between each word pair. The computational demands required for a large language model to manage this vast network of pathways during pre-training and to identify those with the highest probability values are immense, necessitating unprecedented computational resources.

This extreme complexity renders the natural language statistical patterns mastered by large language models fundamentally opaque—even to their developers. How these models process training data containing public personal information remains essentially unknowable—just as the nature of the statistical patterns extracted from such information is. The large language model functions as a comprehensive technical black box. This technical inscrutability makes it impossible for generative AI developers to restrict the processing of public personal information to specific purposes—let alone confine such processing to the original purpose for which the information was disclosed.

---

[25] Cheng Xiao, *On the Legal Regulation of the Processing of Public Personal Information*, (3) China Legal Science 82, 99 (2022).

Consequently, generative AI development inherently struggles to satisfy the requirements of the purpose limitation principle regarding the "reasonable scope" of processing public personal information.

## B. Technical Causes of Generative AI's Challenge to the Contextual Theory

The contextual theory advocates considering the distinctive characteristics of specific contexts when processing public personal information, with the reasonable scope of processing determined according to these contextual differences.[26] When applied to generative AI, this approach would base the determination of a reasonable processing scope on the diverse application contexts of large language models. However, the inherent technical characteristics of generative AI similarly hinder compliance with this context-based reasonableness standard.

Large language models developed through generative AI technology, exemplified by the GPT model, earn the "large" designation primarily due to their extraordinary parametric scale. GPT-3 incorporates 175 billion parameters, while GPT-4, released by OpenAI in 2023, is estimated by scholars to potentially contain up to 1.8 trillion parameters, though the exact figure remains undisclosed.[27] This unprecedented parametric scale enables large language models to exhibit "emergent" capabilities—the ability to autonomously generate effective outputs in response to diverse task prompts without specific prior training.[28] For example, GPT-4, without task-specific training, has successfully passed sophisticated professional examinations, including the U.S. Bar Exam and the LSAT, achieving scores in the top 10th percentile.[29]

This emergent capability enables large language models to generate content that satisfies diverse user information needs while providing exceptional extensibility, facilitating human–machine interaction across numerous domains through adaptation to varied application contexts. This versatility positions these models as prototypes of general artificial intelligence. Indeed, the revolutionary significance of generative AI in the

---

[26] Qi Yingcheng, *A Typological Interpretation of the Rules for Processing Public Personal Information*, 28(5) Law and Social Development 210, 217–219 (2022). Helen Nissenbaum & Wang Yuan (trans.), *Privacy as Contextual Integrity*, (1) Journal of Cyber and Information Law 3, 3–28 (2021).

[27] DYLAN PATEL & GERALD WONG, DEMYSTIFYING GPT-4: THE ENGINEERING TRADEOFFS THAT LED OPENAI TO THEIR ARCHITECTURE, at https://www.semianalysis.com/p/gpt-4-architecture-infrastructure (last visited on March 19, 2024).

[28] JASON WEI *et al*., EMERGENT ABILITIES OF LARGE LANGUAGE MODELS, at https://arxiv.org/abs/2206.07682 (last visited on March 19, 2024).

[29] OPENAI, *supra* note 16.

digital era stems precisely from the capacity of its large language models to function as foundational general-purpose models adaptable to countless specific applications. This technical characteristic means that large language models can be deployed across an essentially unlimited range of contexts—so broad that even their developers cannot anticipate all the specific contexts in which the model's processing of public personal information might ultimately be applied, as these systems are inherently designed for general, rather than narrowly defined, contexts. Consequently, applying contextual theory to determine whether generative AI's processing of public personal information falls within a reasonable scope becomes equally problematic. In the practically infinite application landscape of generative AI, context-specific determinations of a reasonable processing scope for publicly available personal information represent an idealized standard that lacks practical implementability.

It becomes evident that within the technical framework of generative AI, neither the purpose limitation principle nor the contextual theory provides an adequate basis for effectively defining reasonable parameters for the processing of publicly available personal information by large language models. The processing rules established by the Personal Information Protection Law are facing significant challenges, necessitating urgent attention and resolution as generative AI continues to evolve rapidly.

## IV. CHALLENGES TO THE PROTECTION OF SENSITIVE PERSONAL INFORMATION BY GENERATIVE AI AND THEIR TECHNICAL CAUSES

The extensive training datasets employed by generative AI developers contain substantial amounts of sensitive personal information alongside standard personal data. Under Article 28 of the Personal Information Protection Law, sensitive personal information includes "biometric recognition, religious belief, specific identity, medical health, financial accounts, whereabouts tracking, and other information, as well as the personal information of minors under the age of fourteen." This category of information is intrinsically linked to individuals' privacy interests.[30] Consequently, when generative AI developers process sensitive personal information, they risk infringing not only individuals' rights concerning that information but also their fundamental privacy rights. The technical characteristics that render generative AI developers unable to comply with informed consent requirements

[30] Shen Weixing, *Reconstruction of Digital Rights System: Towards a Differential Pattern of Privacy, Information and Data*, (3) Tribune of Political Science and Law 89, 97 (2022).

for standard personal information—as examined in Part Two of this paper—apply equally to the processing of sensitive personal information. Particularly concerning, however, is that generative AI's technical architecture creates risks of infringing upon data subjects' sensitive personal information rights and privacy interests at both the input and output stages of large language model operations.

### A. Technical Causes of Infringement Risks in LLM Inputs

Sensitive personal information incorporated into generative AI training datasets is derived not only from traditional sources, such as purchased data or web crawler extractions, but also from "Machine Learning as a Service" (MLaaS) interfaces. MLaaS refers to cloud-based services offered by generative AI developers that allow users to access large language models remotely without local deployment. This arrangement means that user inputs are necessarily uploaded to and stored on the developers' servers. When users lack a clear understanding of the model's data processing mechanisms, they may inadvertently submit sensitive personal information.[31] These user interactions—comprising both inputs containing potentially sensitive personal information and corresponding model outputs—constitute authentic human–machine interaction data subsequently used by developers for model refinement and retraining. OpenAI's ChatGPT user terms explicitly acknowledge that interaction data will be repurposed as training data for future iterations. Significantly, users cannot selectively delete sensitive personal information they may have unwittingly provided.[32] This creates a cumulative effect, whereby large language models contain sensitive personal information both in their initial training datasets and in supplementary data acquired through iterative user interactions.

When sensitive personal information becomes embedded in large language model training data, the model effectively "memorizes" this information, creating potential leakage vectors and consequent risks to data subjects' rights and privacy.[33] Recent research demonstrates that technical specialists can extract substantial amounts of original training data from various large language models using advanced techniques—inevitably including sensitive personal information contained within that training

---

[31] Zhang Xin, *Data Risks and Governance Paths of Generative Artificial Intelligence*, 41(5) Science of Law 42, 46 (2023).

[32] Zhi Zhenfeng, *Information Content Governance of Large Language Models in Generative Artificial Intelligence*, 41(4) Tribune of Political Science and Law 34, 40 (2023).

[33] Liu Jinrui, *Regulatory Framework for New Risks of Large Generative AI Models*, (2) Administrative Law Review 17, 20 (2024).

data.[34] This potential for privacy infringement prompted sixteen anonymous plaintiffs to file suit against OpenAI and Microsoft on June 28, 2023, alleging that ChatGPT and related products collected private information from millions of individuals—including names, email addresses, payment details, transaction records, chat logs, and search histories—potentially revealing religious beliefs, political views, sexual orientation, and personal preferences in violation of the U.S. Electronic Communications Privacy Act. This litigation indirectly highlights generative AI's capacity to compromise sensitive personal information rights and individual privacy.[35] In response to similar concerns, Italy's Personal Data Protection Authority temporarily suspended ChatGPT's operation in March 2023, permitting its resumption only after OpenAI implemented privacy policy updates and technical modifications. Data protection authorities in France and the United Kingdom have similarly expressed specific concerns regarding the privacy implications of ChatGPT and comparable systems.[36]

## B. Technical Causes of Infringement Risks in LLM Outputs

The output functionality of large language models presents risks to sensitive personal information rights and individual privacy that are equally significant as those posed by their input mechanisms.

This vulnerability stems from large language models' sophisticated information integration capabilities developed during pre-training. While acquiring statistical patterns in natural language, these models simultaneously develop powerful abilities to analyze and synthesize disparate fragments of information from training data into coherent, comprehensive content. This capability enables models to aggregate scattered fragments of personal information across digital environments,[37] facilitating deep user profiling that can extract sensitive personal data and intimate privacy details embedded within seemingly innocuous information fragments. Under the "long-tail effect," this creates substantial risks of compromising data subjects' sensitive personal information rights and fundamental privacy interests.[38] Indeed, this powerful information synthesis

---

[34] MILAD NASR *et al*., SCALABLE EXTRACTION OF TRAINING DATA FROM (PRODUCTION) LANGUAGE MODELS, at https://arxiv.org/abs/2311.17035 (last visited on March 19, 2024).

[35] P.M. v. OpenAI LP, at https://www.courtlistener.com/docket/67535351/pm-v-openai-lp/ (last visited on March 19, 2024).

[36] Fu Hongyu, *Governance Models and Risk Analysis of Generative Artificial Intelligence*, (4) Digital Law 191, 197–198 (2023).

[37] Guo Chunzhen, *supra* note 1, 98.

[38] Bi Wenxuan, *The Dilemma in the Risk Regulation of Generative Artificial Intelligence and Its Resolution: Taking ChatGPT as an Example*, (3) Journal of Comparative Law 155, 159–160 (2023).

capability potentially grants generative AI developers unprecedented, comprehensive access to individuals' personal information portfolios, including highly sensitive data. Furthermore, as previously noted, generative AI represented by GPT models constitutes a nascent form of general artificial intelligence, with their massive parametric scale enabling emergent capabilities applicable across diverse contexts. Consequently, these powerful information integration and analysis capabilities can propagate through various downstream applications, potentially enabling ordinary end users to access others' sensitive personal information and intrude upon their privacy with minimal effort or technical expertise—significantly amplifying infringement risks.[39]

A critical technical limitation compounds these concerns: unlike conventional databases, where specific data entries can be directly modified or deleted, generative AI systems encode learned patterns within model parameters rather than storing individual data points discretely. Consequently, even when developers discover their models outputting sensitive personal information in violation of privacy rights, they cannot simply delete the offending information as they would in a traditional database architecture. Instead, preventing the disclosure of specific sensitive information requires comprehensive model retraining—a resource-intensive process that is not feasible on demand. This technical constraint creates significant obstacles to promptly addressing sensitive information leakage, perpetuating risks to data subjects' rights and privacy interests.

## V. RISK REGULATION PATHWAYS FOR GENERATIVE AI-INDUCED INFRINGEMENT OF PERSONAL INFORMATION

Generative AI's revolutionary technological advances have created multifaceted challenges to China's personal information protection framework and have highlighted potential infringement risks to the integrity of personal information. As generative AI technologies continue to expand across economic and social sectors, these infringement risks potentially threaten not only individual data subjects with devastating consequences but also pose significant security implications for broader society and national interests.[40] In response, traditional "rights-based" approaches to personal information protection have progressively

---

[39] Xu Wei, *On the Legal Status and Liability of Generative Artificial Intelligence Service Providers: Taking ChatGPT as an Example*, (4) Science of Law 69, 77 (2023).

[40] Liu Quan, *Personal Information Protection from the Perspective of Risk Governance*, (2) Journal of Comparative Law 62, 63 (2024).

evolved into "risk-based" methodologies.[41] This conceptual shift has necessitated a corresponding transformation in the primary mechanisms of personal information protection in the generative AI era—specifically, a transition from private law mechanisms centered on civil litigation to public law frameworks emphasizing proactive risk regulation.[42] The private law protection pathway, implemented through judicial proceedings, suffers from inherent limitations, including high transaction costs and remedial delays, often rendering it ineffective for the timely prevention of personal information infringements. By contrast, public law protection mechanisms implemented by administrative authorities offer greater professional expertise and operational effectiveness, providing superior adaptability to the rapidly evolving technological landscape of artificial intelligence. Consequently, amid the flourishing development of generative AI, effectively protecting personal information requires a careful examination of relevant risk regulation approaches to inform both the revision of the Personal Information Protection Law and the formulation of the forthcoming "Artificial Intelligence Law."[43]

### A. "Tolerant Yet Prudent" as Guiding Philosophy of Risk Regulation

While generative AI presents potential risks to the integrity of personal information, this does not justify prohibiting its development. Within the context of the current technological revolution, generative AI represents a cutting-edge innovation requiring robust cultivation and support. Consequently, risk regulation in this domain must maintain a delicate balance between fostering technological advancement and safeguarding personal information—neither sacrificing information protection for technological progress nor impeding innovation for the sake of absolute information security.

This balanced approach is embodied in the regulatory concept of "tolerant yet prudent,"[44] which seeks to achieve a "reconfiguring of obligations and liabilities" within a flexible yet conscientious framework.[45] From the "tolerance" perspective, regulation should facilitate

---

41 Zhang Tao, *Exploring the Dimensions of Risk Control Paths for Personal Information Protection*, (6) Law Science 57, 62–65 (2022).

42 Wang Xixin, *Rethinking the Protection Mechanism of Personal Information Rights: Administrative Supervision or Civil Litigation*, 44(5) Chinese Journal of Law 3, 3 (2022).

43 Zhao Jingwu, *Theoretical Misconceptions and Path Reorientation of Risk Governance for Generative Artificial Intelligence Applications*, (3) Jingchu Law Review 47, 48–50 (2023).

44 Zhang Linghan, *The Legal Positioning and Hierarchical Governance of Generative AI*, 45(4) Modern Law Science 126, 139 (2023). Zhang Xin, *Industry Chain-Oriented Governance: Technological Mechanisms and Governance Logic in the Management of Artificial Intelligence Generated Content*, (6) Administrative Law Review 43, 50–59 (2023).

45 Han Xuzhi, *The Logical Updates and Path Optimization for Governing Generative AI*, (6) Administrative Law Review 30, 37 (2023).

Chinese generative AI developers' lawful access to, and processing of, personal information. If developers were required to obtain explicit informed consent from every data subject before incorporating data into large language model pre-training—regardless of implementation feasibility—China's competitive position in the global technological landscape could be significantly compromised. Therefore, the tolerance dimension of risk regulation advocates for an appropriate relaxation of informed consent requirements, aiming to "control excessive safety redundancy"[46] and enable more efficient use of personal information data for pre-training purposes, thereby realizing the inherent social value of such information.[47]

Concurrently, from the "prudence" perspective, regulation must prioritize the protection of data subjects' personality rights. In the digital economy, personal information encompasses both personality interests and property interests of data subjects.[48] Generative AI's potential infringement risks may affect both dimensions. As a transformative technology, generative AI promises societal and economic benefits that far exceed the property value of individual personal information. Consequently, the protection standards for property rights in personal information may be moderately relaxed to facilitate the growth of the generative AI industry. However, data subjects' personal dignity represents an inviolable threshold that must not be compromised for economic gain. Thus, the prudential dimension of risk regulation requires maintaining stringent protections for personality rights and imposing strict regulations on personal information processing that could compromise these fundamental interests.

In essence, "tolerant yet prudent" represents a foundational regulatory approach that seeks to balance technological innovation with personal information protection. The comprehensive regulatory strategy should facilitate Chinese developers' use of personal information for large language model development, while prioritizing the protection of data subjects' personality rights and appropriately moderating the protection of property rights.

## B. Specific Risk Regulation Pathways

Building upon the "tolerant yet prudent" concept, specific regulatory pathways to address the personal information infringement risks posed by generative AI include:

---

[46] Su Yu, *Legal Risks and Governance Paths of Large Language Models*, 42(1) Science of Law 76, 85 (2024).

[47] Gao Fuping, *Personal Information Protection: From Personal Control to Social Control*, 40(3) Chinese Journal of Law 84, 96 (2018).

[48] Zhang Xinbao, *Rights Allocation for Structural Separation of Data Property Rights*, 45(4) Global Law Review 5, 18 (2023); Liu Deliang, *Proprietary Right Protection of Personal Information*, 29(3) Chinese Journal of Law 80, 80 (2007).

### *1. Adapting Informed Consent Approaches for Generative AI*

Under the current provisions of China's Personal Information Protection Law, most personal information processing requires explicit consent from data subjects, with limited statutory exceptions. The law does not specifically address the unique processing requirements of generative AI. Consequently, the vast amounts of personal information used in large language model pre-training theoretically require prior explicit consent—creating a substantial regulatory barrier to efficient development that necessitates modification.

Consistent with the "tolerant yet prudent" framework focusing on the protection of personality rights, an "implied consent" rule could be adopted for ordinary personal information that generally does not implicate such rights—permitting generative AI developers to process this information unless data subjects explicitly refuse. Simultaneously, sensitive personal information with a greater potential to affect personality rights would remain subject to the explicit consent requirement. This bifurcated approach maximizes development efficiency while maintaining robust protections for information most likely to affect fundamental personality interests.

This adjustment could be implemented through amendments to the Personal Information Protection Law, designating generative AI development as a special exception to standard informed consent requirements. Alternatively, if direct amendment proves challenging, the "Artificial Intelligence Law" currently under development could incorporate provisions designating implied consent as the default rule for processing ordinary personal information in generative AI contexts. Under the principle that "special law prevails over general law," these provisions would take precedence over general personal information processing requirements, thereby effectively implementing the implied consent framework without necessitating amendments to existing legislation.

### *2. Reforming Processing Rules for Publicly Available Personal Information in Generative AI*

The Personal Information Protection Law currently restricts the processing of publicly available personal information to a "reasonable scope"—a limitation that generative AI's technical characteristics render virtually impossible to satisfy under traditional interpretive frameworks, such as the purpose limitation principle or contextual theory.

Strict application of existing provisions would place generative AI developers in continuous legal jeopardy, severely constraining industry development.

Rather than attempting to positively define the "reasonable scope" for generative AI's processing of publicly available information—an approach fraught with practical difficulties—regulation might more effectively establish clear boundaries through negative delineation. Specifically, legislation could establish the protection of data subjects' personality rights as the non-negotiable bottom line for processing publicly available personal information, permitting any activities that do not transgress this threshold. This approach prioritizes the protection of personality rights while relaxing property rights constraints, thereby facilitating generative AI development while maintaining essential safeguards. While this approach may limit individuals' ability to monetize their personal information, it ultimately serves the broader public interest by enabling generative AI advancement and its attendant societal benefits in the digital economy era.

### *3. Developing Administrative Safeguards for Personality Rights in Generative AI*

Current mechanisms for protecting personality rights in personal information in China rely predominantly on civil litigation initiated by affected individuals—a private law remedy framework well-suited for addressing discrete infringements in the pre-generative AI context. However, the generative AI paradigm—through its processing of massive data volumes for large language model pre-training—fundamentally transforms the nature of potential infringements from isolated incidents into systemic, large-scale violations affecting numerous subjects simultaneously. Continued reliance on private law remedies in this environment presents significant shortcomings: not only do such approaches suffer from inherent delays in establishing legal responsibility, but the substantial costs associated with litigation effectively preclude meaningful recourse for most affected individuals, thereby severely limiting the protective function of existing mechanisms.

Therefore, guided by the "tolerant yet prudent" framework, effective administrative regulatory measures must be established to complement private law remedies in the protection of personality rights in the generative AI landscape. Specific enhancement pathways include:

First, establish administrative penalties against generative-AI developers who infringe upon personality rights in personal information. While China's Personal Information

Protection Law establishes penalties for processors who violate statutory information processing requirements, these provisions remain overly general and fail to distinguish between property rights infringements and the more fundamental violations of personality rights. Legislative amendments should establish clear administrative sanctions specifically for personality rights infringements by generative AI developers, while explicitly excluding property rights infringements from this enhanced penalty framework—thereby reinforcing the primacy of personality rights protection as a regulatory priority.

Second, create administrative compensation schemes. While affected individuals technically retain the right to seek compensation through civil tort litigation when personality rights are infringed, the structural power imbalance between individual data subjects and institutional developers creates significant procedural disadvantages. Moreover, the individualized nature of civil litigation means that successful claims benefit only the specific plaintiffs, requiring separate proceedings for each affected subject—an approach marked by profound inefficiency. To address this limitation, administrative authorities should be empowered to issue compensation orders requiring infringing developers to remedy personality rights violations. This approach leverages public administrative authority to reduce barriers to compensation while reinforcing the preferential protection granted to personality rights within the regulatory framework.

Third, establish administrative licensing regime for generative AI development activities. While competitive market forces can accelerate generative AI advancement, unregulated competition inevitably leads to qualitative disparities among developers and creates environments where technical opacity facilitates rights violations. Administrative licensing requirements would restrict generative AI development to organizations that meet predetermined qualifications[49]—for instance, by requiring developers to establish comprehensive internal safeguards for personality rights protection as a precondition for licensure. This approach not only filters out entities lacking adequate protective capabilities but also creates a more clearly defined regulatory landscape that enables more effective administrative oversight of development activities.

---

[49] Sun Qi, *Research on Legal Issues of Governing the Providers of Generative Artificial Intelligence Products*, (7) Political Science and Law 162, 174 (2023).

## VI. CONCLUSION

In summary, generative AI, exemplified by the GPT model, represents a fundamental departure from traditional decision-making artificial intelligence, transcending predictive functionality to achieve content generation based on statistical patterns identified through large language model pre-training. This development signifies the nascent realization of general artificial intelligence and heralds the emergence of transformative technological capabilities in the digital economy era. However, this advancement simultaneously introduces novel risks, with challenges to personal information protection frameworks posing particularly significant concerns. These challenges arise directly from generative AI's technical characteristics and manifest primarily across three dimensions: complications in applying the informed consent rule to personal information processing; difficulties in using existing frameworks for processing publicly available personal information; and heightened vulnerabilities in protecting sensitive personal information. Addressing these challenges requires the implementation of targeted regulatory measures guided by the "tolerant yet prudent" principle, either during the revision of the Personal Information Protection Law or in the formulation of the forthcoming "Artificial Intelligence Law."

Specific recommended measures include: adjusting informed consent requirements to establish differentiated standards for ordinary versus sensitive personal information; reframing rules for processing publicly available personal information to establish personality rights protection as a clear regulatory boundary; and developing administrative regulatory mechanisms specifically tailored to the unique characteristics of generative AI's potential infringements of personality rights. These coordinated regulatory approaches aim to achieve a meaningful balance between fostering generative AI innovation and upholding essential protections for personal information in the emerging digital economy landscape.